RESEARCH

Machine learning and domain decomposition methods - a survey

Axel Klawonn[1,2]*, Martin Lanser[1,2]* and Janine Weber[1,2]*

*Correspondence:
axel.klawonn@uni-koeln.de;
martin.lanser@uni-koeln.de;
janine.weber@uni-koeln.de

[1] Department of Mathematics
and Computer Science,
University of Cologne, Weyertal
86-90, Köln 50931, Germany
[2] Center for Data and Simulation
Science, University of Cologne,
Albertus-Magnus-Platz,
Köln 50923, Germany

Abstract

Hybrid algorithms, which combine black-box machine learning methods with experience from traditional numerical methods and domain expertise from diverse application areas, are progressively gaining importance in scientific machine learning and various industrial domains, especially in computational science and engineering. In the present survey, several promising avenues of research will be examined which focus on the combination of machine learning (ML) and domain decomposition methods (DDMs). The aim of this survey is to provide an overview of existing work within this field and to structure it into domain decomposition for machine learning and machine learning-enhanced domain decomposition, including: domain decomposition for classical machine learning, domain decomposition to accelerate the training of physics-aware neural networks, machine learning to enhance the convergence properties or computational efficiency of DDMs, and machine learning as a discretization method in a DDM for the solution of PDEs. In each of these fields, we summarize existing work and key advances within a common framework and, finally, discuss ongoing challenges and opportunities for future research.

Keywords: Scientific machine learning, Domain decomposition, Survey, Hybrid modelling, Physics-aware neural networks

AMS subject classifications: 65F10, 65N22, 65N55, 68T05, 68T07

1 Introduction

Domain decomposition methods (DDMs) are divide-and-conquer strategies which decompose a given problem into a number of smaller subproblems. This strategy often leads to easier parallelizable algorithms where the subproblems can be solved on different processors (cpu or gpu). Sometimes, an additional problem is needed for parallel scalability, the so-called global problem, or to gather and connect local information obtained from the solution of the local subproblems. DDMs can be useful in different cases, for example, if the original problem is too large to be solved in the available memory, the computing time is too long on a single processor, or the decomposed problem has other preferable properties, for example, being better conditioned, yielding more accurate results, or boosting generalization properties and mitigating the spectral bias, respectively, of machine learning models. DDMs have a long and successful history in constructing parallel scalable preconditioners for linear and nonlinear systems obtained

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 2 of 31

from discretized partial differential equations. Recently, there has been an increased interest in using DDMs in combination with machine learning algorithms, especially in the field of scientific machine learning (SciML).

In recent years, the rapidly growing research area of SciML has drawn increasing interest and attention in various fields of applications, as, for example, in computational fluid mechanics [20, 29], optics and electromagnetic applications [9, 63], and earthquake detections [102]. Originally, the term SciML was introduced and made publicly known by the report [1] prepared for the U.S. Department of Energy. The core idea of SciML is to combine algorithms from supervised or unsupervised machine learning with other domain-specific methods or knowledge to develop new, hybrid methods. In particular, the authors of [1] have identified six priority research directions in SciML which are, among others, domain-awareness, interpretability, and robustness of the considered method. With respect to practical applications, this means that a key factor of SciML methods is to connect "black-box" approaches such as deep learning with well-understood and theoretical approaches from science and expert knowledge.

One framework that fits very well to the priority research directions of SciML are physics-informed neural networks (PINNs) [93]. PINNs are a specific machine learning technique usually used to solve forward or inverse problems involving Partial Differential Equations (PDEs). Similar to other deep learning models, PINNs approximate PDE solutions by training a neural network such that a given loss function is minimized. However, the characteristic feature of PINNs is that knowledge with respect to the underlying physics behind the data is integrated into the loss function as well as, possibly, initial or boundary conditions of the computational domain's boundary. For a detailed overview of existing research and applications around PINNs as well as remaining challenges and open research questions, we refer to [12].

