

SPINN: Sparse, Physics-based, and partially Interpretable Neural Networks for PDEs[☆]



Amuthan A. Ramabathiran^{a,b,*}, Prabhu Ramachandran^{a,b}

^a Department of Aerospace Engineering, Indian Institute of Technology Bombay, Mumbai - 400076, Maharashtra, India

^b Center for Machine Intelligence and Data Science (C-MInDS), Indian Institute of Technology Bombay, Mumbai - 400076, Maharashtra, India

ARTICLE INFO

Article history:

Available online 28 July 2021

Keywords:

Physics-based neural networks
Sparse neural networks
Interpretable machine learning
Partial differential equations
Meshless methods
Numerical methods for PDEs

ABSTRACT

We introduce a class of Sparse, Physics-based, and partially Interpretable Neural Networks (SPINN) for solving ordinary and partial differential equations (PDEs). By reinterpreting a traditional meshless representation of solutions of PDEs we develop a class of sparse neural network architectures that are partially interpretable. The SPINN model we propose here serves as a seamless bridge between two extreme modeling tools for PDEs, namely dense neural network based methods like Physics Informed Neural Networks (PINNs) and traditional mesh-free numerical methods, thereby providing a novel means to develop a new class of hybrid algorithms that build on the best of both these viewpoints. A unique feature of the SPINN model that distinguishes it from other neural network based approximations proposed earlier is that it is (i) interpretable, in a particular sense made precise in the work, and (ii) sparse in the sense that it has much fewer connections than typical dense neural networks used for PDEs. Further, the SPINN algorithm implicitly encodes mesh adaptivity and is able to handle discontinuities in the solutions. In addition, we demonstrate that Fourier series representations can also be expressed as a special class of SPINN and propose generalized neural network analogues of Fourier representations. We illustrate the utility of the proposed method with a variety of examples involving ordinary differential equations, elliptic, parabolic, hyperbolic and nonlinear partial differential equations, and an example in fluid dynamics.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

There has been a flurry of activity in the recent past on the application of machine learning algorithms to solve Partial Differential Equations (PDEs). Unlike traditional methods like the finite element, finite volume, finite difference, and mesh-free methods, Deep Neural Network (DNN) based methods like Physics Informed Neural Networks (PINN) [29] and the Deep-Ritz method [7] circumvent the need for traditional mesh-based representations and instead use a DNN to approximate solutions to PDEs. The idea of using DNNs to solve PDEs is not new [19,20], but their usage has exploded in the recent past. A non-exhaustive list of other deep learning approaches for solving PDEs include [2,38,12,24,36,42,22,3,40,5,25,21]. A drawback with such DNN based techniques, apart from their marked inefficiency in comparison with traditional mesh based methods for lower dimensional PDEs, is the fact that they are difficult to interpret and involve many arbitrary choices related to the network architecture.

[☆] Author names listed alphabetically. Both authors contributed equally to the work.

* Corresponding author.

E-mail addresses: amuthan@aero.iitb.ac.in (A.A. Ramabathiran), prabhu@aero.iitb.ac.in (P. Ramachandran).

In this work, we propose a new class of Sparse, Physics-based, and partially Interpretable Neural Network (SPINN) architectures to solve PDEs that are both interpretable, in a sense to be made precise later, and efficient. DNNs have been studied in the context of meshless methods in works such as [14,42], and connections between DNNs and meshless methods have been noted in works like [6,4]. The key idea behind SPINN, which distinguishes it from other works cited above, is the observation that certain meshless approximations can be directly transcribed into a sparse DNN. We demonstrate that such a simple re-expression of meshless approximations as sparse DNNs allows us to bridge the emerging field of scientific machine learning and the well established methods of traditional scientific computing.

To set the stage for introducing SPINN, we note that a connection between ReLU DNNs and piecewise linear finite element approximations was proved in [15]. This shows that basis functions with compact support can be represented as a DNN. We generalize this to represent kernel functions in meshless methods as DNNs. We use this to construct the SPINN architecture, which is a new and partially interpretable DNN. This is significant in light of the notorious interpretability problem that attends the use of DNNs. We point out that once the DNN is replaced by a sparse equivalent network based on the aforementioned reinterpretation of meshless methods, we use an approach similar to PINN in modeling the boundary conditions and the loss function directly using the PDE. A further novelty of our method is that it naturally suggests neural network analogues of commonly used transforms such as the Fourier and wavelet transforms. We illustrate how Fourier decomposition can be accomplished using special sparse architectures in one dimension, and suggest natural neural network generalizations that go beyond traditional transformations.

The interpretation of deep neural networks in general, and PINN in particular, as a basis decomposition is not new. For instance, the authors in [4] propose new training algorithms for deep neural networks based on this adaptive basis interpretation, and show applications to PINN too. In [9], the authors use this interpretation to develop greedy algorithms for function approximation using neural networks. We remark that while the interpretation of DNNs via basis functions is not new, what distinguishes SPINN is that fact that we take advantage of this interpretation to provide a new class of sparse network architectures that generalize traditional meshless methods along the lines of deep neural networks. Further, the basis functions that we employ in SPINN are shifted and scaled versions of a single *parent* basis function, unlike a collection of distinct basis functions that one would get from using DNNs. We clarify this issue further in the discussion on interpretability.

We further distinguish SPINN from other methods in the literature like [41,44,26] that first extract system specific features from the inputs, use these as inputs to a DNN, and finally extract the desired solution from the outputs of this DNN. In contrast, the primary point of departure of the method we propose is that the inputs are retained as such, but domain specific features are abstracted into kernels in a manner that the overall network architecture is much sparser than fully connected DNNs. We present many examples to illustrate this in the sequel.

One of the primary advantages of SPINN is that it serves as a natural bridge between traditional meshless methods and methods that use DNNs, including both of these as special cases. This permits us to develop solution methodologies for PDEs that are based on neural networks and yet are potentially computationally competitive with traditional methods. Further, the fact that our proposed method is partially interpretable opens up new roads for traditional computational methods to be enhanced using insights from DNNs, and vice versa. For instance, we also present in this work a new class of implicit finite difference and neural network based solution schemes for time dependent partial differential equations. We thus point out the unique role of SPINN in unifying both traditional and modern numerical methods, while at the same time generalizing them. Further, the fact that SPINN generalizes meshless methods using DNNs facilitates differentiable programming, which is difficult to implement with traditional methods.

The rest of the paper is structured as follows: we begin with an introduction of the SPINN architecture by highlighting the exact relation between certain meshless representations and DNNs. We also show how Fourier representations of functions can be handled in the same framework. We then present a variety of examples involving ordinary and partial differential equations to illustrate the method. We conclude with a discussion of the key ideas presented in this work, along with directions for future investigations. Additional details about some of the simulations are presented in the appendices. The code developed to implement SPINN along with an automation script that reproduces every result reported here can be found at <https://github.com/nn4pde/SPINN>.

2. Methodology

We present here details of the SPINN architecture in a systematic fashion starting with an illustration of how certain meshless methods using radial basis functions can be exactly represented as a deep neural network. We subsequently use this to develop the general SPINN framework that goes beyond traditional meshless methods, and highlight the precise sense in which SPINN is interpretable. We then provide various details about the way the SPINN architecture is used in solving PDEs. We also propose a generalization of SPINN based on Fourier decomposition.

2.1. Meshless approximation using radial basis functions

To set the stage for the introduction of Sparse Physics-based and partially Interpretable Neural Networks (SPINNs), we focus on the problem of finding a solution $u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ ($d \geq 1$) of the partial differential equation $\mathcal{N}(x, u(x), \nabla u(x), \dots) =$

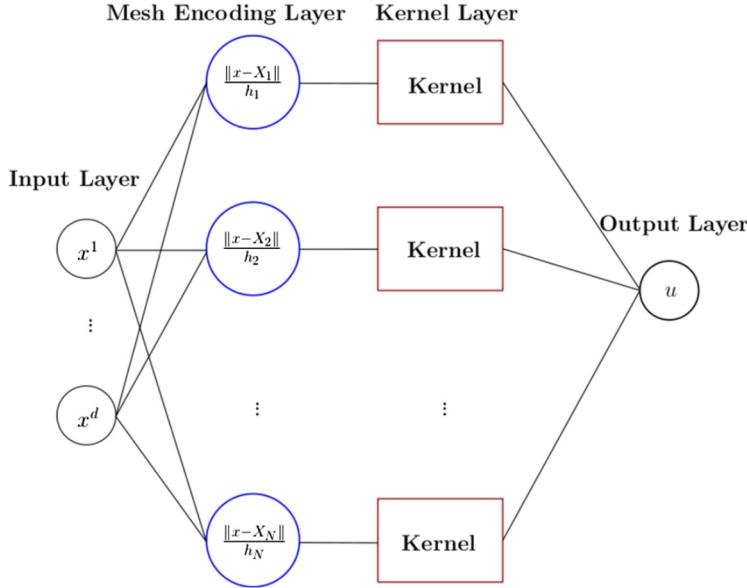


Fig. 1. Simplified SPINN architecture. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$0, x \in \Omega$, with specified Dirichlet and Neumann boundary conditions on the domain boundary $\partial\Omega$. Among the many numerical techniques that have been developed to solve such equations we focus in particular on a class of meshless methods that approximate the solution of the differential equation in terms of Radial Basis Functions (RBFs): for every $x = (x^1, \dots, x^d) \in \Omega$,

$$u(x) = \sum_{i=1}^N U_i \varphi \left(\frac{\|x - X_i\|}{h_i} \right). \quad (1)$$

The various terms in the approximation (1) are to be understood as follows. $(X_i \in \mathbb{R}^d)_{i=1}^N$ represent nodes in the domain Ω . $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ represents an RBF kernel. The variables $(h_i \in \mathbb{R})_{i=1}^N$ are appropriately defined measures of width of the RBF kernels centered at the nodes (X_i) . Finally, the coefficients $(U_i \in \mathbb{R})_{i=1}^N$ represent the nodal weights associated with the basis functions centered at $(X_i)_{i=1}^N$. In the sequel we also consider a variant of the meshless approximation (1) that enforces the partition of unity property; such an approximation takes the form

$$u(x) = \left(\sum_{j=1}^N \varphi \left(\frac{\|x - X_j\|}{h_j} \right) \right)^{-1} \sum_{i=1}^N U_i \varphi \left(\frac{\|x - X_i\|}{h_i} \right). \quad (2)$$

It is noted that the partition of unity meshless representation (2) satisfies the important property that it can represent any constant function on Ω exactly.

2.2. Meshless approximation reinterpreted as a sparse DNN

The key idea behind SPINNs is the fact that meshless approximations like (1) and (2) can be *exactly* represented as specially structured sparse DNN. We first discuss how the meshless approximation (1) can be written as a sparse DNN; the corresponding sparse DNN representation for the partition of unity approximation (2) is constructed analogously. The meshless approximation (1) can be thought of as a DNN with an architecture as shown in Fig. 1.

We first transform the input $x \in \Omega$ to the vector $(\|x - X_i\|/h_i)_{i=1}^N$ via hidden layers which we call the *mesh encoding layer*, shown in blue in Fig. 1. In more detail, the mesh encoding layer first transforms the input $x \in \Omega$ to a hidden layer with Nd neurons that have input weights

$$\left(\underbrace{\frac{1}{h_1}, \dots, \frac{1}{h_1}}_{d \text{ terms}}, \underbrace{\frac{1}{h_2}, \dots, \frac{1}{h_2}}_{d \text{ terms}}, \dots, \underbrace{\frac{1}{h_N}, \dots, \frac{1}{h_N}}_{d \text{ terms}} \right),$$

and biases being the Nd vector

$$(-X_1^1, \dots, -X_1^d, -X_2^1, \dots, -X_2^d, \dots, -X_N^1, \dots, -X_N^d),$$

and with the function $\text{sqr} : \mathbb{R} \rightarrow \mathbb{R}$ defined as $\text{sqr}(z) = z^2$ as their activation functions. The output of the first hidden layer of the mesh encoding layer is thus the Nd -vector

$$\left(\frac{(x - X_1^1)^2}{h_1^2}, \dots, \frac{(x - X_1^d)^2}{h_1^2}, \frac{(x - X_2^1)^2}{h_2^2}, \dots, \frac{(x - X_2^d)^2}{h_2^2}, \dots, \frac{(x - X_N^1)^2}{h_N^2}, \dots, \frac{(x - X_N^d)^2}{h_N^2} \right).$$

This is then transformed to another hidden layer consisting of N neurons each of which takes d inputs with weights 1 and has the function $\text{sqrt} : \mathbb{R} \rightarrow \mathbb{R}$ defined as $\text{sqrt}(z) = \sqrt{z}$ as the activation function to produce the N -vector

$$\left(\frac{\|x - X_1\|}{h_1}, \frac{\|x - X_2\|}{h_2}, \dots, \frac{\|x - X_N\|}{h_N} \right).$$

This vector is then passed to the *kernel layer*, shown in brown in Fig. 1, that consists of N neurons with unit input weights and the RBF kernel φ as the activation function. The output of the kernel layer, which is the vector $(\varphi(\|x - X_i\|/h_i))_{i=1}^N$, is then linearly combined using weights (U_i) , which are the coefficients of the meshless approximation (1), to compute the final output $u(x)$ according to the ansatz (1). This demonstrates that the meshless ansatz (1) is exactly representable as a DNN with a special architecture as described above. We wish to highlight two important aspects of this architecture: (i) it is sparse; the number of connections and trainable parameters of this network are much smaller than a DNN with the same number of hidden layers and neurons, and (ii) the trainable coefficients of this network, namely the vectors (h_i) , (X_i) and (U_i) , are interpretable directly in terms of the meshless ansatz (1).

2.3. SPINN architecture

The foregoing discussion naturally motivates the introduction of generalized meshless approximations where the kernel is represented using a DNN. For instance, a partition of unity meshless approximation of the form (2) with the kernel replaced by a DNN is shown in Fig. 2. It can be seen that the mesh encoding layers, shown in blue in Fig. 2, are identical to that described earlier in the context of the DNN equivalent of (1). The primary difference is that instead of using an RBF kernel as the activation function in the kernel layer, a standard DNN with any differentiable activation function, shown in brown in Fig. 2, is used as the kernel; we call this the *kernel network* in the sequel. It is worth pointing out that the *same* kernel network is used for each of the outputs of the mesh encoding layer, in conformity with the meshless ansatz (2). This drastically reduces the number of connections in the SPINN architecture in comparison with a DNN network having a similar architecture. We therefore see that the SPINN architecture proposed here is more efficient than conventional DNN based methods such as Deep Ritz [7], PINN [29], etc. Note also that the number of neurons in the mesh encoding layer is exactly the same as the number of nodes used in the meshless discretization. We thus have physically inspired means to fix the size of the hidden layers in SPINN, unlike other DNN based approaches like PINN where the size of the hidden layers is arbitrary. Further, except for the parameters of the kernel network, the remaining learnable parameters of the network are fully interpretable. In fact, once the SPINN model is trained, it is straightforward to extract the corresponding meshless ansatz. The parameters of the kernel network do not require to be interpreted since the kernel network itself can be thought of as a generalized RBF kernel.

To emphasize the interpretability of SPINN using as an example the architecture shown in Fig. 2, which is an instance of the meshless ansatz (2) with a kernel network in place of the RBF kernel φ . The interpretability of SPINN is to be understood as follows:

- (i) The weights connecting the input layer to the first hidden layer encode the widths (h_i) of the meshless kernels located at each node.
- (ii) The biases of the first hidden layer, on the other hand, encode the coordinates (X_i^1, \dots, X_i^d) of the nodes, which are the centers of the meshless kernels.
- (iii) The kernel network, which in this case is a DNN, stands for the RBF kernel φ that the meshless ansatz uses.
- (iv) Finally, the weights connecting the output of the kernel network to the output layer encode the coefficients (U_i) of the meshless ansatz.

We thus see that the SPINN network is interpretable in the sense outlined above. In particular, given a SPINN model, it is straightforward to extract a meshless approximation using the procedure elaborated above. This provides a principled rationale for choosing the internal architecture of SPINN based on the corresponding meshless representation. This is in sharp contrast to methods like PINN, Deep Ritz, etc., where the internal architecture of the DNNs used is arbitrary. All the non-interpretability of SPINN is thus reduced to the non-interpretability of a *much smaller* kernel network. It bears emphasis nevertheless that the kernel network itself is interpretable as a whole as a basis function.

We also highlight the fact that since the positions of the nodes and the widths of the kernel associated with the nodes are trainable parameters of the network, the learning algorithm implicitly encodes mesh adaptivity as part of the training process. We note in particular that for problems with large gradients, the widths of the kernels naturally develop a multiscale hierarchy during training, as will be demonstrated in the *Results* section.