One drawback of PINNs and also other neural network architectures, in general, is the large computational complexity of the related optimization problem for large problem domains and multi-scale problems as well as the resulting long training times. Hence, a wide range of attempts have been made to distribute and parallelize the training of different network architectures; see [6, 90, 112] for a comprehensive overview. In general, the wide range of techniques used in parallel and distributed deep learning can roughly be categorized into *data parallelism*, that is, partitioning the input samples, *model parallelism*, that is, horizontally partitioning the network structure, and *pipelining*, that is, vertically partitioning the network structure and hence, by layer; see also ([6], Sect. 6). From an abstract perspective, model parallelism can be and often is interpreted as a form of DDMs [91, 111].

In this paper, our aim is to review existing work on combining DDMs with different machine learning models. In the following, we focus on model parallelism and pipelining when using domain decomposition in machine learning algorithms. Let us note that, in a wider sense, also data parallelism can be interpreted as a form of domain decomposition such that the global data space is decomposed into smaller subspaces operating only on parts of the data. However, simple data parallel training methods are not the focus of this survey. The remainder of the paper is organized as follows. In Section 2, we provide a general description of DDMs for the solution of PDEs and introduce some necessary notation. Subsequently, in Section 3, we briefly summarize the main idea

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 3 of 31

and central algorithmic formulae of multilayer perceptrons, PINNs, and the Deep Ritz method. Note that both, Sections 2 and 3, can be skipped by the experienced reader who is already familiar with DDMs and PINNs and who can directly proceed with Section 4. In the present paper, we categorize all cited work into two main classes:

i)Methods using domain decomposition within machine learning models (Section 4)
ii)Machine learning-enhanced DDMs (Section 5).

Additionally, we classify all methods in group i) into two subgroups. First, we consider all approaches using domain decomposition for the acceleration and parallel training of PINNs; see Section 4.1. Second, in Section 4.2, we consider approaches using domain decomposition in machine learning algorithms other than PINNs, in particular, classical supervised and unsupervised algorithms. With respect to work within class ii), one can also subdivide the approaches into two subclasses; see also [38]. In this case, the first class consists of methods where machine learning is used to improve the convergence properties or the computational efficiency within classical DDMs. This is typically done by learning or approximating optimal interface conditions or by learning other optimal parameters; see Section 5.1. In the second class, different types of neural networks are used as discretization methods and replace classical local subdomain or coarse solvers based on finite elements or finite differences; see Section 5.2. For a schematic overview of the described categorization and the structure of this paper, see Fig. 1.

Note that for some approaches which combine domain decomposition with machine learning the categorization is not uniquely defined. For example, D3M [69] and Deep-DDM (Deep Domain Decomposition Method) [75] can be interpreted as replacing subdomain solvers within a DDM by neural networks but, at the same time, also as a DDM within neural network training. Hence, we have made the following classifications to the best of our knowledge and with regard to the subject focus of the concrete considered work.
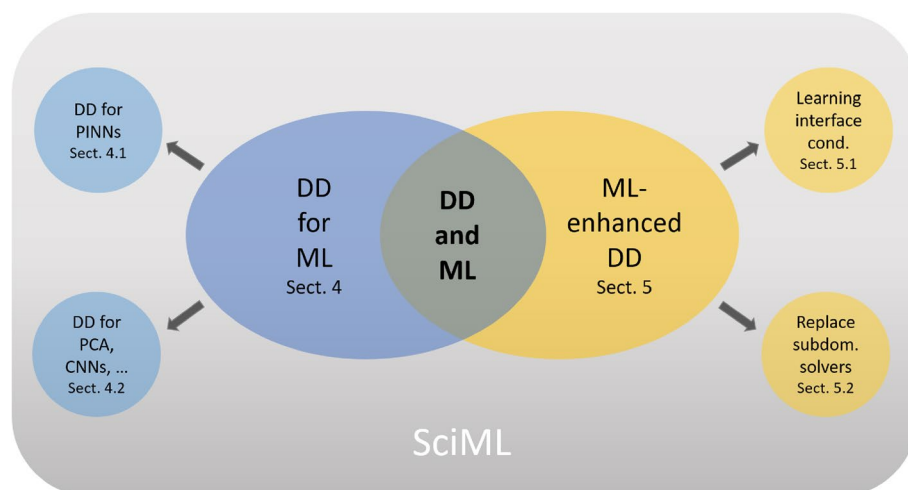
Fig. 1 Schematic representation of the structure of this review paper. All reviewed work on the combination of machine learning and domain decomposition methods are clustered into the two main groups domain decomposition (DD) for machine learning (ML), Section 4, and machine learning-enhanced domain decomposition, Section 5

2 Domain decomposition methods for the solution of PDEs

In this section, we briefly introduce the main principles of overlapping and nonoverlapping DDMs as well as some necessary notation.