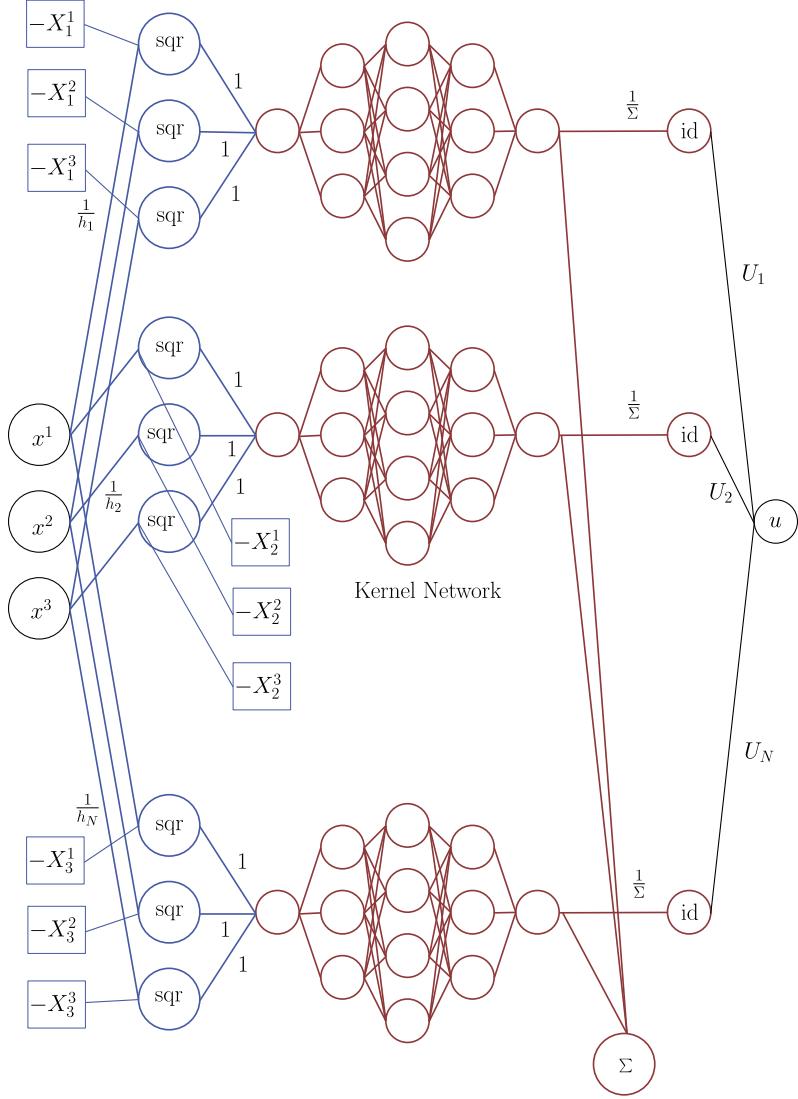


Fig. 2. A detailed view of SPINN with DNN kernel. Note that the kernel networks are exact copies of each other. They are shown separately to emphasize the sparsity of the connection between the mesh encoding and kernel layers.

2.4. Kernel functions

The restriction that the kernel is a radial basis function can be easily removed. It is well known [15] that one-dimensional piecewise linear finite element basis functions can be written exactly in terms of the ReLU basis functions; details are provided in Appendix A. However, ReLU functions are not differentiable, and this can pose problems for their use in ODEs and PDEs of order two or higher. Motivated by the connection between ReLU activation functions and hat functions, we propose a new class of basis functions that are infinitely differentiable. We note first that the softplus function

$$\rho(x) = \log(1 + \exp x), \quad (3)$$

provides a smooth approximation of the ReLU function, as shown in Fig. 3. The softplus function can now be used to create a basis function with *almost* compact support in a manner analogous to the construction of piecewise linear finite element basis functions using ReLU functions. Specifically, we note that the function

$$N(x) = \frac{1}{\rho(1)} \rho(1 + 2 \log 2 - \rho(x) - \rho(-x)) \quad (4)$$

resembles kernels used in meshless approximations; we call this the *softplus hat kernel*. The constants are chosen such that $N(0) = 1$. A graph of the softplus hat function is shown in Fig. 4a.

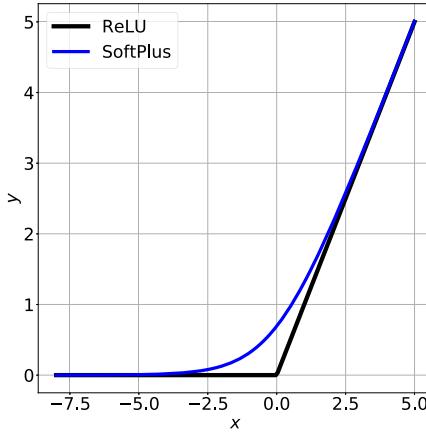


Fig. 3. Comparison of the softplus and ReLU functions.

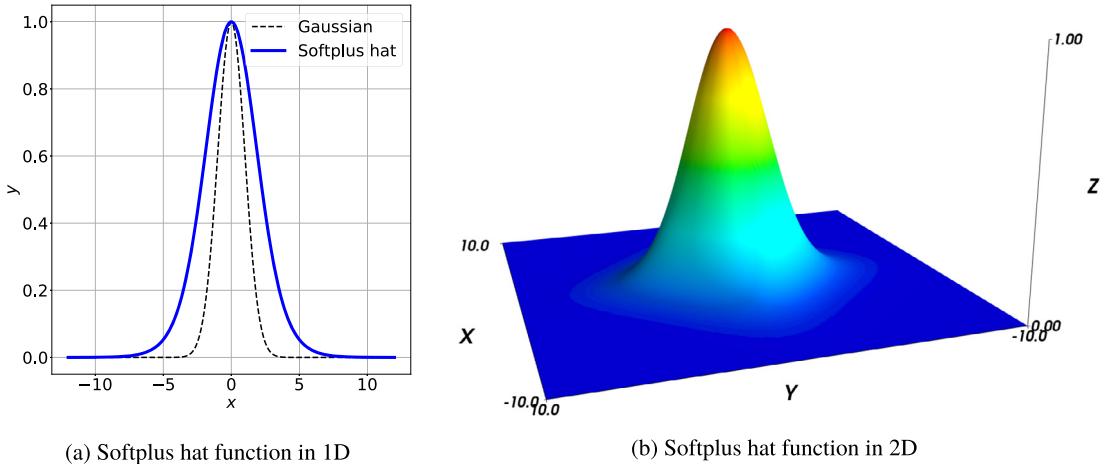


Fig. 4. Softplus based kernel functions in 1 and 2 dimensions.

What makes the softplus hat kernel interesting is the fact that it can be exactly represented as a two layer neural network. The input x first feeds into a hidden layer with two neurons, with weights $(1, -1)$, bias 0 and softplus activation function. The output of this hidden layer $(\rho(x), \rho(-x))$ is then linearly combined using a second hidden layer with one neuron with weights $(-1, -1)$, bias $1 + 2 \log 2$ and softplus activation function. The output of this layer is therefore $\rho(1 + 2 \log 2 - \rho(x) - \rho(-x))$. Finally, it is straightforward to scale this by a factor of $1/\rho(1)$ to get the final output $N(x)$. We thus see that the softplus hat function is indeed transcribable exactly as a neural network.

It is straightforward to generalize this to higher dimensions. For instance, a d dimensional softplus hat kernel is given by

$$N(x^1, \dots, x^d) = \frac{1}{\rho(1)} \rho \left(1 + 2d \log 2 - \sum_{k=1}^d (\rho(x^i) + \rho(-x^i)) \right). \quad (5)$$

We emphasize that the higher dimensional softplus hat kernel functions (5) are also representable directly as a two layer neural network; this is illustrated for two dimensional softplus hat kernels in Fig. 5. A graph of softplus hat kernel in two dimensions is shown in Fig. 4b. Higher dimensional softplus hat functions are also exactly represented by an equivalent neural network.

Though the softplus hat kernels do not have compact support, they approach zero quickly outside a small neighborhood of the central node, just like Gaussian kernels. Thus they are expected to have the same performance as Gaussian RBF kernels. This is indeed borne out by the numerical experiments, as will be demonstrated shortly.

We wish to point that the softplus hat kernel is new and has not been used before to the best of our knowledge. With the choice of these softplus hat kernel functions, meshless approximations that go beyond radial basis functions are constructed along the same lines mentioned before. We present many examples using the softplus hat kernel in the Results section.

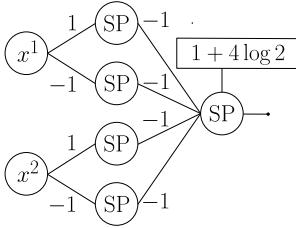


Fig. 5. Softplus hat kernel represented as a neural network.

2.5. Loss definition

With this choice of architecture, solving the PDE, $\mathcal{N}(x, u, \nabla u, \dots) = 0$, is easily accomplished using a collocation technique similar to the one used in PINNs [29]. We also note in passing that for PDEs that are obtained as the Euler-Lagrange equations of a known functional, the loss function can be formulated using quadratures of the integral of the functional along with penalty terms to enforce boundary conditions, as is carried out in the Deep Ritz method [7].

To illustrate the loss functions used in training the SPINN models, consider the special case of a second order PDE of the form

$\mathcal{N}(x, u(x), \nabla u(x), \nabla^2 u(x)) = 0, \quad x \in \Omega,$
 $u = u_0 \text{ on } \partial_0 \Omega, \quad \nabla u \cdot n = g_0 \text{ on } \partial \Omega \setminus \partial_0 \Omega,$

n is the outward unit normal to $\partial \Omega$. The generalization to
 defining the SPINN network for this problem is chosen as

$$L((h_i), (X_i), (U_i)) = w_i \sum_{i=1}^{M_i} \mathcal{N}(\xi_i, u(\xi_i), \nabla u(\xi_i), \nabla^2 u(\xi_i)) + w_d \sum_{i=1}^{M_d} (u(\eta_i) - u_0(\eta_i))^2 + w_n \sum_{i=1}^{M_n} (\nabla u(\xi_i) \cdot n(\xi_i) - g_0(\xi_i))^2.$$

Here w_i , w_d and w_n are constants that enforce the loss in the interior of the domain, the Dirichlet boundary, and the Neumann boundary, respectively. The set of points $(\xi_i)_{i=1}^{M_i}$, $(\eta_i)_{i=1}^{M_d}$ and $(\zeta_i)_{i=1}^{M_n}$ in the interior of Ω , and on the Dirichlet and Neumann boundaries on Ω , respectively, are the sampling points where the loss is evaluated. The sampling points are chosen to be uniformly distributed points inside the domain and on its boundary. For regular domains, we choose a fine uniform grid and identify the vertices of the grid as the interior and boundary sampling points. For irregular domains, we generate a fine triangulation and choose the vertices of the triangulation as the sampling points. Note that the precise mode of generation of the sampling points is not as important as the fact that every region of the domain is sufficiently represented by the sampling points. We choose a uniformly spaced distribution of the sampling points in the examples, but more general distributions that are tailor made for specific problems can be easily incorporated in the current framework.

In cases where a variational form

$$I(u) = \int_{\Omega} f(x, u(x), \nabla u(x)) dx \quad (7)$$

is available for the PDE, the interior loss can alternatively be defined directly in terms of an appropriate quadrature approximation of the integral in (7). Dirichlet boundary conditions are imposed using penalty terms as in the strong form collocation case described earlier. Neumann boundary conditions, on the other hand, are directly integrated into the definition of the integral loss functional.

2.6. Time dependent PDEs

We present two different approaches for time dependent PDEs using SPINN. The first employs a space-time grid of dimension $d + 1$ and uses exactly the same ideas presented above to solve a time dependent PDE. This is similar to the spacetime DNN used in works like [29]. Alternatively, a hybrid finite difference and SPINN method, which we call FD-SPINN, can be employed which performs time marching using conventional finite difference methods and performs spatial

discretization at each time step using the SPINN architecture. It is worth mentioning that both explicit and implicit finite difference schemes are subsumed in FD-SPINN. This can also be generalized to multi-step methods as is done in works like [30]. Both the space-time and the FD-SPINN methods will be illustrated later on.

We illustrate an implicit FD-SPINN algorithm in the special case of the 1D heat equation $u_t = u_{xx}$; a similar formalism applies for other time dependent PDEs. Choosing a time step Δt , we denote by $(u^k(x))_{k=1}^{N_t}$ the approximations to the solution $u(x, k\Delta t)$; here N_t is chosen such that $N_t \Delta t \simeq T$. The following first order implicit time difference scheme is used in this work:

$$\frac{u^{n+1}(x) - u^n(x)}{\Delta t} = \frac{d^2 u^{n+1}(x)}{dx^2}, \quad n = 1, \dots, N_t.$$

It is to be noted that the parameters of the network are *not* re-initialized at each time step. Rather, the converged parameters from the previous time step are used as the initial values of the parameters for the current time step. We wish to emphasize that the spatial derivatives are computed *exactly* using automatic differentiation since the spatial approximation uses SPINN. Thus this implicit scheme is different in comparison to traditional time marching schemes. The loss in the interior of the domain $[0, L]$ is computed as the squared residue of the foregoing equation. The second algorithm uses SPINN for both space and time discretization. The implementation of the space-time SPINN solution is similar to the implementation of second order PDEs described earlier.

2.7. Fourier-SPINN

Another advantage of the SPINN architecture is that it suggests natural generalizations of familiar decompositions of functions. To make this precise, consider the Fourier expansion of a function $u : [a, b] \rightarrow \mathbb{R}$, namely

$$u(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos k\omega x + \sum_{k=1}^{\infty} b_k \sin k\omega x,$$

where $\omega = 2\pi/(b-a)$. It is straightforward to reinterpret this as a neural network with one hidden layer that has the sinusoidal functions as the activation functions of the neurons. The input x is transformed using a hidden layer with $2N$ neurons to the scaled inputs

$$(x, 2x, \dots, Nx, x, 2x, \dots, Nx) \in \mathbb{R}^{2N}.$$

The activation functions of the $2N$ neurons in the hidden layer are chosen as

$$\underbrace{(\cos \omega z, \cos \omega z, \dots, \cos \omega z)}_{N \text{ terms}}, \underbrace{(\sin \omega z, \sin \omega z, \dots, \sin \omega z)}_{N \text{ terms}}.$$

The biases of the hidden layer are uniformly set to zero. The output of this hidden layer is then linearly combined to produce the output

$$u(x) = U_0 + \sum_{k=1}^N U_k \cos k\omega x + \sum_{k=1}^N V_k \sin k\omega x,$$

which is just the Fourier representation of u . The Fourier representation is then used in conjunction with an appropriately defined loss function to solve a given differential equation for u . It bears emphasis that the weights and biases of this neural network are fully interpretable as in the case of the meshless approximation discussed earlier.

We wish to point out that the specific version of Fourier-SPINN that we propose here is different from other Fourier based methods like [41,26], where the authors extract Fourier features of the inputs, which are then fed as inputs to a DNN. In contrast, the method we propose here directly transcribes a Fourier decomposition as a neural network architecture, thereby suggesting many generalizations. It is possible however to combine both these approaches and study hybrid variants; we do not pursue those directions in this paper.

We also note in passing that wavelet transforms, see for instance [43], could also be represented using SPINN. As a natural generalization one could replace the sinusoidal functions by a DNN thereby providing a neural network generalization of the Fourier transform. Investigations of such extensions are planned for future work. We remark that the examples studied using SPINN presented later already show multiresolution capabilities in the specific manner in which it adapts the nodes and kernel widths.

2.8. Details of implementation

The SPINN architecture proposed in this work is easily implementable using standard NN libraries. We provide PyTorch [28] implementations of all the examples considered in this work at <https://github.com/nn4pde/SPINN>.

We present here certain details of the implementation. We classify the nodes associated with the SPINN model as fixed and free nodes. Fixed nodes are typically used on the (Dirichlet) boundaries. Free nodes, on the other hand, are used inside the domain of definition of the problem. Both fixed and free nodes are designed to have variable kernel widths, but the free nodes are also free to move both inside and outside the domain. A separate set of sampling points on the interior and boundary of the domain are also used. These are the points where the interior and boundary losses are evaluated using the SPINN model. We provide options for both full sampling and random sampling. In random sampling, a random subset of an initially generated set of sampling points is chosen for each iteration of the loss minimization algorithm. It is to be noted that what we call as random sampling is the same as mini-batching in the machine learning literature. The weights and biases of both the hidden and output layers are set to zero by default. This implies in particular that when using full sampling and either the Gaussian or softplus hat kernels, the SPINN algorithm is fully deterministic. Thus there are only two sources of stochasticity in the current implementation of SPINN: (i) randomness due to initialization of the DNN kernel weights and biases when DNN kernels are used, and (ii) randomness due to sampling of the interior sampling points when random sampling is used. We do not sample on the boundary and use full boundary sampling always. This is because of the fact that boundary conditions are more delicate to impose in the current framework. The optimization of the SPINN models is carried out using any of the well known optimizers. All the examples presented in this work were carried out using the Adam optimizer [18] implemented in PyTorch.

3. Results

We now present solutions of a variety of ordinary and partial differential equations using SPINN. We implement the SPINN architecture using PyTorch [28], and the code is available at <https://github.com/nn4pde/SPINN>. All our results are automated using `automan` [32] for easy reproducibility. We have provided details of various parameters used to arrive at the solutions. However, we wish to point out that our entire code including all the examples shown here is available at the aforementioned Github repository, a perusal of the `automate.py` file should provide full details of the various parameters used. All our simulations are thus *fully reproducible*. We also point out that the error values for *each of the simulations* across training steps are automatically recorded and archived in the `outputs/` folder when the codes are run using the procedure outlined in the repository. Unless stated otherwise, all simulations use full sampling. When random sampling is used, the relevant details are specified explicitly.

3.1. Ordinary differential equations

To validate the SPINN method, we first consider ordinary differential equations (ODEs) with different boundary conditions. We study a few ODEs primarily to illustrate various aspects of the SPINN algorithm in a simpler setting before moving on to PDEs. Except for the cases involving variational SPINN, strong form collocation is used to compute the loss function in all examples shown here.