DDMs [91, 111] are highly scalable, iterative methods for the solution of sparse, linear systems of equations as, for example, arising from the discretization of PDEs. DDMs have been shown to be numerically stable for many practically relevant problems and are designed for the application on parallel computers. Generally speaking, they rely on a divide-and-conquer strategy and decompose the original global problem into a number of smaller subproblems. Mathematically, this corresponds to a subdivision of the domain into a number of smaller subdomains, where the local problems can be solved in parallel on different processors of a parallel computer. Additionally, a small global problem is often needed for numerical and parallel scalability. Certain mathematical conditions at the interface of neighboring subdomains have to be satisfied such that the solution of the original problem is recovered.

For a generic description of DDMs, we consider a boundary value problem of the general form

$$\begin{aligned}
\mathcal{L}(u) &= f \quad \text{in} \quad \Omega \\
\mathcal{B}(u) &= g \quad \text{on} \quad \partial\Omega
\end{aligned} \tag{1}$$

on the domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, where $\mathcal{L}$ is a linear, second-order, elliptic differential operator and $\mathcal{B}$ represents the boundary conditions.

In order to compute a numerical solution of boundary value problem (1), we discretize the variational formulation of the given linear PDE with an appropriate numerical method, for example, conforming finite elements. This results in a, usually sparse, linear system of equations

$$K_g u_g = f_g. \tag{2}$$

The basic idea of DDMs is then, instead of directly solving the completely assembled system of Eq. (2), to decompose the problem into smaller subproblems and thus, solving smaller systems of equations which are assembled on local parts of the domain exclusively. In the following, we denote these local parts of the domain as subdomains and we assume to have a decomposition of Ω into $N \in \mathbb{N}$ nonoverlapping subdomains Ω_i, $i = 1, ..., N$, which fulfills $\overline{\Omega} = \bigcup_{i=1}^{N} \overline{\Omega}_i$; see also Fig. 2. Moreover, we define $\Gamma := \left(\bigcup_{i=1}^{N} \partial\Omega_i \right) \setminus \partial\Omega$ as the interface between neighboring subdomains, that is, degrees of freedom shared by at least two subdomains. In order to define overlapping DDMs, we additionally introduce overlapping subdomains Ω_i', $i = 1, ..., N$, where the overlap δ is a measure for the amount of local information shared between neighboring subdomains; see also Fig. 2 (left) for a visualization.

However, simply solving decoupled local subdomain systems independently from each other might result in discontinuities along the interface Γ and will not result in the correct solution of the global (original) system. Hence, in order to obtain the correct global (original) solution which is continuous across the different subdomains, also communication between the different subproblems is necessary, that is, a global coupling between the different subdomains has to be ensured. With respect on how the global exchange of information across the original domain Ω is ensured, we distinguish between one- and two-level
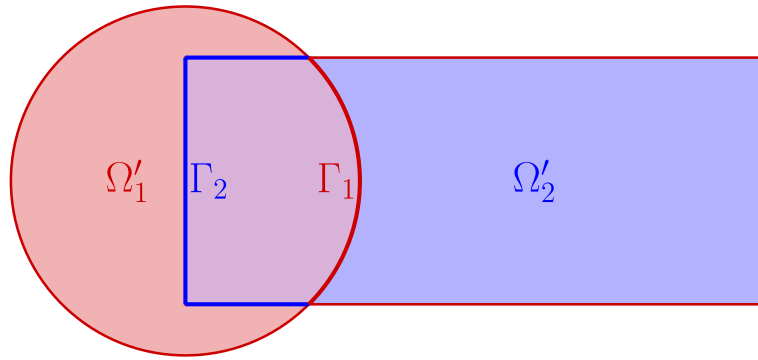
Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 5 of 31
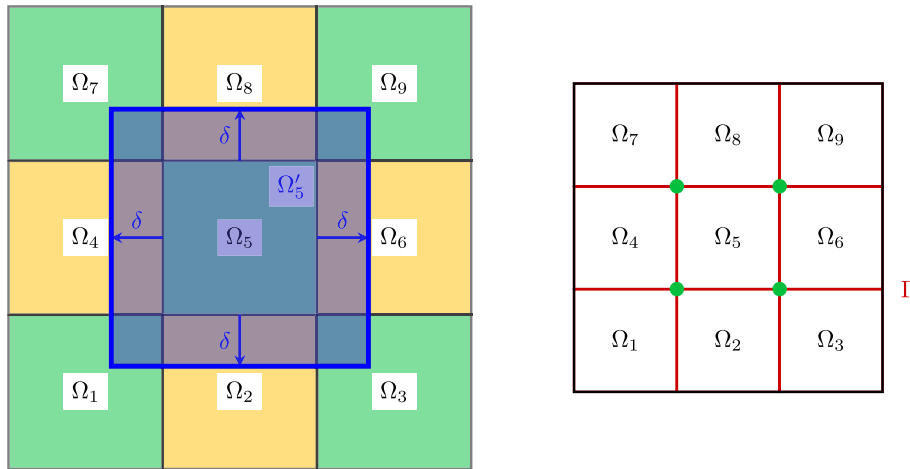
Fig. 2 Decomposition of a domain $\Omega \subset \mathbb{R}^2$ into nine nonoverlapping subdomains Ω_i, $i = 1, \dots, 9$. **Left:** The overlapping domain decomposition is obtained by extending the nonoverlapping subdomains by the width δ. The resulting overlapping subdomain Ω'_5 is marked in blue. **Right:** The interface Γ is marked in red

DDMs. One-level DDMs, where information is only exchanged between neighboring subdomains, result in a rate of convergence which deteriorates with a growing number of subdomains when used for the approximate solution of elliptic systems of PDEs; see, for example, [111]. A remedy is obtained using two-level DDMs by the setup and the solution of an additional, small global problem which needs to be solved in each iteration of the iterative solver. In particular, in two-level DDMs, the additional global problem ensures the fast global transport of information between the different subdomains as well as the scalability and robustness of the iterative method. In the context of two-level DDMs, we refer to this global problem as *coarse problem* and to the related solution space as *coarse space*. Let us note that the idea of a coarse problem is also transferred to domain decomposition approaches for, e.g., training PINNs; cf. Section 4.1. In classic DDMs, a coarse space is often defined by certain weighted averages or translations and rotations along the interface between neighboring subdomains. Additionally, also more sophisticated coarse spaces exist which are, for example, obtained by solving certain eigenvalue problems and which are problem-dependent or, in other words, adaptive. However, a detailed overview of existing classic and adaptive coarse spaces in DDMs would be beyond the scope of this review article.

To provide a better insight into how DDMs work, we now describe the classical alternating Schwarz approach [95, 103] as an example for overlapping DDMs in details. Although the method is not restricted to it, we consider a decomposition into only two overlapping subdomains as shown in Fig. 3. A restriction of the boundary value problem (1) to the subdomains results in the two local problems

$$
\begin{aligned}
\mathcal{L}(u_i) &= f && \text{in} && \Omega'_i \\
\mathcal{B}(u_i) &= g_i && \text{on} && \partial\Omega'_i \setminus \Gamma_i \\
u_i &= u|_{\Gamma_i} && \text{on} && \Gamma_i
\end{aligned}
\tag{3}
$$

with $i = 1, 2$. Here, clearly $g_i = g$, $i = 1, 2$, on the boundary $\partial\Omega \cap \partial\Omega'_i$, $i = 1, 2$, but especially the choice of the boundary condition on Γ_i is crucial. For the choice in (3)