3.1.1. ODE with Dirichlet boundary conditions

To begin with we consider the ODE

$$\frac{d^2u(x)}{dx^2} + 1 = 0, \quad x \in (0, 1), \\ u(0) = u(1) = 0. \quad (8)$$

The exact solution for the ODE (8) is easily computed as

$$u(x) = \frac{1}{2}x(1 - x). \quad (9)$$

The solution computed using SPINN with Gaussian kernel, $n = 1, 3, 7$ uniformly placed internal nodes, two fixed boundary nodes on the left and right boundaries $x = 0$ and $x = 1$ respectively are shown in Fig. 6. The nodal positions learnt by the SPINN algorithm are also shown therein. For all these simulations $n_S = 6n$ sampling points were used to compute the interior loss, while two boundary nodes, one at each boundary, were used to compute the boundary loss. Both the nodes and sampling points were initially uniformly distributed over the domain. As can be seen from Fig. 6 the nodes cluster towards regions where the solution gradient is higher. A learning rate of 10^{-4} was used in all these simulations. For the case when $n = 3$, we present the solutions obtained by using the Gaussian, Softplus and NN kernels in Fig. 7. The kernel network was chosen as a fully connected NN with 2 hidden layers with 5 neurons each. We note that this is a much smaller network than what is typically employed in PINNs. The parameters of the kernel network are initialized randomly, unlike the other two kernels where all parameters are set to zero initially. Curiously, while both the Gaussian and softplus hat kernels show symmetric node placements, the node distribution for the case when the kernel is chosen as a neural network is asymmetric. To understand this better we show in Fig. 8 the kernel function that SPINN learns. It is seen that the kernel learnt by SPINN in this case is actually asymmetric, unlike the Gaussian and softplus hat kernels. This causes a corresponding asymmetry in the nodal positions observed in Fig. 7c.

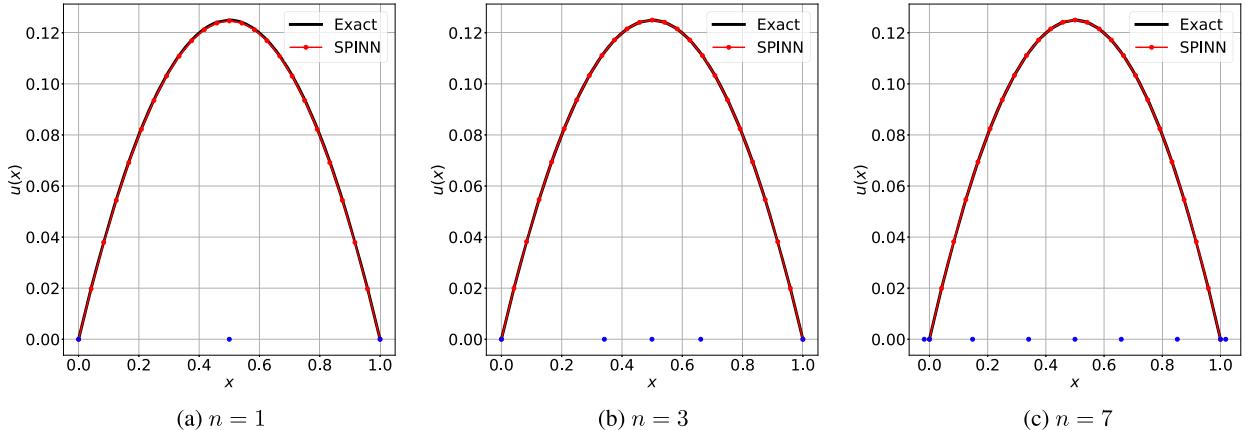


Fig. 6. Solution of the ODE (8) using SPINN with Gaussian kernel using 1, 3 and 7 interior nodes. The nodal positions learnt by SPINN are shown as blue circles along the x axis. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

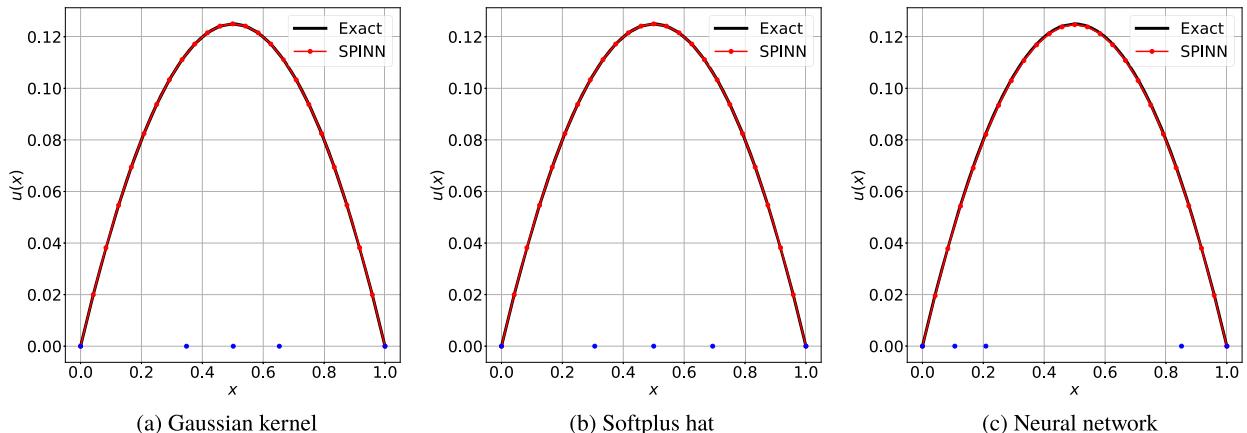


Fig. 7. Solution of the ODE (8) using SPINN with different kernels. All simulations use $n = 3$ interior nodes. The nodal positions learnt by SPINN are shown as blue circles along the x axis. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

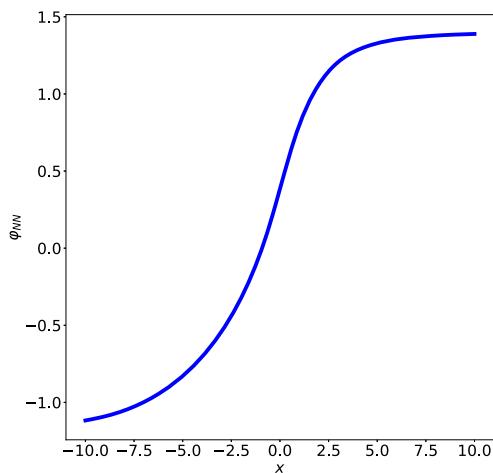


Fig. 8. Basis learnt by SPINN with neural network kernel. Note that the basis is asymmetric. This is reflected in the asymmetric placement of the nodes in Fig. 7c.

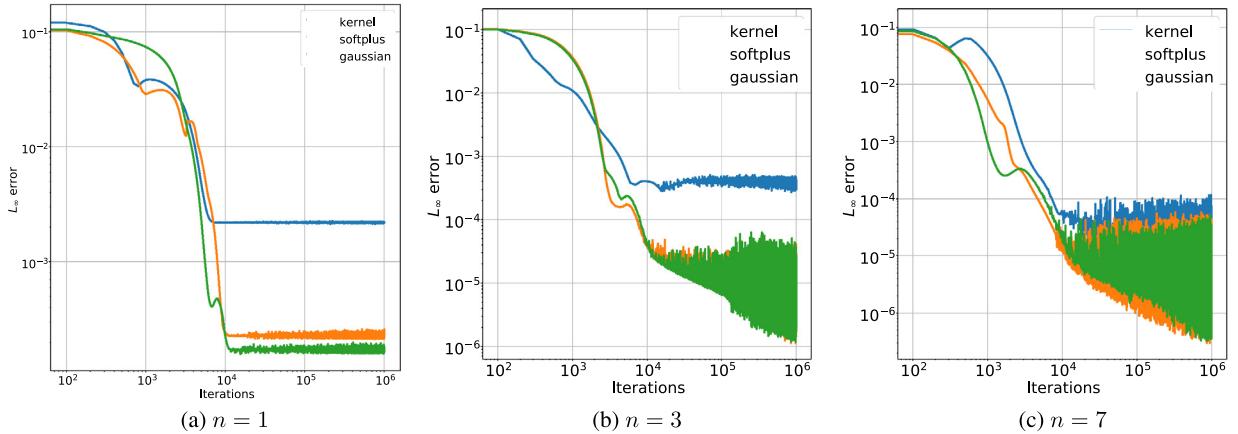


Fig. 9. Evolution of L_∞ error during training for ODE (8) using SPINN with different kernels for $n = 1, 3, 7$ nodes. The simulations were performed for 10^6 training steps to study the long term error behavior.

We discuss the accuracy of the SPINN algorithm for this problem next. The evolution of the L_∞ error during the training of the SPINN model with $n = 3$ internal nodes and different kernels is shown in Fig. 9. All simulations shown in this figure were carried out for 10^5 training steps to study the long term behavior of the L_∞ errors in each of the cases. We find that the Gaussian kernel has the best performance in general, with the softplus hat kernel having a similar performance. The neural network kernel, however, does not perform well for smaller number of internal nodes. We note that the errors for the $n = 3$ and $n = 7$ cases dip below 10^{-4} , which is satisfactory considering the fact that these simulations employ single precision floating point numbers.

For all the kernels, it is observed that the error plot shows three distinct regimes. Focusing on Fig. 9a, for instance, an examination of the intermediate solutions during the iteration reveals that the first phase characterized by high loss and slow convergence correlates with the SPINN model minimizing the interior loss. The second phase in Fig. 6a characterized by a rapid drop in the error correlates with the SPINN model learning the boundary conditions of the ODE. The final phase is characterized by the error saturating and exhibiting oscillations due to the inability of the algorithm to perform better given the various aspects of the discretization.

3.1.2. ODE with large gradients

We consider next the ODE

$$\begin{aligned} u''(x) + x(\exp(-(x - (1/3))^2/K) - \exp(-4/9K)) &= 0, \quad x \in (0, 1), \\ u(0) = u(1) &= 0. \end{aligned} \tag{10}$$

In ODE (10), we choose $K = 0.01$. The exact solution for the ODE (10) is

$$u(x) = -\frac{1}{K^2} \left(K \left(\frac{4}{3} - 6x \right) + 4x \left(x - \frac{1}{3} \right)^2 \right) \exp \left(-\frac{1}{K} \left(x - \frac{1}{3} \right)^2 \right). \tag{11}$$

The exact solution (11) informs us that the solution has sharp gradients, with the gradient being sharper the lower the value of K . We chose this example to study the ability of SPINN to capture large gradients in a simplified setting. The solution obtained with SPINN using $n = 1, 3, 7$ nodes is shown in Fig. 10. For all the simulations, $n_S = 20n$ interior sampling points and two boundary sampling points are used. As in the previous case, both the nodes and sampling points are distributed uniformly initially. A learning rate of 10^{-4} is used and the network is trained for 10^5 iterations. The solution corresponding to the choice of different kernels with $n = 3$ interior nodes, along with the evolution of the L_∞ errors is shown in Fig. 11. As in the case of the previous example, the clustering of the nodes near high gradient regions of the solution and the three phases of the error plots are clearly seen.

3.1.3. Treatment of Neumann boundary conditions

We study next the following ODE with both Dirichlet and Neumann boundary conditions:

$$\begin{aligned} u''(x) + \pi^2 u(x) &= \pi \sin \pi x, \quad x \in (0, 1), \\ u(0) = 0, \quad u'(1) &= \frac{1}{2}. \end{aligned} \tag{12}$$

The exact solution to ODE (12) is

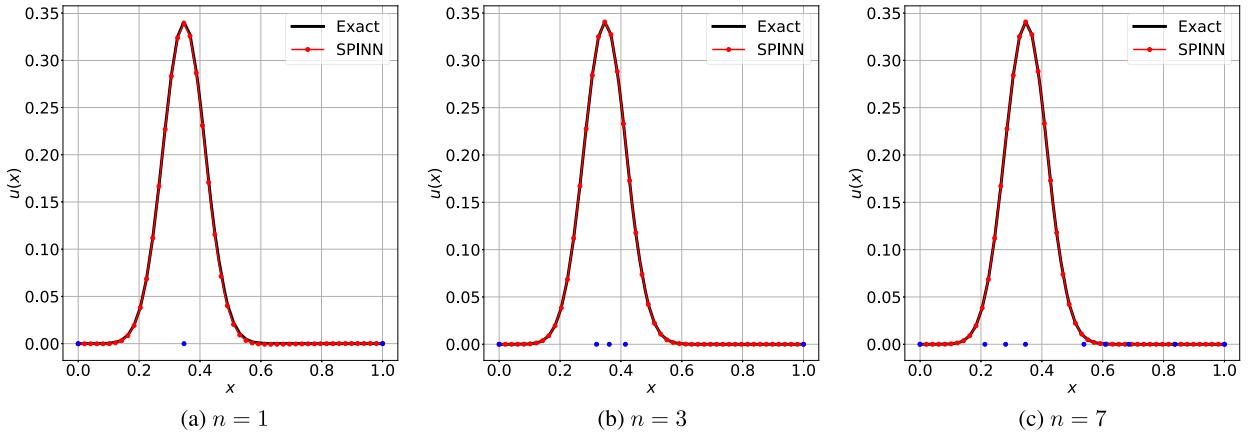


Fig. 10. Solution of ODE (10) using SPINN with Gaussian kernel and $n = 1, 3, 7$ interior nodes. The nodal positions learnt by SPINN are shown as blue circles along the x axis. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

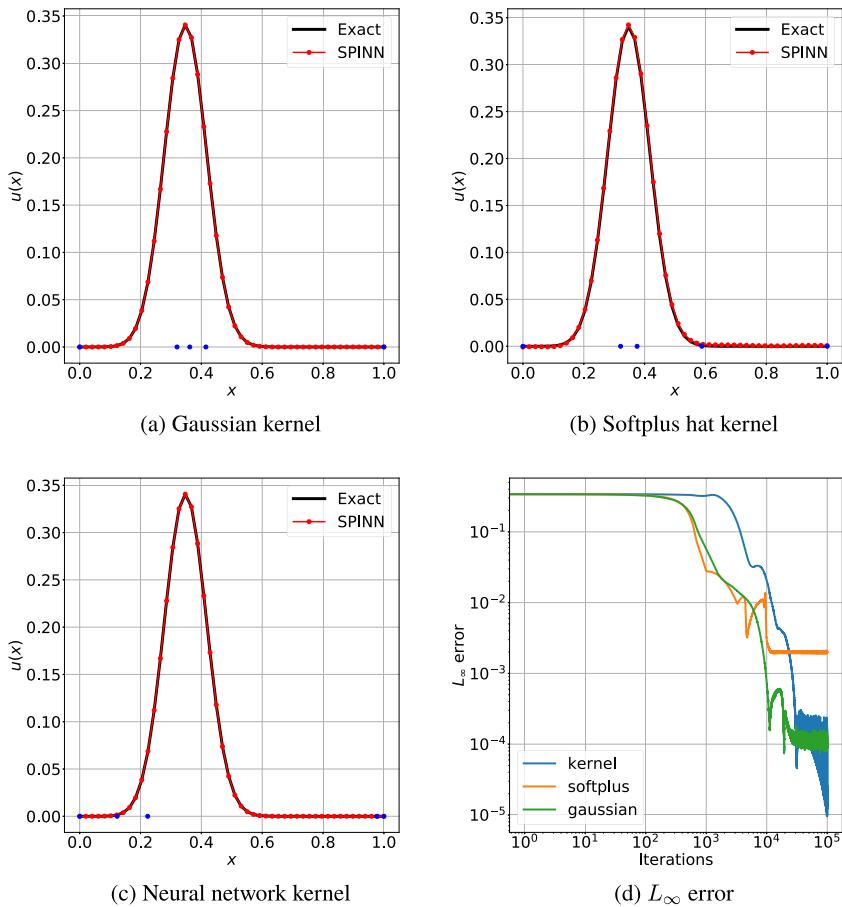


Fig. 11. Solution of ODE (10) with different kernels. $n = 3$ internal nodes are used for all these simulations. The position of the nodes learnt by the SPINN algorithm is also shown as filled blue circles on the x axis. The evolution of the L_∞ error for all three choices of kernels is also shown. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$$u(x) = -\frac{1}{2}x \cos \pi x. \quad (13)$$

The solution computed by SPINN using Gaussian kernel and $n = 3, 5, 7$ internal nodes is shown in Fig. 12. In each case, $n_S = 15n$ internal sampling points and two boundary sampling points were used. Both the nodes and samples were distributed uniformly initially. To handle the Neumann boundary condition at $x = 1$, no fixed node is used there. The reason for this is

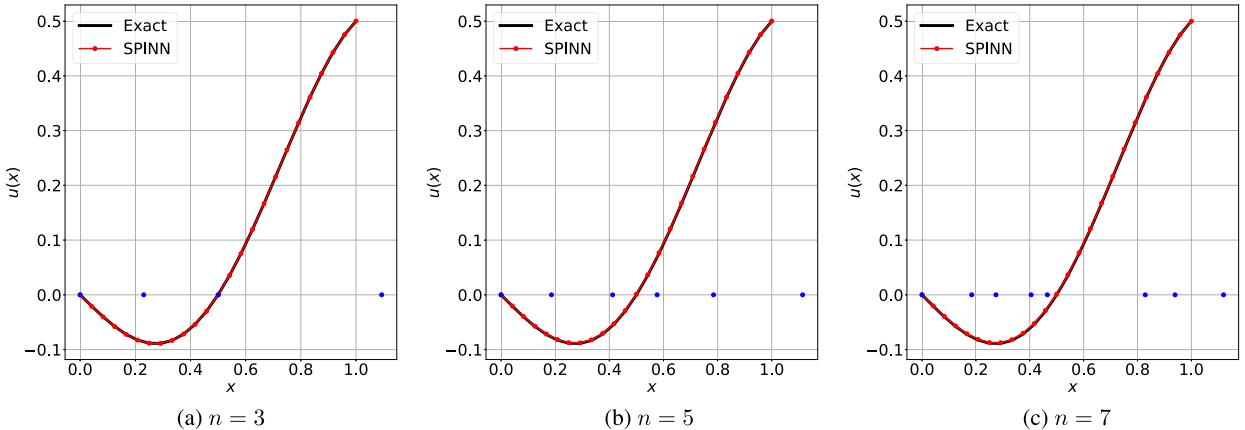


Fig. 12. Solution of ODE (12) using SPINN with Gaussian kernel and $n = 3, 5, 7$, interior nodes. No fixed node is placed on the Neumann boundary. The nodal positions learnt by SPINN are shown as blue circles along the x axis. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

that the kernels employed in this work are symmetric in nature and hence have vanishing gradients at their centers. This in turn leads to an infinite indeterminacy if a fixed node is placed on a Neumann boundary. The interior nodes are free to move outside the domain to accommodate the Neumann condition. This is indeed seen in the SPINN solution shown in Fig. 12. The solutions shown in Fig. 12 were computed using random sampling with sampling ratio $f = 0.2$. The learning rate was chosen as 10^{-3} . Unlike the previous simulations which were run for a fixed number of training steps, the simulations shown in Fig. 12 were terminated once the L_∞ error reached a threshold value of 10^{-3} , or when the number of iterations exceeded 10^6 . As can be seen from Fig. 12, the SPINN algorithm is able to handle Neumann boundary conditions effectively.

We present the effect of random sampling on the performance of SPINN in Fig. 13. The corresponding errors for the case of $n = 5$ interior nodes, with all other parameters fixed as mentioned above, are shown in Fig. 14. We also illustrate the effect of sampling ratio f on the convergence of the SPINN algorithm in Fig. 14a. The L_∞ error plot is shown in Fig. 14a. We note that since the algorithm is non-deterministic on account of the use of random sampling, there is a finite probability of obtaining a solution that gets stuck in a metastable state. To illustrate this we performed five different runs corresponding to a sampling ratio of $f = 0.2$ - the results are shown in Fig. 14b. It can be seen that while one of the initial conditions of the five trials gets stuck in a metastable state, the remaining four converge to the true solution. Thus the algorithm is sensitive to a small extent on initial conditions when random sampling is used.

3.1.4. Variational SPINN

The examples shown so far used collocation on the strong form of the differential equation for the loss computation. SPINNs can also be designed with variational principles directly if they are available. As a proof of concept, we recall from basic variational calculus that the Euler-Lagrange equation of the functional

$$I(u) = \int_0^1 \frac{1}{2} \left(\frac{du(x)}{dx} \right)^2 dx - \int_0^1 f(x)u(x) dx,$$

where $u : [0, 1] \rightarrow \mathbb{R}$ with suitable regularity and such that $u(0) = u(1) = 0$, is precisely the differential equation

$$\frac{d^2 u(x)}{dx^2} + f(x) = 0,$$

defined over the domain $(0, 1)$ with zero Dirichlet boundary conditions. The functional associated with the differential equation is chosen as the loss function. Since the functional is defined via integrals, an appropriate choice of quadrature is required. For the examples shown here a simple Riemann sum over a uniform partition of the interval $[0, 1]$ is used to evaluate the integral. Basic examples are shown in Fig. 15. We note that the variational formulation of SPINN is sensitive to the order of quadrature used. A coarse discretization results in a loss of stability for the true solution. The correct solution is however captured when the discretization is fine enough.

3.1.5. Fourier-SPINN

We also use this opportunity to illustrate Fourier-SPINN. Preliminary results showing the solution of ODE (10) using Fourier SPINN with strong form collocation and Gaussian kernel are shown in Fig. 16. The solution obtained required the use of $m = 50$ modes. The learning rate was set to 10^{-4} . The corresponding L_∞ error plot is also shown in Fig. 16. We note that using fewer modes results in oscillatory solutions, as expected from a Fourier fit. We remark that the Fourier version of SPINN is useful for problems with periodic boundary conditions.

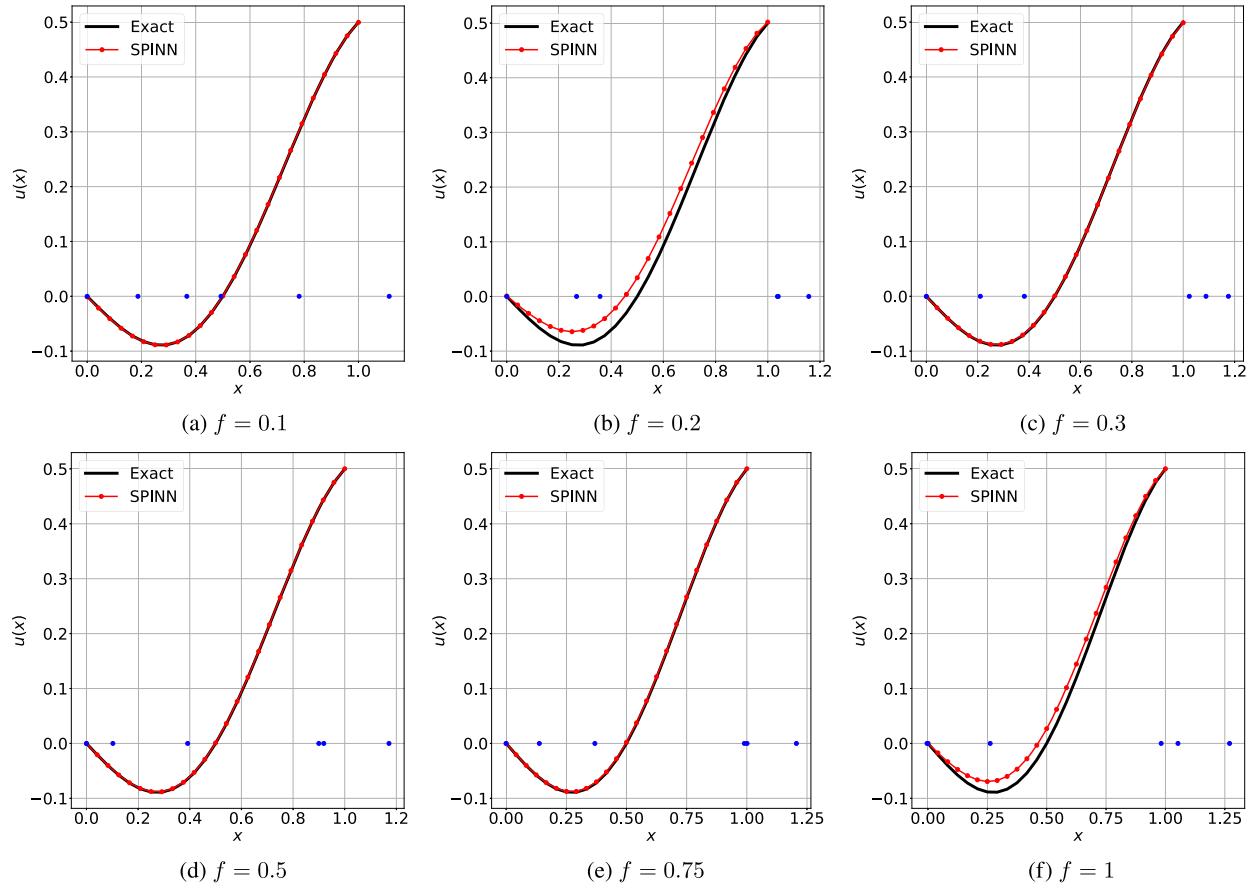


Fig. 13. Effect of random sampling on the SPINN solution to ODE (12) using Gaussian activation function. The number of nodes is fixed at $n = 5$ and the total number of samples is fixed at $n_s = 15n$. The various graphs shown here correspond to different fractions, f , of the samples used to evaluate the loss at each iteration.

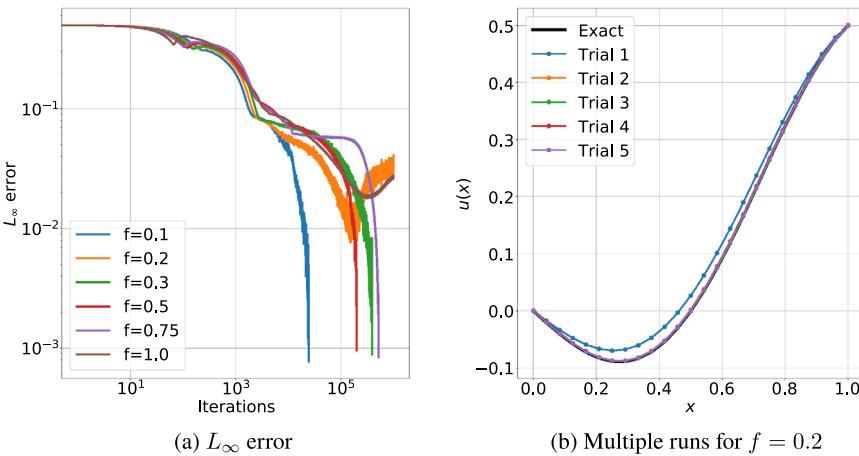


Fig. 14. L_∞ error for SPINN solution of ODE (12) for various sampling ratios. The sampling ratio f is varied, while the number of nodes $n = 5$ and number of samples $n_s = 15n$ were fixed. The simulations were terminated when the errors reached a threshold value of 10^{-3} , or if the number of iterations exceeded 10^6 . The figure on the right shows the results of multiple runs for the same set of parameters, with sampling ratio fixed at $f = 0.2$.

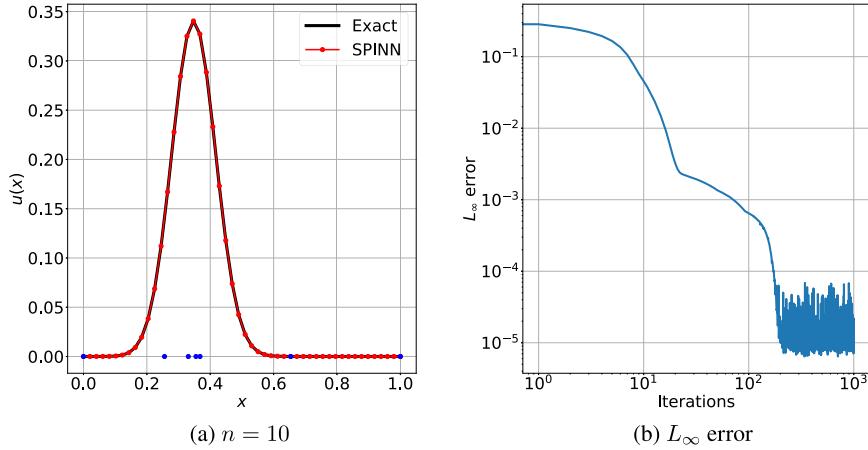


Fig. 15. Variational SPINN for ODE (10). The solution obtained by SPINN along with the nodal positions are shown on the left. The evolution of the L_∞ error during training is shown on the right.

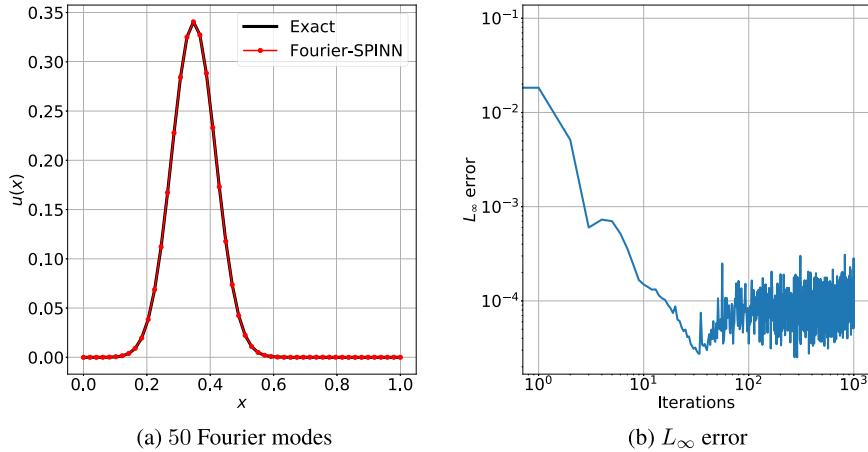


Fig. 16. Fourier-SPINN solution of ODE (10) with 50 modes, along with the L_∞ error during training.

3.1.6. ODE with highly oscillatory solution

As a final example involving ODEs, we now consider the following ODE:

$$\frac{d^2u(x)}{dx^2} + 4\pi^2 \sin(2\pi x) + 250\pi^2 \sin(50\pi x), \quad x \in (0, 1), \quad (14)$$

$$u(0) = u(1) = 0.$$

The exact solution to ODE (14) is easily computed as

$$u(x) = \sin(2\pi x) + 0.1 \sin(50\pi x). \quad (15)$$

This ODE was studied using PINN in [41] as an example of an ODE with multiscale features. This is an example of a problem where PINN performs poorly. In [41], the authors provide a solution of this problem using a PINN with a 5 layer fully connected network with 200 neurons in each layer and report poor convergence and an inability of PINN to learn the correct solution even after 10^7 iterations. In sharp contrast, we illustrate below the superior performance of SPINN. We present SPINN solutions of (14) using Gaussian kernel, a small neural network kernel and using Fourier-SPINN. In all three cases, we find that the SPINN algorithm is not only accurate but also converges to the correct solution with much smaller number of iterations.

The SPINN solution ODE (14) using Gaussian, softplus hat and neural network kernels is shown in Fig. 17. The corresponding L_∞ error is also shown in Fig. 17. For the Gaussian and softplus hat kernels, $n = 100$ internal nodes, $n_S = 800$ sampling points and a learning rate of 5×10^{-4} was used. For the neural network kernel, a small fully connected two layer networks with 5 neurons each was used. For all these simulations, the iterations were terminated when the L_∞ error dipped below 10^{-3} . It is worthwhile noting that the SPINN algorithm with different choices of kernels is able to satisfactorily capture the exact solution in $O(10^4)$ iterations while the standard PINN algorithm as reported in [41] fails to converge

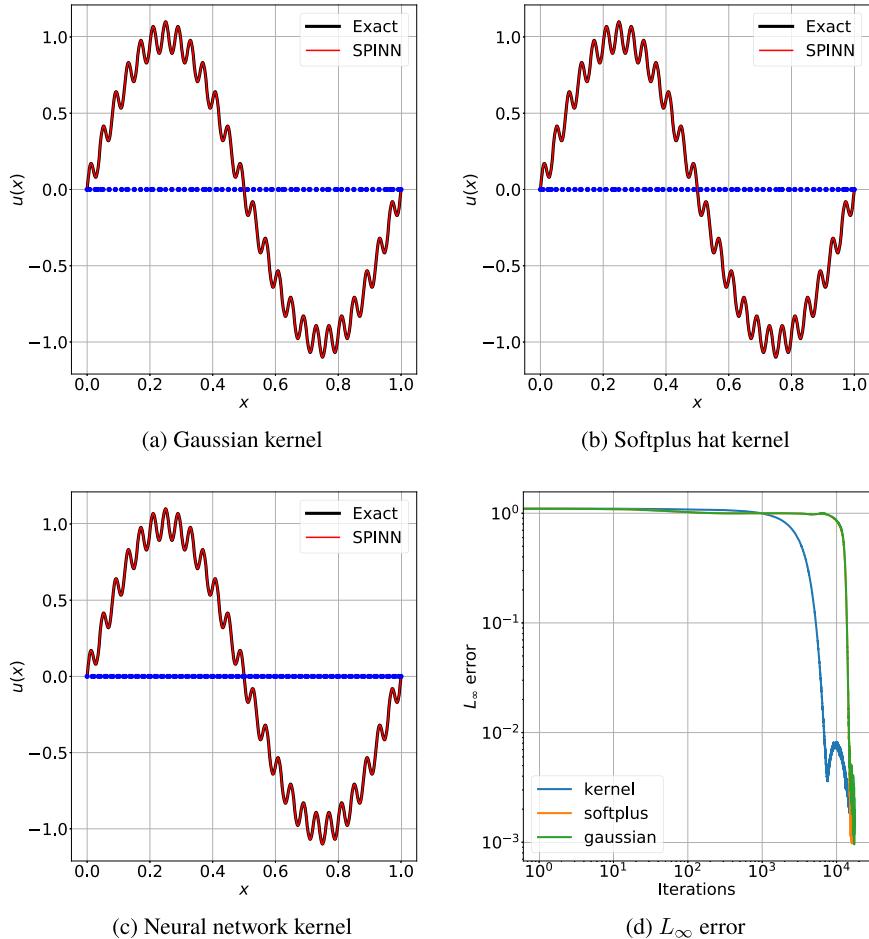


Fig. 17. Solution of ODE (14) using Gaussian, softplus hat and neural network kernels. The nodal positions learnt by the corresponding SPINN models are also shown. The figure on the bottom right shows the L_∞ error for these simulations. The simulations were terminated when the L_∞ error dipped below 10^{-3} .

even after 10^7 iterations. For completeness, we also show the solution of (14) computed using Fourier-SPINN with 200 Fourier modes, and remaining parameters as before, in Fig. 18. It is clearly seen that the Fourier-SPINN model too captures the exact solution within an error of approximately 10^{-3} in $O(10^2)$ iterations. This example thus highlights the strength of SPINN in comparison to PINN.

We plot the solution obtained using PINN with 5 hidden layers and 200 neurons in each layer as mentioned in [41] and compare the solution obtained with SPINN. The same parameters for SPINN as used above are used here. We iterate for a total of 20000 iterations for both methods. For the PINN implementation, we use 2000 sampling points with sampling ratio 0.25, and a learning rate of 5×10^{-4} . The results are shown in Fig. 19. The L_∞ errors as a function of iterations are also shown. As can be clearly seen in the figure, and as reported in [41], the PINN solution fails to converge in sharp contrast to the SPINN solution which not only converges but is also more efficient. The SPINN solution takes around 216 seconds to execute on a CPU whereas the PINN solution takes 1047 seconds. This shows that for this problem SPINN is around 5 times as fast and much more accurate.