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 6 of 31

Fig. 3 Decomposition of the domain Ω into two overlapping subdomains Ω_1 and Ω_2. The boundary of the overlapping region is partitioned into Γ_1 and Γ_2 accordingly

Loop over $k = 1, 2, 3, \dots$ until convergence

Solve local problem 1:

$$\mathcal{L}(u_1^{(k)}) = f \text{ in } \Omega_1'$$

$$\mathcal{B}(u_1^{(k)}) = g \text{ on } \partial\Omega_1' \setminus \Gamma_1$$

$$u_1^{(k)} = u_2^{(k-1)}\Big|_{\Gamma_1} \text{ on } \Gamma_1$$

Solve local problem 2:

$$\mathcal{L}(u_2^{(k)}) = f \text{ in } \Omega_2'$$

$$\mathcal{B}(u_2^{(k)}) = g \text{ on } \partial\Omega_2' \setminus \Gamma_2$$

$$u_2^{(k)} = u_1^{(k)}\Big|_{\Gamma_2} \text{ on } \Gamma_2$$

EndLoop

Fig. 4 Classical alternating Schwarz algorithm to iteratively solve the boundary value problem (1)

where u is the global solution of (1), the solutions u_1 and u_2 of the two local problems (3) will be the restrictions of u to the subdomains. Unfortunately, u is not known a priori and the correct boundary condition on Γ_i has to be computed in an iterative process. Therefore, in the classical Schwarz approach, both local problems are solved alternatingly and the information between both subdomains is exchanged via Γ_1 and Γ_2 approximating $u|_{\Gamma_i}$ in (3) by restricting the recent solution on the other subdomain to Γ_i. Initializing $u_2^{(0)}$, for example, with zero or randomly, the alternating classical Schwarz approach can be written as a simple iterative process shown in Fig. 4. Let us remark that the local solution process can be chosen arbitrarily in this algorithm, that is, analytically, using a numerical discretization with finite elements or finite differences, or using a machine learning based discretization, as, for example, a PINN approach.

Let us note that in the field of solution methods for partial differential equations a large variety of DDMs, for example, overlapping, nonoverlapping, one-level, two-level, algebraic, and non-algebraic ones, exist and potentially all can be combined

with various ML approaches. To learn more about solution methods based on DDMs we recommend [103], which is still a good overview.

3 Multilayer perceptrons, PINNs, and Deep Ritz

In this section, we briefly provide some preliminaries with respect to multilayer perceptrons, including a short outlook on convolutional neural networks and residual neural networks, see Section 3.1, as well as PINNs and the Deep Ritz method; see Section 3.2. For more details on deep learning and more specialized network architectures, we refer to, for example, [10, 27, 36]. Let us note that for experienced readers, who are already familiar with multilayer perceptrons, PINNs, and the Deep Ritz method, it might be more convenient to skip the following section and proceed directly with Section 4.

3.1 Feedforward multilayer perceptrons

In general, neural networks or multilayer perceptrons, respectively, are used within supervised learning to approximate a functional relation between given input and output data. Usually, the overall aim is to minimize a loss function which measures the difference between the ground truth of the given data and the output, that is, estimated classification or regression values, of the neural network. The following description of multilayer perceptrons is loosely based on [37, 97, 114]. Generally, a feedforward neural network can be interpreted as a directed, weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a set of nodes $\mathcal{V}$, a set of edges $\mathcal{E}$, and a weight function $w : \mathcal{E} \to \mathbb{R}$; see, for example, ([97], Chapt. 20.1). An exemplary visualization of the graph of a dense feedforward neural network is presented in Fig. 5. A neural network is assumed to be organized in layers, that is, the set of nodes $\mathcal{V}$ can be represented as the union of nonempty, disjoint subsets $\mathcal{V}_i \subset \mathcal{V}, i = 0, \ldots, N + 1$. For feedforward neural networks, these sets are defined such that for each edge $e \in \mathcal{E}$ there exists an $i \in \{0, \ldots, N\}$ with e being an edge between a
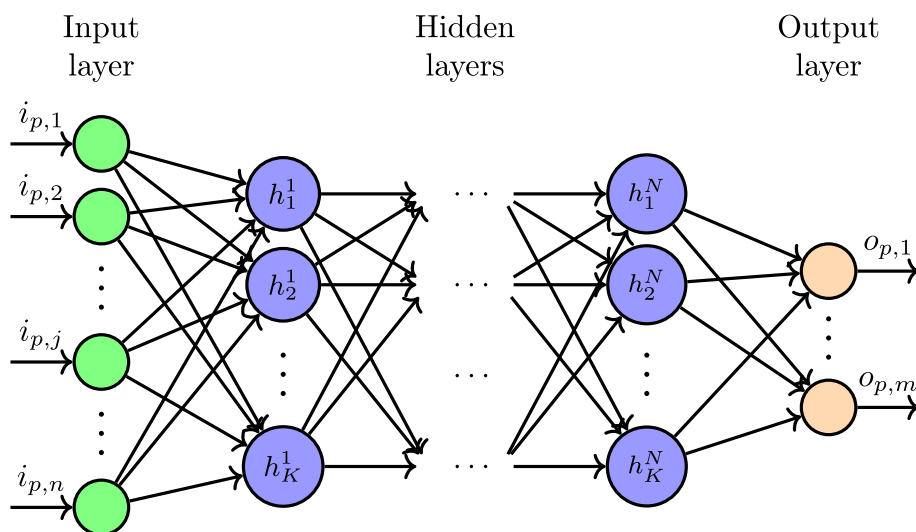
Fig. 5 Structure of a dense feedforward neural network with *n* input nodes (marked in green), *m* output nodes (marked in orange), and *N* hidden layers with *K* neurons per layer (marked in blue). Figure taken from ([114], Fig. 7.1)

node in $\mathcal{V}_i$ and one in $\mathcal{V}_{i+1}$; see ([97], Chapt. 20.1). In general, different layers can perform different transformations on their input. A neural network usually starts with an input layer (marked in green in Fig. 5), where the features $\{i_p\}_{p=1}^P$ of the external data are used as input data, and concludes with an output layer (marked in orange in Fig. 5). We denote the layers in between the input and the output layer as hidden layers and the nodes in a neural network as neurons.

Mathematically, the relation between two consecutive layers is the conjunction of a linear mapping defined by the weight function w and a nonlinear activation function α. In particular, the nonlinearity of the activation function enables a neural network to approximate highly complex relations between the input and output data.

Generally, for a generic activation function α, the output of the k-th layer of the neural network can be written as

$$y = \alpha^k \left(x, W^k, b^k \right),$$

where $W^k = (w_{ij}^k)_{i,j}$ and b^k are the weight matrix and the bias vector, respectively. Note that an entry w_{ij}^k of W^k corresponds to the value of the weight function w associated with the corresponding edge between layer $\mathcal{V}_{k-1}$ and $\mathcal{V}_k$. Then, the application of a complete neural network with N hidden layers to an input vector $i \in I$ is given by

$$\begin{aligned}
h^1 &= \alpha^1 \left(i, W^1, b^1 \right), \\
h^{k+1} &= \alpha^{k+1} \left(h^k, W^{k+1}, b^{k+1} \right), \quad 1 \le k < N, \\
o &= \left(W^{N+1} \right)^T h^N + b^{N+1}
\end{aligned} \tag{4}$$

where h^k is the output of the k-th hidden layer and $o \in O$ is the (final) output vector. The computation of the output vector o is usually performed without an additional application of the activation function. Note that for specific neural network architectures, the consecutive relation between two layers can be more complicated than written in (4). In Section 4.2, we consider also DDMs for the training of convolutional neural networks (CNNs) and residual neural networks (ResNets). In CNNs, convolutions, that is, kernel functions, and pooling functions are gradually applied to input data with a grid-like structure such that the information within the data is aggregated and condensed throughout the network layers. For more mathematical details on CNNs, we refer to, for example, [64], ([27], Chapt. 9), and ([10], Chapt. 5). For ResNets, skip connections that perform identity mappings are added between distant blocks of hidden layers of a multilayer perceptron or a CNN. Hence, the connected block of hidden layers approximates the residual function with respect to the input data which can help to mitigate the degradation problem of very deep neural networks; see [43] for further details.