3.2. PDEs in two dimensions

We now present a few examples solving PDEs in two dimensions. We focus in particular on linear elliptic PDEs, and discuss the solution of the Poisson equation in both regular and irregular domains.

3.2.1. Poisson equation

As a basic illustration of SPINN's capability to solve PDEs, we consider the Poisson equation

$$\nabla^2 u(x, y) = 20\pi^2 \sin 2\pi x \sin 4\pi y, \quad x, y \in D = (0, 1) \times (0, 1), \\ u(x, y) = 0, \quad x, y \in \partial D. \quad (16)$$

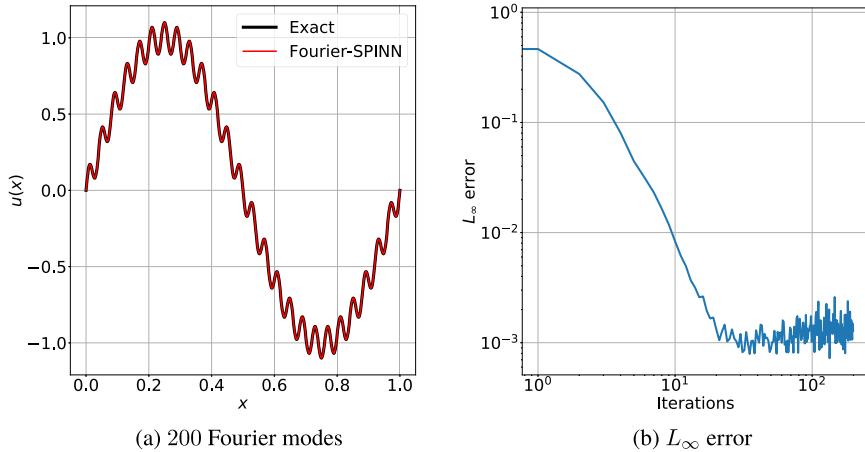


Fig. 18. Solution of ODE (14) using Fourier-SPINN with 200 Fourier modes. The corresponding L_∞ plot is also shown.

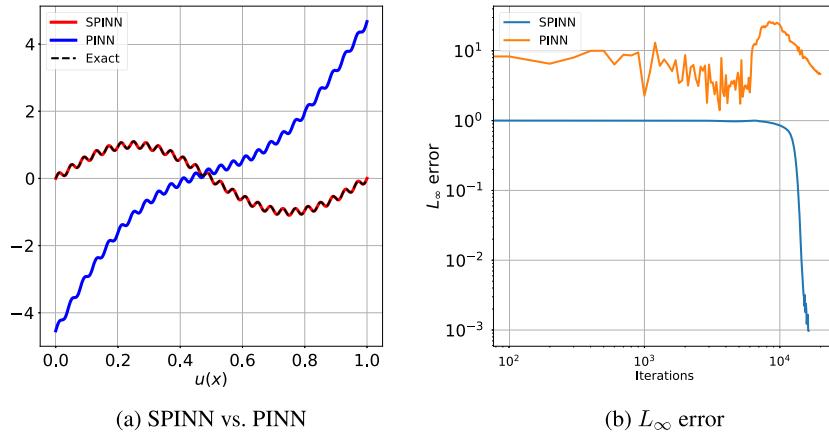


Fig. 19. Solution of ODE (14) using SPINN compared with that obtained using PINN. The corresponding L_∞ plot is also shown.

The exact solution to the PDE (16) is given by

$$u(x, y) = \sin 2\pi x \sin 4\pi y. \quad (17)$$

We first present the convergence of SPINN with different activation functions with around 100 internal points, 40 fixed boundary nodes, 400 internal sampling points and 80 boundary sampling points, in Fig. 20. The learning rate was chosen as 10^{-3} and a total of 5×10^4 iterations. The L_∞ error as a function of the iteration number is plotted for different kernels in Fig. 21a. As before, we observe two distinct regimes in the error graph corresponding to learning the interior and learning the boundary conditions. While all three kernels provide a good solution, the softplus hat kernel performs better than the other two for this PDE. The convergence of the SPINN solution as a function of the number of interior nodes used is shown in Fig. 21b; we observe that the error decreases with increase in the number of nodes, as expected.

3.2.2. Square slit problem

As a second example we consider the Poisson equation on a square domain with a slit:

$$\begin{aligned} \nabla^2 u(x, y) + 1 &= 0, \quad D = (-1, 1) \times (-1, 1) \setminus [0, 1], \\ u(x, y) &= 0, \quad x, y \in \partial D. \end{aligned} \quad (18)$$

This PDE does not have an exact solution, but it has known asymptotic properties around the origin [7]. The solution obtained using SPINN is shown in Fig. 22a. For this simulation, around 200 interior nodes, 70 fixed boundary nodes, 600 interior sampling points and 110 boundary sampling points were used. The learning rate was chosen as 10^{-3} . The computations were performed for 10^4 iterations. The corresponding nodal positions are shown in Fig. 22b. It is seen that the SPINN algorithm learns the optimal position of the nodes and the size of the kernels at the nodal positions appropriately. In more detail, we expect the initially uniform nodes to rearrange themselves so as to better reflect the local geometry of the domain, which in this case is most notable near the slit. Since the solution varies rapidly over a small spatial distance close

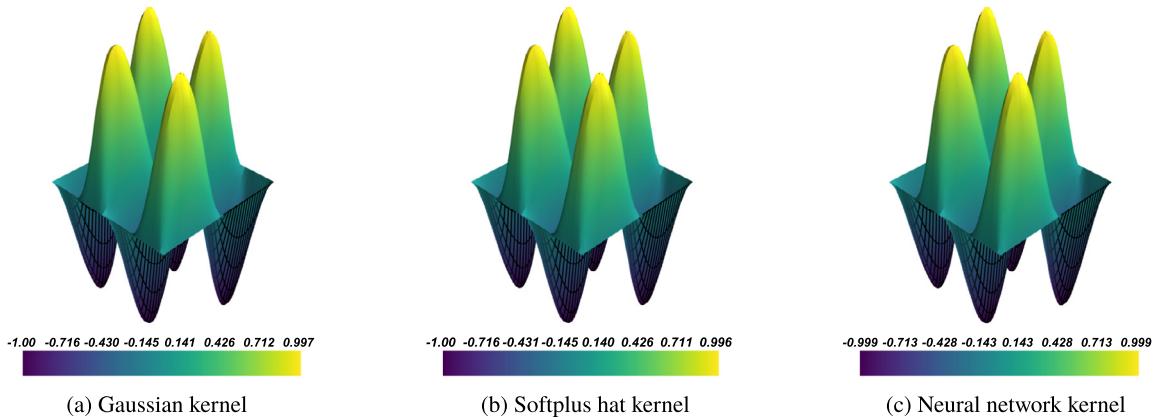


Fig. 20. Solution of PDE (16) using SPINN with different kernels. All simulations have approximately 100 interior nodes, and about 400 interior sampling points. The exact solution is shown in each case as a wiremesh, and the solution computed using SPINN is shown using continuous surface maps.

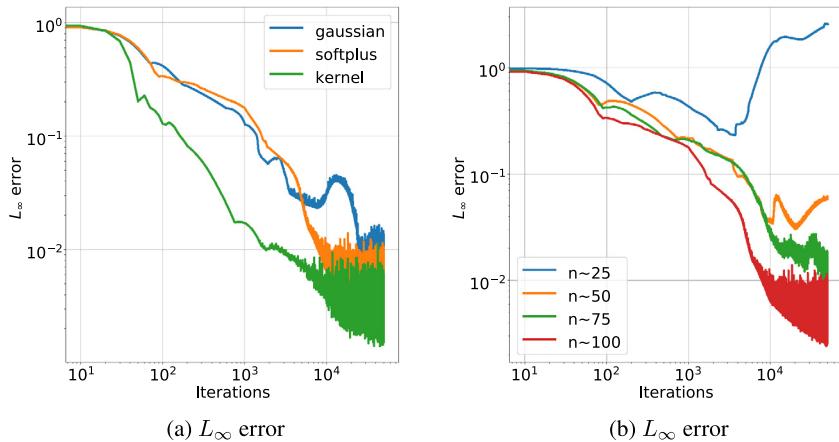


Fig. 21. Evolution of L_∞ errors for the SPINN solution of PDE (16) for different kernels is shown in the figure on the left. The parameters used are identical to that used to generate Fig. 20. The effect of the number of interior nodes on the L_∞ error is shown in the figure on the right. As expected, the errors decrease with larger number of internal nodes.

to the slit, we expect the nodes to have smaller widths. This indeed is what we observe in our simulations and can be seen clearly in Fig. 22b. Notice how the radius of the circles around the nodes is smaller closer to the slit, thereby indicating the adaptivity of the SPINN solution to local gradients in the solution. A comparison of the error of the SPINN solution with respect to a reference finite element solution using a very fine mesh is shown in Fig. 22c.

The mesh for the finite element solution was created using Gmsh [10]. The finite element solution was computed using Fenics [1,23].

3.2.3. Irregular domains

The examples shown so far feature domains with regular geometric shapes. The SPINN algorithm, however, works well on arbitrarily shaped domains too. The solution of the Poisson equation $\nabla^2 u(x, y) + 1 = 0$ on an irregularly shaped domain is shown in Fig. 23a, and a reference finite element solution computed using a fine mesh is shown in Fig. 23c. The distribution of the nodes in this case is shown in Fig. 23b. The simulation used around 450 internal nodes, 110 fixed boundary nodes, 1800 internal sampling points and 220 boundary sampling points. The learning rate was chosen as 10^{-3} and the iterations were carried for 5×10^3 training steps. The L_∞ error of the SPINN solution was found to be around 4.9×10^{-3} .

3.3. Time dependent PDEs

We present a few applications of SPINN to solve time dependent PDEs, namely the heat equation (parabolic PDE), linear advection equation (hyperbolic PDE), Burgers equation (shock formation) and Allen-Cahn equation (periodic boundary conditions and sharp gradients). We chose the various examples to highlight specific features of SPINN as highlighted in the brackets.

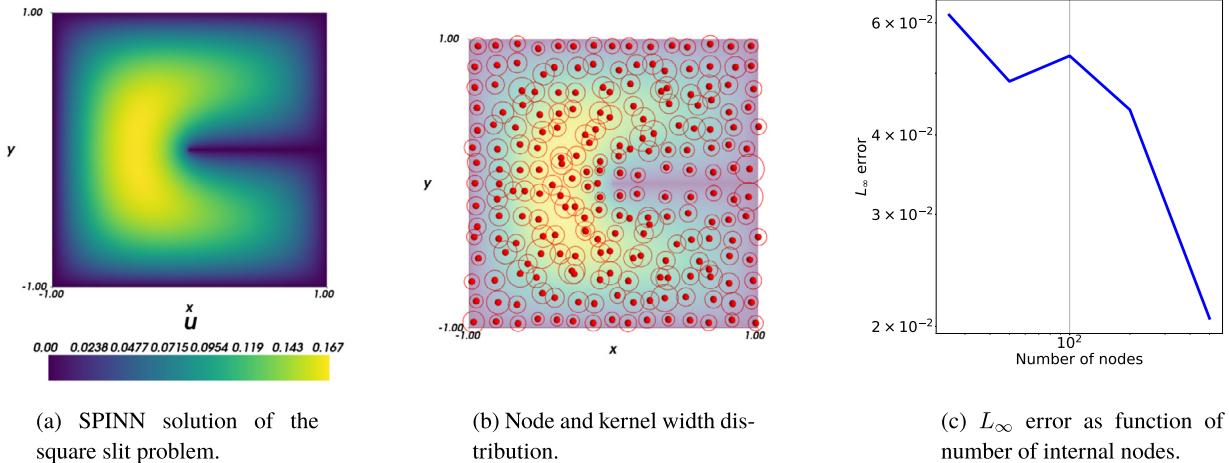


Fig. 22. Solution of the square slit problem. The node and kernel width distributions for the solutions learnt by SPINN are shown, along with a plot of the L_∞ error as a function of the number of nodes.

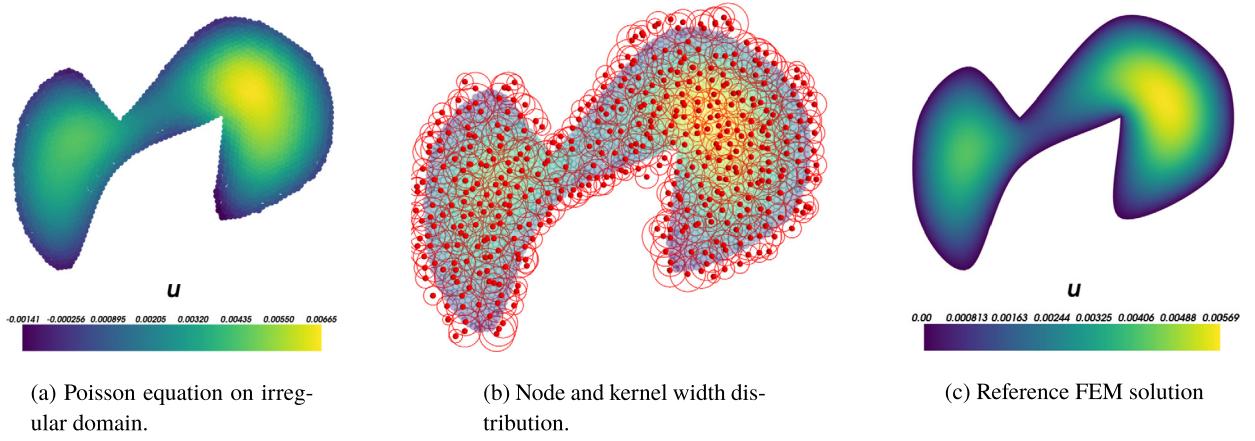


Fig. 23. Illustration of the applicability of the SPINN algorithm to solve PDEs defined on complex geometries. Both the nodal positions and reference finite element solutions are shown for comparison.

3.3.1. Heat equation

We now present examples involving time dependent PDEs. To start with, we consider the one-dimensional heat equation

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} &= c^2 \frac{\partial^2 u(x, t)}{\partial x^2}, \quad x \in (0, L), t \in [0, T], \\ u(x, 0) &= f(x), \quad x \in (0, L), \\ u(0, t) &= u(L, t) = 0, \quad t \in [0, T]. \end{aligned}$$

We record for reference the exact solution to the heat equation displayed above. The coefficients $(b_k)_{k=1}^\infty$ are first defined as the Fourier coefficients of f : $f(x) = \sum_{k=1}^\infty b_k \sin k\pi x$. The coefficients (b_k) are easily computed as $b_k = \frac{2}{L} \int_0^L f(x) \sin \frac{k\pi x}{L} dx$. The exact solution of the heat equation is then computed as

$$u(x, t) = \sum_{k=1}^{\infty} b_k \exp(-\alpha_k^2 t) \sin k\pi x, \quad \alpha_k = \frac{k\pi c}{L}.$$

The example considered in the paper uses the following inputs: $f(x) = 2 \sin \pi x$, $c = 1$, $L = 1$ and $T = 0.2$. The coefficients (b_k) are easily computed as $b_1 = 2$, $b_k = 0$, $k = 2, 3, \dots$

We consider two different methods to solve the heat equation. First, we solve it using the implicit FD-SPINN algorithm

$$u^{n+1} = u^n + c^2 \Delta t u_{xx}^{n+1}.$$

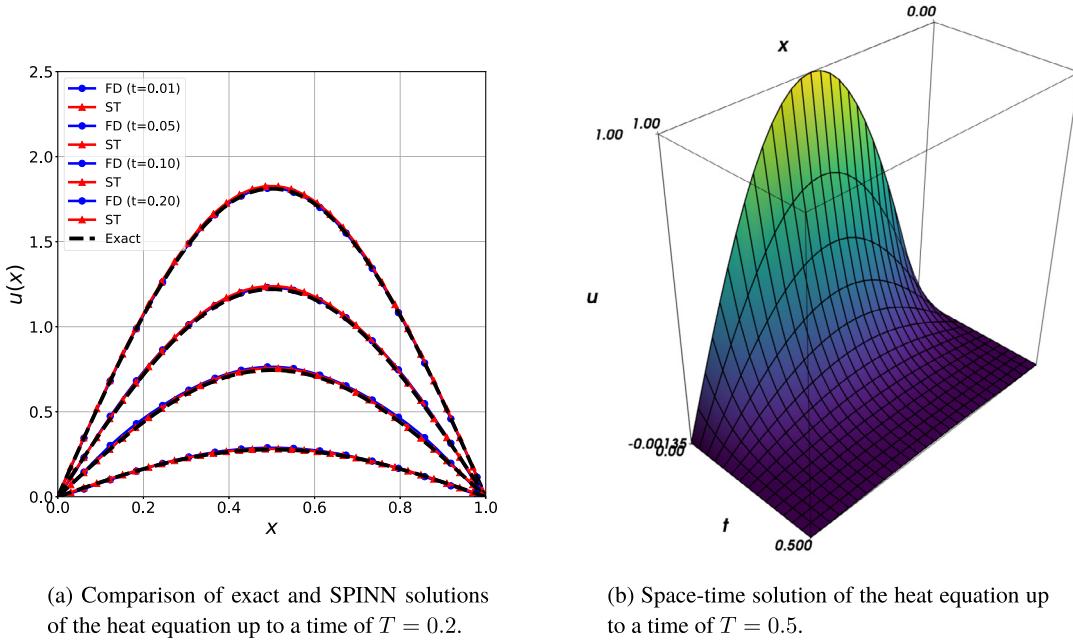


Fig. 24. Solution of the one dimensional heat equation using both the space-time version of SPINN and a first order implicit FD-SPINN algorithm.

It is emphasized that the spatial derivative on the right hand side of the FD-SPINN algorithm displayed above is computed exactly using automatic differentiation, in contrast to typical finite difference schemes which employ a numerical discretization of the derivative operator. We use a learning rate of 10^{-3} , a time step of 0.0025, 20 interior nodes, 200 sample points that are fully sampled. We iterate either until a loss of 2×10^{-5} is attained, or up to 5000 iterations. The L_∞ error at $T = 0.2$ is $O(10^{-2})$.

We show time snapshots of the solution in Fig. 24a. We also solve this problem as a space-time PDE by employing SPINN to simultaneously approximate the solution in space and time. In this case, we solve the problem until $T = 0.5$. We use 100 internal nodes, and 400 sample points which are fully sampled. The learning rate is 10^{-3} . We train the network for 10000 iterations. The L_∞ error is around 2×10^{-2} . Both approaches have similar computational time. The solution is compared with the exact solution in Fig. 24a and the space-time solution is shown in Fig. 24b; the exact solution is also shown as a wireframe for comparison. In Fig. 25, the final positions of the internal nodes learnt by the space-time SPINN and FD-SPINN methods are shown. It is seen that the nodes move to capture the correct solution and that this may result in the nodes even moving outside of the domain of the solution.

3.3.2. Linear advection equation

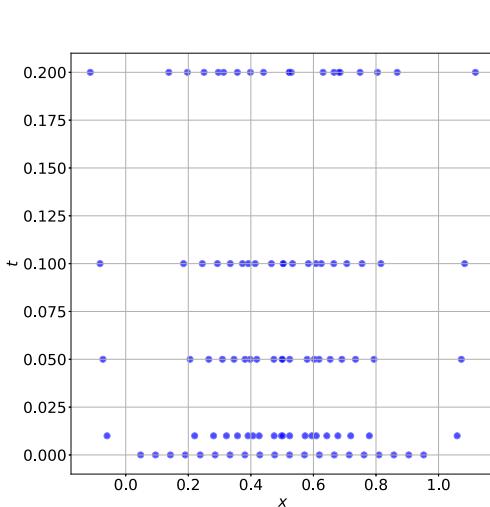
We study next the linear advection equation. The linear advection equation is a classic hyperbolic PDE commonly solved when testing new finite volume algorithms. We solve the following equation

$$\begin{aligned} \frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} &= 0, \quad x \in \mathbb{R}, t \in [0, T], \\ u(x, 0) &= u_0(x), \quad x \in \mathbb{R} \\ u(x, t) &= 0, \quad |x| \rightarrow \infty. \end{aligned}$$

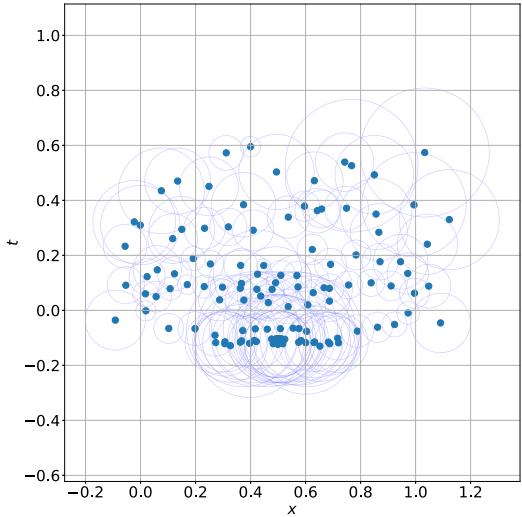
The exact solution is $u(x, t) = u_0(x - at)$. We consider a simple Gaussian pulse, $u(x, 0) = e^{-(x+\mu)/2\sigma^2}$, where $\mu = 0.3$ and $\sigma = 0.15$ and consider the evolution of this, with $a = 0.5$ and $T = 1$. Given the almost compact nature of the initial condition we solve the PDE in a finite spatial domain $[-1, 1]$ and over the time interval $[0, 1]$. Since the initial condition almost vanishes on the boundaries, we set the boundary conditions at $x = -1$ and $x = 1$ uniformly to zero. While this is strictly not correct, we expect the error associated with this to be negligible. We note however that this restriction can be removed by using the Fourier SPINN model to implement the corresponding spatially periodic model.

As done earlier for the heat equation, we solve the problem using both a first order implicit FD-SPINN as well as a space-time SPINN. For the FD-SPINN case we use 15 internal nodes, 100 sample points, and a timestep of 0.0025. We iterate either until a loss of 10^{-6} is attained, or up to 5000 iterations, using a learning rate of 2.5×10^{-3} . For the space-time SPINN we use 40 nodes, with 800 sample points, a learning rate of 10^{-3} and train the network for 5000 iterations.

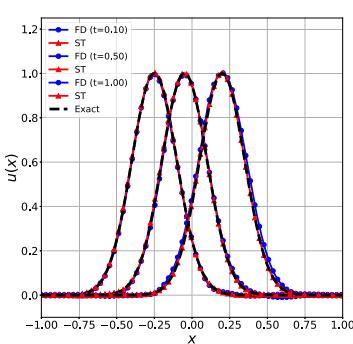
Time snapshots of the solution with a Gaussian pulse as initial condition are shown in Fig. 26a. The L_∞ error at the end of the simulation is $O(10^{-2})$. In Fig. 26b the space-time solution is compared with the exact solution, shown in wireframe.



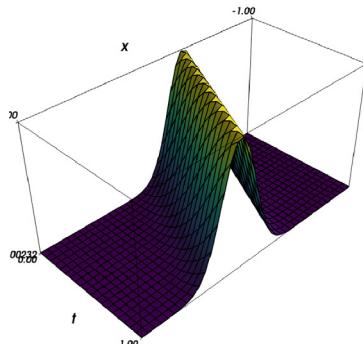
(a) First order Implicit FD-SPINN



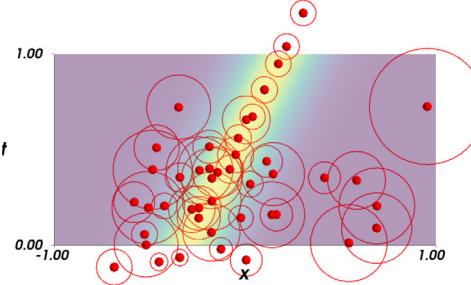
(b) Space-time SPINN

Fig. 25. Location of the nodes for the heat equation for space-time SPINN and FD-SPINN.

(a) Comparison of exact and SPINN solutions



(b) Space-time solution



(c) Node and kernel width distributions

Fig. 26. Solution of the linear advection equation using both FD-SPINN and space-time SPINN. The alignment of the nodes and the kernel widths as shown in Fig. 26c illustrates both the implicit mesh adaptivity and the interpretability of SPINN.

The L_∞ error at the end of the simulation is $O(10^{-2})$ for the space-time SPINN as well. The location of the interior nodes along with the kernel widths is shown in Fig. 26c. The nodes are initially placed uniformly. It is worth emphasizing that the nodes adapt in time and space to capture the features of the solution, which in this case is a traveling wave. We also point out that widths of the kernel are narrow around the peak of the wave while they are broad away from the peak thereby demonstrating mesh-adaptivity. This also illustrates the interpretability of the SPINN algorithm; the position of the nodes and their corresponding widths, as shown in Fig. 26c gives an immediately visual representation of the weights and biases of the SPINN model. We would like to point out that such an interpretation is seldom achievable using conventional methods like PINN.

We note that both the methods require roughly the same order of computational time. For problems with smooth solutions, either of the methods is effective.

3.3.3. Burgers' equation

We next consider a classic non-linear, time-dependent hyperbolic PDE in one spatial dimension, namely the inviscid Burgers' equation. The inviscid and viscous Burgers' equations are important hyperbolic PDEs that arise in many applications. This problem is interesting because it is non-linear, and develops a shock (a discontinuity) in finite time for initially smooth solutions. The inviscid Burgers' equation along with the initial condition reads,

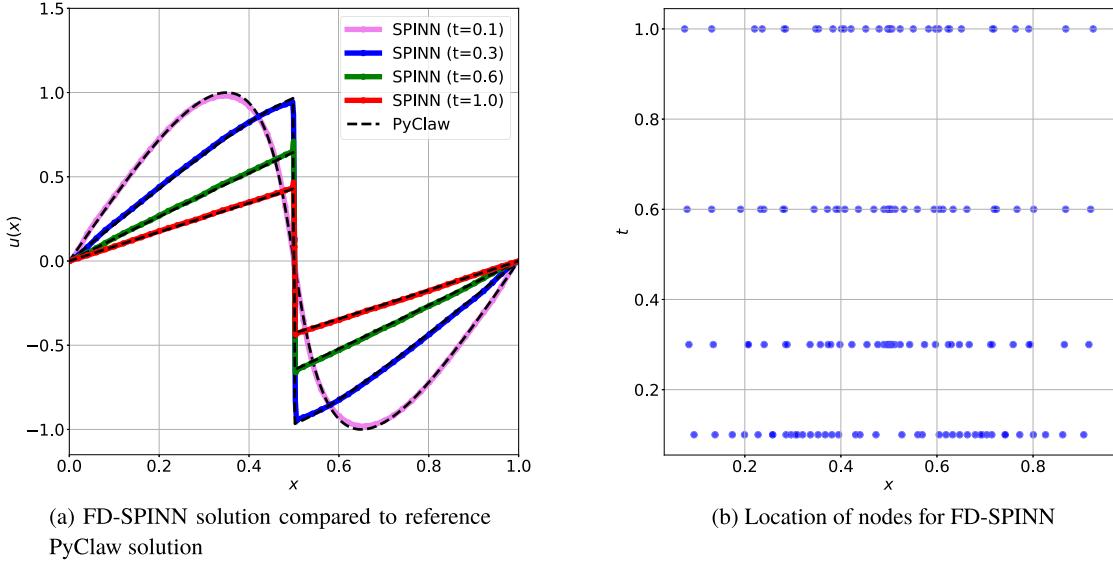


Fig. 27. Comparison of the implicit FD-SPINN solution of the Burgers' equation with reference solution computed using PyClaw for Case1. The location of the nodes is also shown. The shock forms at $x = 0.5$; the corresponding clustering of nodes illustrates the adaptivity of the SPINN algorithm.

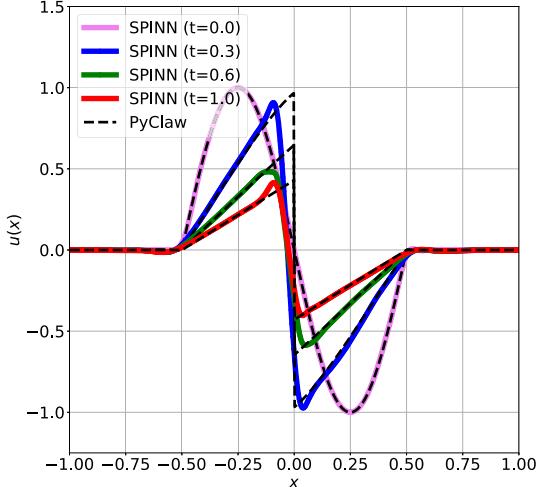
$$\begin{aligned} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} &= 0, \quad x \in [0, 1], t \in [0, T], \\ u(x, 0) &= \sin(2\pi x), \quad x \in [0, 1] \\ u(0, t) &= u(1, t) = 0. \end{aligned}$$

We call this “Case1”. We solve the problem using first order implicit FD-SPINN using Gaussian kernel with 40 internal nodes. We use 400 sampling points (with full sampling) and a learning rate of 10^{-4} . We iterate upto a loss of 10^{-6} or 5000 steps for each timestep and employ a timestep of 0.01. Time snapshots of the solution at different times are shown in Fig. 27a and compared against reference solution obtained using PyClaw [17], a popular and high-quality finite volume package. A shock develops at $x = 0.5$ at around $t = 0.25$ which is clearly captured well by SPINN. The L_1 error is 4.7×10^{-3} and this is computed with reference to the PyClaw solution. We show in Fig. 27b the position of the nodes learnt by the FD-SPINN algorithm. It can be seen that the nodes initially cluster around the peaks of the sine function but cluster around $x = 0.5$, which is the location of the shock for $t > 0.2$ thereby demonstrating the nodal adaptivity implicit in the SPINN algorithm. We also point out that the widths of the kernel at the shock locations are much smaller than the corresponding kernels centered at nodes away from the shock. What is remarkable is that even for a fully inviscid problem as we consider here, and despite using smooth kernels, in this case the Gaussian, the FD-SPINN method is able to capture the shock accurately. This thus demonstrates (i) the adaptivity of the SPINN method; the nodes closer to the shock front have a much smaller kernel width in comparison to nodes away from the shock as expected, and (ii) the ability of SPINN to capture discontinuities in the solution. We present a space-time solution for Case1 in Fig. 28a where we employ 400 internal nodes, 2000 sampling points, a learning rate of 10^{-3} , and train for 10^4 iterations. The results are not as accurate as the FD-SPINN and display some dissipation and the L_1 error is 1.8×10^{-2} . We believe that this is a consequence of an insufficient number of internal nodes. The solution does capture the overall features quite well. We observe that there is no viscosity in this problem; indeed, the solution does display on occasion some oscillatory behavior for both FD-SPINN and ST-SPINN. Since FD-SPINN allows us to employ a limited number of nodes more effectively, it is much more efficient in capturing the correct solution with a smaller number of nodes.

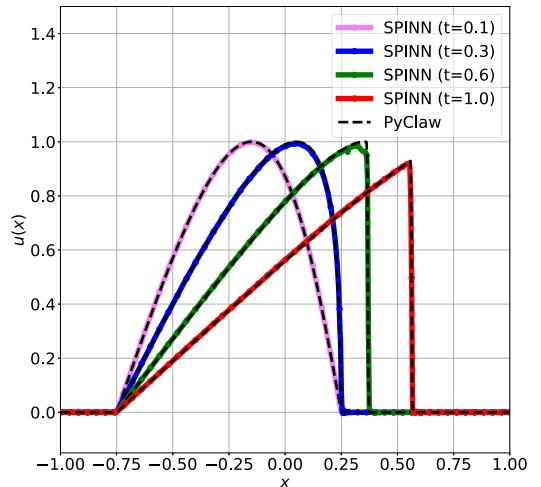
We also present a space-time solution for the Burgers' equation over the domain $x \in [-1, 1]$ using the following initial condition:

$$u_0(x) = \begin{cases} \sin(\pi(x + 0.75)), & -0.75 \leq x \leq 0.25, \\ 0, & \text{otherwise.} \end{cases}$$

We call this “Case2”. This problem also develops a discontinuity. The solution obtained using the space-time SPINN version with a Gaussian kernel is shown in Fig. 29a and compared with a reference simulation again obtained using PyClaw [17]. We employ around 400 nodes in this case, with a learning rate of 2×10^{-3} , 2000 sampling points, and iterate for 10000 steps. The corresponding distribution of nodes and their widths are shown in Fig. 29b. The L_1 error is 1.2×10^{-2} . As remarked earlier in the case of linear advection, we see that the nodes align along the shock front, and the widths of the nodes are correspondingly smaller near the shock front to ensure that the shock is captured adequately. Note also that the

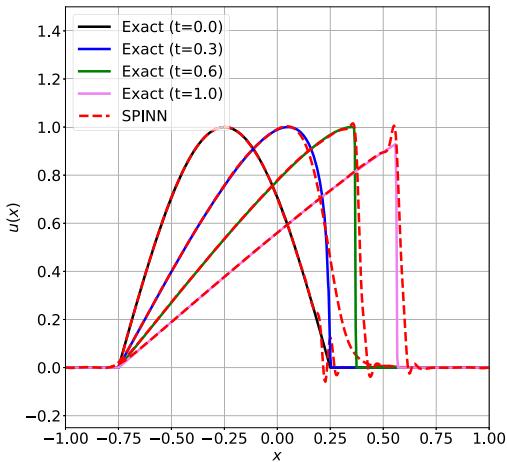


(a) Space-time SPINN solution for Case1

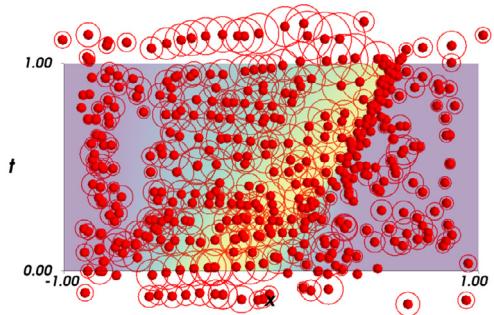


(b) FD-SPINN solution for Case2

Fig. 28. Comparison of the space-time SPINN solution for Case1 and the FD-SPINN solution for Case2 of the Burgers' equation with reference solution computed using PyClaw.



(a) Space-time SPINN solution



(b) Location and width of nodes

Fig. 29. Solution of the inviscid Burgers' equation for Case2 using space-time SPINN with Gaussian kernel and about 400 nodes.

alignment of the nodes starts after $t \simeq 0.5$, which is when the shock forms. It can also be seen that the solution captures the correct shock speed. However, despite the large number of nodes used, we note that there are some oscillations and diffuse behavior near the shock front, unlike the FD-SPINN case. We expect these problems to be solved when the number of nodes is increased, or when a small non-zero artificial viscosity is introduced; we will be investigating this in more detail in a future work. We however point out that unlike works like [29] where the viscous Burgers equation is solved, albeit with a small viscosity, we solve the fully inviscid problem and still get very good results. With the addition of a small viscosity, we expect the performance of both FD-SPINN and space-time SPINN to be better. It is important to note that even the FD-SPINN solution can display oscillations in these cases - the tolerance for iterations has to be set low enough that these are removed. We present a FD-SPINN solution for Case2 in Fig. 28b. The results are in good agreement although at $t = 0.6$ some slight oscillations can be seen. This case was run with 150 nodes, 600 sampling points, a learning rate of 10^{-4} , and trained each timestep upto a tolerance of 10^{-7} or 10000 iterations with a timestep of 0.005. The L_1 error is 1.9×10^{-3} . As can be seen, the new method produces very good results even for an inviscid problem.

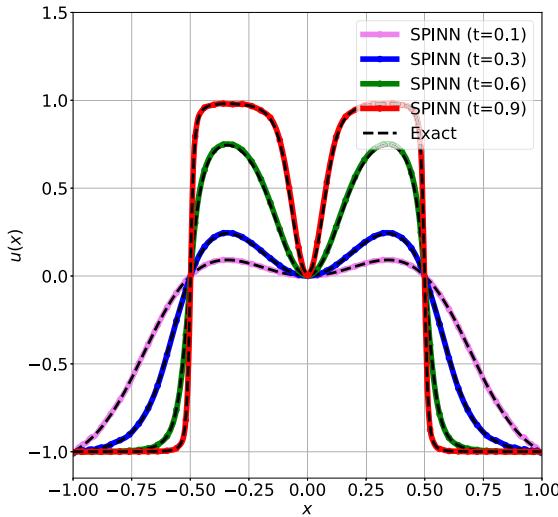


Fig. 30. FD SPINN solution for the Allen-Cahn problem compared with the exact solution available from [29].

3.3.4. Allen-Cahn equation

We present next a solution of the following Allen-Cahn equation with periodic boundary conditions:

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} - 0.0001 \frac{\partial^2 u(x, t)}{\partial x^2} + 5u(x, t)^3 - 5u(x, t) = 0, \quad x \in (-1, 1), t \in (0, 1), \\ u(x, 0) = x^2 \cos \pi x, \\ u(-1, t) = u(1, t) \\ u_x(-1, t) = u_x(1, t). \end{aligned} \tag{19}$$

We consider this PDE for two reasons: (i) to illustrate the fact that periodic boundary conditions can be handled easily with SPINN, and (ii) to make a comparison with the PINN solution for the same equation as carried out in [29]. We solve the Allen-Cahn PDE (19) using the implicit finite difference version of SPINN. We use 100 nodes, with 500 samples, and a learning rate of 10^{-3} , we iterate until a loss of 10^{-6} is reached or 5000 iterations are complete and use a timestep of 0.005. This problem completes execution on a CPU in about 4 minutes. We compare the result with the solution made available in [29]. The results are shown in Fig. 30. The loss is consistently less than 10^{-6} and eventually is around 10^{-8} . The L_1 error is 4.4×10^{-3} which we compute using the reference solution given in [29]. This shows the excellent accuracy obtained for this problem with only 100 SPINN nodes.

3.4. Fluid dynamics

Our final example involves solving the steady incompressible viscous Navier-Stokes equations in two spatial dimensions. We consider the classic lid-driven cavity problem [11]. Consider a unit square placed on the x-axis in the region $[0, 1] \times [0, 1]$ and filled with a fluid with a density $\rho = 1$. Let the Cartesian velocity components be given by (u, v) and the pressure of the fluid be p . The governing differential equations for the fluid when it attains a steady flow is,

$$\begin{aligned} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0, \\ u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \nabla^2 u, \\ u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} &= -\frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \nabla^2 v, \end{aligned}$$

where ν is the kinematic viscosity of the fluid. The first equation is the conservation of mass (also called the continuity equation) and the subsequent two are the momentum equations. The boundary conditions are given as,

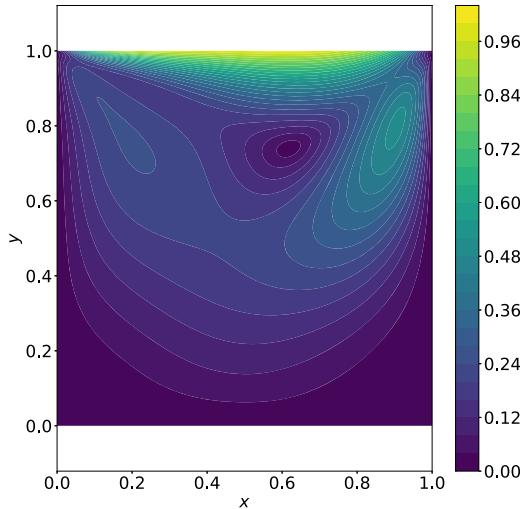


Fig. 31. Velocity magnitude of flow inside the driven cavity with a fluid having kinematic viscosity $\nu = 0.01$ using SPINN with a Gaussian kernel.

Table 1
Comparison of L_∞ error of SPINN and PySPH solution of centerline velocity components with respect to the solution reported in [11].

Method	L_∞ error: u	L_∞ error: v
SPINN 100 nodes	0.073718	0.035425
SPINN 200 nodes	0.022839	0.006325
SPINN 300 nodes	0.012010	0.004969
PySPH 50 × 50 particles	0.042251	0.022198

$$u(x, 0) = 0$$

$$u(x, 1) = 1$$

$$u(0, y) = 0$$

$$u(1, y) = 0$$

$$\frac{\partial p}{\partial n} = 0,$$

where n is the normal vector along the boundary. The velocity v is zero on the boundary of the square. The walls are modeled with no-slip boundary conditions consistent with the behavior of a viscous fluid. The Reynolds number, Re for this problem is defined as $Re = 1/\nu$. There is no exact solution for this problem but there are many numerical solutions available. We compare the horizontal and vertical profile of the velocity along the center-line of the square with the classic results of [11] at a Reynolds number of 100.

The solution of this problem requires that SPINN returns a vector representing the solution (u, v, p) at each point where the solution is desired, unlike the previously studied scalar differential equations. The magnitude of the velocity computed using SPINN with a Gaussian kernel is shown in Fig. 31. In Figs. 32a and 32b the velocity profile along the center-lines are compared with those of [11] as the number of internal nodes is varied. We obtain good results with around 225 internal nodes for this problem. The optimizer is run for 10000 iterations with a learning rate of 5×10^{-4} . This example was chosen to demonstrate the ability of SPINN to model a non-linear system of PDEs that frequently arise in the modeling of physical systems. In Figs. 32c and 32d the velocity profiles are computed using an SPH simulation using an initial 10×10 , 25×25 and 50×50 grid of particles for the fluid and compared with the results of [11]. These simulations were made using the standard cavity example available as part of the open source PySPH package [33]. The lid driven cavity problem has been simulated using the EDAC-SPH scheme [34] where the NS equations are evolved in time and driven to a steady state. As can be seen by a close inspection of the results, the solution obtained using SPINN with 225 nodes is more accurate than that obtained by the SPH simulation using 2500 particles. This suggests that SPINN behaves like a higher-order method.

We report the L_∞ error by using the data provided in [11] as a reference. We compute the u and v velocities along the centerline from the SPINN and PySPH solutions and then compute the values at the corresponding location from [11]. We compute the L_∞ error using these values. We show the computed values in Table 1. As can be clearly seen, the results of SPINN are very accurate even with a very small number of nodes.

The algorithms used for SPINN and SPH are considerably different, however, a comparison of the execution times provides some insight into the relative computational performance of the methods. We note that PySPH employs a highly optimized

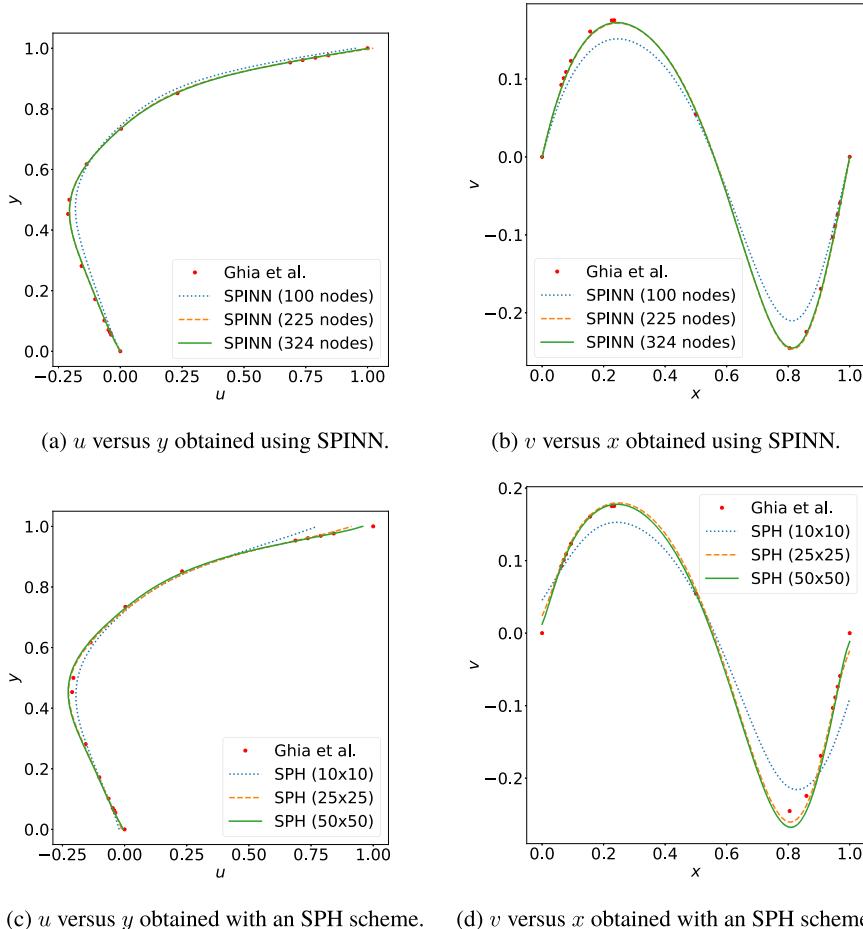


Fig. 32. Comparison of the centerline velocity obtained for the lid-driven cavity problem using SPINN and with an SPH scheme as compared with [11].

implementation whereas the SPINN implementation is largely a proof of concept and in this particular case inefficiently computes the influence of all nodes at a given point. The SPINN case with 225 nodes and 10000 iterations takes around 825 seconds to complete. The SPH simulation with 2500 particles for 33000 iterations takes 517 seconds of computational time on a single core of an Intel i7 CPU running at 2.9 GHz. This shows that for a similar accuracy, SPINN is competitive with a traditional meshless method.

4. Discussion

The primary purpose of the various examples presented earlier is to provide a proof-of-concept demonstration that the SPINN model works well for a large class of ODEs and PDEs. There are however many aspects which can be improved. We discuss some of these here; we will be investigating these systematically in forthcoming publications.

4.1. Interpretability of SPINN

In this section, we further clarify the precise sense in which we say that SPINN is interpretable. Once a specific kernel has been chosen, either in the form of basis functions like RBF, softplus hat, neural network kernel, etc., the SPINN approximation reduces to the ansatz (1). The interpretability of this ansatz when choosing simple basis functions like RBF and softplus hat is evident. In the case when the kernel network is chosen as an entire neural network, we note that we do not provide a detailed interpretation of the weights and biases of the kernel network. Such an interpretation, however, is not necessary in the current setting since all that we require is that the kernel network, which is typically much smaller than a full DNN that would be employed in methods like PINN, provides a particular choice of basis φ in (1). This is to be compared with traditional approximations, like Daubechies wavelets for instance, where the precise shape of the mother wavelet is not *interpretable* in the classical sense, but the wavelet transform itself as a whole is interpretable. This is the sense in which we consider SPINN to be interpretable when using neural network kernels - given a particular choice of basis function, which is what the kernel network encodes, the specific way in which shifted and scaled versions of this

basis function are linearly combined to provide the SPINN approximation provides a representation that is interpretable in exactly the same sense that traditional representations are. The difference between our interpretation of SPINN and basis function interpretations of PINN as in [4] is that by confining all non-interpretability to a much smaller neural network kernel, we are able to peer into the workings of the SPINN algorithms in a manner that can be done with traditional meshless methods. Indeed, as we have demonstrated with various examples, we can post-process the SPINN solution to extract nodal positions and local widths of the basis functions. This in turn permits us to highlight the precise sense in which the SPINN algorithm handles regions of sharp gradients or shocks by locally adapting a single parent basis function. As seen in ODE (14) which has a highly oscillatory solution, there is no natural means to design the architecture of PINNs. In sharp contrast, in SPINN we choose the number of nodes in a manner that is similar to how the discretization is chosen based on a Nyquist sampling condition in traditional meshless methods. Such a detailed analysis and network architectural choices involving local information is not easily accomplished when employing methods like PINN, even when interpreted as global basis functions.

4.2. Boundary conditions

For all the variants of SPINN presented in this work, the loss is defined as a weighted sum of the interior and boundary losses. The boundary loss, in particular, is enforced as a penalty term that is proportional to the square of the predicted and imposed boundary values. The constant of proportionality, however, is arbitrarily chosen and varies for each problem. This is a well known limitation of penalty based techniques to enforce constraints. While Dirichlet boundary conditions are relatively easy to enforce, capturing Neumann boundary conditions require careful choice of the fixed and free nodes. For instance, a fixed node at a Neumann boundary when using Gaussian kernels will lead to an infinite indeterminacy of the corresponding coefficient since the slope of the Gaussian kernel is zero at its center. This translates to convergence issues with the loss minimization algorithm. As a simple solution, we use fixed nodes only on the Dirichlet boundaries. We observe that the nodes at times move outside the domain to enforce the boundary condition properly. However, we point out that there are other means to enforce boundary conditions like having a fixed layer of nodes along the boundary, as is done in some particle based methods like Smoothed Particle Hydrodynamics (SPH) [45]. Alternatively, one could build in the boundary conditions directly into the solution, as is done in works such as [2].

4.3. Time dependent PDEs

In the context of FD-SPINN algorithms for time dependent PDEs, we used a first order scheme for time marching. However it is in principle straightforward to implement higher order schemes to control the time discretization errors using methods such as those proposed in [37]. We also point out that the FD-SPINN algorithms presented here are different from the corresponding finite difference schemes since the spatial derivatives are handled *exactly* using automatic differentiation. Automatic differentiation has been used in the context of finite element discretization problems in both static [39] and dynamic [31] PDEs, but the implicit finite difference schemes we propose here provide a systematic means to develop a variety of efficient time marching schemes.

4.4. Computational efficiency

The current implementation of SPINN has been designed as a proof-of-concept. While the problems considered in this work can be solved with very little computational effort, a naive implementation will not scale well for larger problems. This is because the method currently evaluates the effect of *all* the nodes on a given sample point making the problem $O(NM)$ given N nodes and M samples. We have investigated performing spatial convolutions using the pytorch-geometric package [8] to accelerate this computation by restricting the interaction of sampling points only with their nearest nodes, thereby reducing the problem to $O(nM)$ where n is the typical number of neighbors. While preliminary results are encouraging and allow us to use more nodes, there are some limitations and constraints that need to be investigated further.

We note however that SPINN is more efficient than PINN as we have demonstrated in some of the results here. Many applications involving PINN that we find in the literature use DNNs with many hidden layers and $O(100)$ neurons in each layer. In contrast, we note that the size of the SPINN architecture, even when using kernel networks, is significantly smaller in comparison. A simple estimate is provided here to illustrate this difference better: for the SPINN architecture as shown in Fig. 2 the number of connections is $O(N)$, where N is the number of nodes in the meshless representation, while for a fully connected network of the same size the corresponding number of connections is $O(N^2)$.¹ We remark that though PINNs need not necessarily have the same size of the corresponding SPINN architecture, as noted above, the size of typical DNNs used in PINN is much larger than what we use in SPINN. This highlights the sparsity of the SPINN model, and hence the corresponding gain in computational efficiency. Over and above this, we note that by choosing the kernel to be compact functions, SPINN can in principle be made as efficient as a traditional meshless method. This is a significant computational advantage.

¹ The actual numbers for the case shown in Fig. 2 are $2dN + (3 + 12 + 12 + 3) + N = (2d + 1)N + 30$, while the number of connections in the corresponding DNN is $(Nd^2 + dN^2 + (3N^2 + 12N^2 + 12N^2 + 3N^2 + N) = (30 + d)N^2 + (d^2 + 1)N$. Here d is the physical dimension of the problem.

It is difficult to perform a direct performance comparison between a traditional meshless method and SPINN as it is currently implemented. The algorithms are very different and this makes a direct comparison not very meaningful. However, as we have done in section 3.4, the SPINN solution is more accurate than an optimized SPH solution using an order of magnitude more computational elements (225 versus 2500) and requires about 1.5 times the computational time on the same computational hardware. This is quite a remarkable result.

As discussed earlier, our current implementation is not optimized and relies on `pytorch`. Many open source meshless frameworks, like PySPH [33] that was used in our comparison, are highly optimized. On the other hand, it is easy to see that the function approximation using SPINN is of the same computational order as a conventional meshless method. Based on some preliminary computations using the geometric convolutions from `pytorch-geometric`, we observe that the evaluation of the solution using the neural network for the interior and boundary points takes about 20% of the entire computational cost. The remainder of the time is taken by the computation of the derivatives, losses, and optimizer. Therefore if we are able to improve the computational performance of the forward operation and the resultant derivatives, which is certainly possible, we can obtain a performance that is comparable with that of a conventional meshless method. This along with the fact that neural networks are easily implemented on GPUs makes for a compelling case for SPINN.

4.5. Finite element based extensions of SPINN

We have used meshless approximations to motivate the development of the SPINN architecture in this work. However, conforming mesh based representations like finite elements could also be used to develop the corresponding neural network generalizations, although the connection is not straightforward. The difficulty arises because of the fact that enforcing mesh conformity places more constraints on the architecture. There are theoretical results elaborating the relation between finite element models and DNNs. For instance in [15] the authors show how every piecewise linear finite element model can be mapped to a ReLU DNN, while in [27] higher order finite elements have been investigated along similar lines. The advantage in using meshless representations over conforming mesh representations based on the finite elements is that the corresponding DNN has more flexibility in how the nodes move, and how the kernel widths adapt. In addition it also allows for a variety of generalizations like the Fourier and wavelet generalizations of SPINN.

5. Conclusion

To conclude, we have presented a new class of sparse DNNs which we call SPINN - Sparse, Physics-based and partially Interpretable Neural Networks - which naturally generalize classical meshless methods while retaining the advantages of DNNs. The key features of SPINN are:

- (i) It is interpretable (in the specific sense discussed earlier and summarized next), in sharp contrast to DNN based methods.
- (ii) All the non-interpretability is reduced to a much smaller kernel network in contrast to a full DNN, thereby permitting an interpretation of the trained network in a manner similar to meshless methods.
- (iii) It is efficient in comparison with a DNN with the same number of neurons on account of its sparsity.
- (iv) It is physics-based since the loss function depends directly on the strong form of a PDE or its variational form.
- (v) It implicitly encodes mesh adaptivity as part of its training process.
- (vi) It is able to resolve non-smoothness in solutions, as in the case of shocks.
- (vii) It suggests new classes of numerical algorithms for solving PDEs. For instance, the Fourier-SPINN extension was illustrated in this work.

We have demonstrated these aspects of SPINN using a variety of ODEs and PDEs, and thus present SPINN as a novel numerical method that seamlessly bridges traditional meshless methods and modern DNN-based algorithms for solving PDEs. Recognizing this link opens fresh avenues for developing new numerical algorithms that blend the best of both worlds. Even a mere re-expression of meshless algorithms as a SPINN model makes it easier to enhance traditional meshless methods along the lines of differentiable programming. We conclude by emphasizing that the strength of SPINN lies in its interpretability, its backwards-compatibility with traditional discretizations, and forward-compatibility with DNN based methods.

CRediT authorship contribution statement

Both Amuthan Ramabathiran and Prabhu Ramachandran contributed equally to: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, and Writing of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. ReLU networks and piecewise linear finite element approximation

In this section, we illustrate the connection between SPINN and DNN representations for PDEs that have been previously studied. To keep the discussion concrete we focus on the special case of one spatial dimension. Consider the problem of minimizing the functional $I : H_0^1([0, 1]) \rightarrow \mathbb{R}$,

$$I(u) = \frac{1}{2} \int_0^1 \left(\frac{du(x)}{dx} \right)^2 dx - \int_0^1 f(x)u(x) dx, \quad (20)$$

where $f \in L^2([0, 1])$. A standard argument shows that the Euler-Lagrange equation corresponding to the minimization of the functional (20) is the second order ODE:

$$\begin{aligned} \frac{d^2u(x)}{dx^2} + f(x) &= 0, \quad x \in [0, 1], \\ u(0) &= 0, \quad u(1) = 0. \end{aligned} \quad (21)$$

To illustrate this, we consider the special case of a piecewise linear finite element approximation of the solution of (21). The basis functions for a piecewise linear finite element approximation can be equivalently thought of as ReLU network with one hidden layer. For this case, a convenient SPINN architecture is the following ReLU network with one hidden layer:

$$u(x) = \sum_{i=0}^N w_i \text{ReLU}(x - x_i), \quad (22)$$

Following [15], we outline the relation between a piecewise linear finite element representation of a function $u : [a, b] \rightarrow \mathbb{R}$ and neural networks with ReLU activation functions. Letting $(x_i)_{i=0}^N$ be a partition of $[a, b]$, such that $x_0 = a$, $x_N = b$, and $x_i < x_{i+1}$ for $0 \leq i < N$, the piecewise linear basis function $N_i(x)$, for $1 < i < N$, is given by

$$N_i(x) = \begin{cases} \frac{x_i - x}{x_i - x_{i-1}}, & x \in [x_{i-1}, x_i], \\ \frac{x - x_i}{x_{i+1} - x_i}, & x \in [x_i, x_{i+1}]. \end{cases} \quad (23)$$

An important observation that connects the finite element approximation of u with ReLU neural networks is the fact that the basis function in (23) can be written as a linear combination of ReLU functions in the following form:

$$N_i(x) = \frac{1}{h_i} \text{ReLU}(x - x_{i-1}) - \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \text{ReLU}(x - x_i) + \frac{1}{h_{i+1}} \text{ReLU}(x - x_{i+1}), \quad (24)$$

where we use the symbols $h_k = x_k - x_{k-1}$, $1 < k < N$, to denote the lengths of the various elements in a given partition of $[a, b]$. Using (23) and (24), we can write the piecewise linear finite element approximation of u , namely

$$u(x) = \sum_{i=0}^N N_i(x) U_i, \quad (25)$$

as

$$u(x) = \sum_{i=0}^{N-1} (\theta_{i+1} - \theta_i) \text{ReLU}(x - x_i), \quad (26)$$

where $\theta_i = (U_i - U_{i-1})/h_i$. The representation (26) informs us that a piecewise linear finite element approximation of 1d functions is equivalent to a DNN with one hidden layer with weights and biases consistent with (26).

In [15], the authors proposed a hybrid method wherein they use the representation (26), with the weights (θ_i) computed using the standard finite element method holding the mesh fixed, and the biases (x_i) are computed by minimizing the loss I as a function of the biases (x_i) for fixed values of the weights (θ_i) . This staggered approach, however, does not take full advantage of the variety of stochastic gradient algorithms that have been developed for DNNs. In contrast, the SPINN architecture which we propose in this work does not use a staggered approach, and is more efficient.

Even in this simple case, we note the following features: (i) the weights connecting the input layer, which just takes in x , and the hidden layer, are equal to 1 uniformly, (ii) the biases of the hidden layer are directly interpretable as the position of the nodes of the finite element discretization, and (iii) the weights connecting the hidden layer and the output layer are interpretable in terms of the nodal values of the corresponding finite element solution. We also see that the number of neurons in the hidden layer is just the number of interior nodes in the finite element discretization. This is in sharp contrast to the approach followed in Physics Informed Neural Networks (PINN) [29], or the Deep-Ritz method [7], which employ dense neural networks, and hence are not easily interpretable.

Appendix B. Code design

The source code for SPINN is freely available at <https://github.com/nn4pde/SPINN>. We use the Python programming language, the PyTorch [28] library, and NumPy [13]. In addition the code employs the following libraries for visualization and plotting, matplotlib [16] is used for the simpler plots and Mayavi [35] is used for more complex three-dimensional plots. We use the pytorch-geometric [8] package to demonstrate the use of geometric convolutions as a means to accelerate the performance, however this is an optional requirement. Finally, every plot shown in this manuscript is completely automated using the automan [32] package.

Our code follows a simple object-oriented design employing the following objects:

- (i) The PDE base class and its sub-classes provide the common methods that one would need to override to define a particular set of ordinary/partial differential equations to solve along with their boundary conditions.
- (ii) Subclasses of `torch.nn.Module` manage the SPINN models.
- (iii) The Optimizer class manages the loss optimization.
- (iv) The Plotter base class provides routines for live plotting and storing the results.
- (v) The App base class manages the creation and execution of the objects mentioned above to solve a particular problem.

Each problem demonstrated has a separate script which can be executed standalone along with the ability to show the results live. The code is written in a manner such that every important parameter can be configured through the command line. This is used in our automation script, `automate.py`, which uses the `automan` framework to automate the creation of every figure we present in this work. We provide the name of the specific scripts for the different differential equations in the sections that follow; these can be found in the `code` sub-directory of the repository. Further, the individual parameters used to obtain the various plots can be found by perusing `automate.py`. It bears emphasis that every single result presented here is fully reproducible and can be generated by running a single command.

References

- [1] Martin S. Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E. Rognes, Garth N. Wells, The FEniCS project version 1.5, *Arch. Numer. Softw.* 3 (100) (2015).
- [2] Jens Berg, Kaj Nyström, A unified deep artificial neural network approach to partial differential equations in complex geometries, *Neurocomputing* 317 (2018) 28–41.
- [3] Zhiqiang Cai, Jingshuang Chen, Min Liu, Xinyu Liu, Deep least-squares methods: An unsupervised learning-based numerical method for solving elliptic PDEs, *J. Comput. Phys.* 420 (2020) 109707.
- [4] Eric C. Cyr, Mamikon A. Gulian, Ravi G. Patel, Mauro Perego, Nathaniel A. Trask, Robust training and initialization of deep neural networks: An adaptive basis viewpoint, in: Jianfeng Lu, Rachel Ward (Eds.), *Proceedings of The First Mathematical and Scientific Machine Learning Conference, 20–24 Jul 2020, PMLR*, in: *Proceedings of Machine Learning Research*, vol. 107, Princeton University, Princeton, NJ, USA, 2020, pp. 512–536.
- [5] Vikas Dwivedi, Balaji Srinivasan, Physics informed extreme learning machine (PIELM): A rapid method for the numerical solution of partial differential equations, *Neurocomputing* 391 (2020) 96–118.
- [6] W. E, C. Ma, L. Wu, Machine learning from a continuous viewpoint, I, *Sci. China Math.* 63 (2020) 2233–2266.
- [7] E. Weinan, Bing Yu, The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems, *Commun. Math. Stat.* 6 (2018) 1–12.
- [8] Matthias Fey, Jan E. Lenssen, Fast graph representation learning with PyTorch Geometric, in: *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- [9] Daria Fokina, Ivan Oseledets, Growing axons: greedy learning of neural networks with application to function approximation, *e-prints*, page, arXiv: 1910.12686, October 2019.
- [10] C. Geuzaine, J.-F. Remacle, Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities, *Int. J. Numer. Methods Eng.* 79 (2009) 1309–1331.
- [11] U. Ghia, K.N. Ghia, C.T. Shin, High-Re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method, *J. Comput. Phys.* 48 (1982) 387–411.
- [12] Jiequn Han, Arnulf Jentzen, E. Weinan, Solving high-dimensional partial differential equations using deep learning, *Proc. Natl. Acad. Sci. USA* 115 (34) (2018) 8505–8510.
- [13] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, Travis E. Oliphant, Array programming with NumPy, *Nature* 585 (7825) (September 2020) 357–362.
- [14] Cuiyu He, Xiaozhe Hu, Lin Mu, A mesh-free method using piecewise deep neural network for elliptic interface problems, *CoRR*, arXiv:2005.04847 [abs], 2020.
- [15] Juncai He, Lin Li, Jinchao Xu, Chunyue Zheng, Relu deep neural networks and linear finite elements, *J. Comput. Math.* 38 (3) (2020) 502–527.
- [16] J.D. Hunter, Matplotlib: a 2d graphics environment, *Comput. Sci. Eng.* 9 (3) (2007) 90–95.
- [17] David I. Ketcheson, Kyle T. Mandli, Aron J. Ahmadia, Amal Alghamdi, Manuel Quezada de Luna, Matteo Parsani, Matthew G. Knepley, Matthew Emmett, PyClaw: Accessible, Extensible, Scalable Tools for Wave Propagation Problems, *SIAM J. Sci. Comput.* 34 (4) (November 2012) C210–C231.
- [18] Diederik P. Kingma, Jimmy Ba, Adam: a method for stochastic optimization, Published as a conference paper at the in: *3rd International Conference for Learning Representations*, San Diego, 2015, 2014.
- [19] I.E. Lagaris, A. Likas, D.I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, *IEEE Trans. Neural Netw.* 9 (5) (1998) 987–1000.
- [20] I.E. Lagaris, A.C. Likas, D.G. Papageorgiou, Neural-network methods for boundary value problems with irregular boundaries, *IEEE Trans. Neural Netw.* 11 (5) (September 2000) 1041–1049.
- [21] Kookjin Lee, Nathaniel A. Trask, Ravi G. Patel, Mamikon A. Gulian, Eric C. Cyr, Partition of unity networks: deep hp-approximation, *CoRR*, arXiv: 2101.11256 [abs], 2021.

- [22] Ziqi Liu, Wei Cai, Zhi-Qin John Xu, Multi-scale deep neural network (MscaleDNN) for solving Poisson-Boltzmann equation in complex domains, *Commun. Comput. Phys.* 28 (5) (2020) 1970–2001.
- [23] Anders Logg, Kent-Andre Mardal, Garth N. Wells, et al., *Automated Solution of Differential Equations by the Finite Element Method*, Springer, 2012.
- [24] Zichao Long, Yiping Lu, Xianzhong Ma, Bin Dong, PDE-net: learning PDEs from data, in: Jennifer Dy, Andreas Krause (Eds.), *Proceedings of the 35th International Conference on Machine Learning*, Stockholm Sweden, 10–15 Jul 2018, PMLR, in: *Proceedings of Machine Learning Research*, vol. 80, 2018, pp. 3208–3216.
- [25] Lu Lu, Xuhui Meng, Zhiping Mao, George Em Karniadakis, DeepXDE: a deep learning library for solving differential equations, *SIAM Rev.* 63 (1) (2021) 208–228.
- [26] Lu Lu, Raphael Pestourie, Wenjie Yao, Zhicheng Wang, Francesc Verdugo, Steven G. Johnson, Physics-informed neural networks with hard constraints for inverse design, preprint, arXiv:2102.04626, 2021.
- [27] J.A.A. Opschoor, P.C. Petersen, Ch. Schwab, Deep ReLU networks and high-order finite element methods, Technical Report 2019-07, Seminar for Applied Mathematics, ETH, Zürich, Switzerland, 2019.
- [28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, Soumith Chintala, PyTorch: an imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019, pp. 8024–8035.
- [29] Maziar Raissi, Paris Perdikaris, George E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
- [30] Maziar Raissi, Paris Perdikaris, George Em Karniadakis, Multistep neural networks for data-driven discovery of nonlinear dynamical systems, preprint, arXiv:1801.01236, 2018.
- [31] Amuthan Arunkumar Ramabathiran, S. Gopalakrishnan, Automatic finite element formulation and assembly of hyperelastic higher order structural models, *Appl. Math. Model.* 38 (11) (2014) 2867–2883.
- [32] Prabhu Ramachandran, automan: a python-based automation framework for numerical computing, *Comput. Sci. Eng.* 20 (5) (Sep./Oct. 2018) 81–97.
- [33] Prabhu Ramachandran, Aditya Bhosale, Kunal Puri, Pawan Negi, Abhinav Muta, Dinesh Adepu, Dileep Menon, Rahul Govind, Suraj Sanka, Amal S. Sebastian, Ananya Sen, Rohan Kaushik, Anshuman Kumar, Vikas Kurapati, Mrinalgouda Patil, Deep Tavker, Pankaj Pandey, Chandrashekhar Kaushik, Arkopal Dutt, Arpit Agarwal, PySPH: a Python-based framework for smoothed particle hydrodynamics, preprint, arXiv:1909.04504, 2020.
- [34] Prabhu Ramachandran, Kunal Puri, Entropically damped artificial compressibility for SPH, *Comput. Fluids* 179 (30) (January 2019) 579–594.
- [35] Prabhu Ramachandran, Gaël Varoquaux, Mayavi: 3D visualization of scientific data, *Comput. Sci. Eng.* 13 (2) (2011) 40–51.
- [36] E. Samaniego, C. Anitescu, S. Goswami, V.M. Nguyen-Thanh, H. Guo, K. Hamdia, X. Zhuang, T. Rabczuk, An energy approach to the solution of partial differential equations in computational mechanics via machine learning: concepts, implementation and applications, *Comput. Methods Appl. Mech. Eng.* 362 (2020) 112790.
- [37] Xing Shen, Xiaoliang Cheng, Kewei Liang, Deep Euler method: solving ODEs by approximating the local truncation error of the Euler method, e-prints, page, arXiv:2003.09573, March 2020.
- [38] Justin Sirignano, Konstantinos Spiliopoulos, DGM: a deep learning algorithm for solving partial differential equations, *J. Comput. Phys.* 375 (2018) 1339–1364.
- [39] E. Tijskens, D. Roose, H. Ramon, J. De Baerdemaeker, Automatic differentiation for solving nonlinear partial differential equations: an efficient operator overloading approach, *Numer. Algorithms* 30 (2002) 259–301.
- [40] Jihong Wang, Zhi-Qin John Xu, Jiwei Zhang, Yaoyu Zhang, Implicit bias with Ritz-Galerkin method in understanding deep learning for solving PDEs, CoRR, arXiv:2002.07989 [abs], 2020.
- [41] Sifan Wang, Hanwen Wang, Paris Perdikaris, On the eigenvector bias of Fourier feature networks: from regression to solving multi-scale PDEs with physics-informed neural networks, *Comput. Methods Appl. Mech. Eng.* 384 (2021) 113938.
- [42] Zhongjian Wang, Zhiwen Zhang, A mesh-free method for interface problems using the deep learning approach, *J. Comput. Phys.* 400 (2020) 108963.
- [43] John R. Williams, Kevin Amarantunga, Introduction to wavelets in engineering, *Int. J. Numer. Methods Eng.* 37 (14) (1994) 2365–2388.
- [44] A. Yazdani, L. Lu, M. Raissi, G.E. Karniadakis, Systems biology informed deep learning for inferring parameters and hidden dynamics, *PLoS Comput. Biol.* 16 (11) (2020) e1007575.
- [45] Ting Ye, Dingyi Pan, Can Huang, Moubin Liu, Smoothed particle hydrodynamics (SPH) for complex fluid flows: recent developments in methodology and applications, *Phys. Fluids* 31 (1) (2019) 011301.