3.2 PINNs and Deep Ritz

The key concept of PINNs is to integrate domain specific knowledge in form of physical laws or domain expertise modeled by ordinary or partial differential equations into neural networks. In particular, this is done by enhancing the loss function with the residual error of a certain differential equation; see [93] for more details. Hence, the objective is

to obtain solutions which do not only fit some given training data but also satisfy a given ordinary or partial differential equation in a least square sense. Let us note that, in principle, also other loss functions can be used to minimize the residual loss, for example, the L_1-loss function. The following descriptions are loosely based on ([28], Sect. 4).

For the mathematical derivation of a physics-enhanced loss function, we consider again the general boundary value problem in (1) on the domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ with a linear, second-order, elliptic differential operator $\mathcal{L}$ and boundary conditions $\mathcal{B}$. We now aim to define a PINN which solves (1), that is, we aim for a feedforward neural network $\mathcal{N}(\cdot, W, b)$ that fulfills (1) in a least square sense with weights W and biases b. As input data of the neural network, we use a set of collocation points located inside the domain Ω as well as a set on the boundary $\partial\Omega$. In order to obtain a neural network $\mathcal{N}$ solving the boundary value problem (1), we enhance the classic loss function by a point-wise error of the residual of the PDE. Thus, the physics-informed loss function is defined as

$$
\begin{aligned}
\mathcal{M}(W, b) \quad &:= \mathcal{M}_\Omega(W, b) + \mathcal{M}_{\partial\Omega}(W, b) \\
\mathcal{M}_\Omega(W, b) \quad &:= \frac{1}{N_f} \sum_{i=1}^{N_f} |\mathcal{L}\left(\mathcal{N}\left(x_f^i, W, b\right)\right) - f\left(x_f^i\right)|^2 \\
\mathcal{M}_{\partial\Omega}(W, b) &:= \frac{1}{N_g} \sum_{i=1}^{N_g} |\mathcal{B}\left(\mathcal{N}\left(x_g^i, W, b\right)\right) - g\left(x_g^i\right)|^2,
\end{aligned}
\tag{5}
$$

with collocation points x_f^i, $i = 1, ..., N_f$, located in the domain Ω and collocation points x_g^i, $i = 1, ..., N_g$, located on the boundary $\partial\Omega$. During the training of the neural network, the derivatives occurring in the operators $\mathcal{L}$ and $\mathcal{B}$ are evaluated using the backward propagation algorithm and automatic differentiation [4]. Taking advantage of this, the training of the neural network then consists of solving the optimization problem

$$
\{W^*, b^*\} := \arg \min_{\{W, b\}} \mathcal{M}(W, b)
\tag{6}
$$

using a stochastic gradient approach based on mini-batches built from all collocation points. An exemplary PINN is shown in Fig. 6. Note that the loss term $\mathcal{M}_\Omega(W, b)$ enforces the PINN to satisfy the condition $\mathcal{L}(u) = f$ in a least square sense, while $\mathcal{M}_{\partial\Omega}(W, b)$ enforces the boundary condition.

An approach that is very closely related to PINNs is the Deep Ritz method [22] which uses dense neural networks (DNNs) for solving PDEs in variational form. Here, we will only give a very short overview of the Deep Ritz method. For a more detailed description, we refer the reader to the original publication [22] and to the overview paper ([38], Sect. 3.2); in the latter reference, a more detailed description in the context of domain decomposition methods is given. The Deep Ritz method is based on the Ritz approach, where a given PDE as in (1) is reformulated as an equivalent minimization problem as in (7) which is then discretized and solved by a DNN in combination with a numerical integration method. For a formal description of the Deep Ritz method, we assume that (1) is self-adjoint in the following. Then, solving the PDE is equivalent to the solution of the minimization problem

$$
\min_u \mathcal{E}(u) \quad \text{s.t.} \quad \mathcal{B}(u) = g \quad \text{on} \quad \partial\Omega
\tag{7}
$$
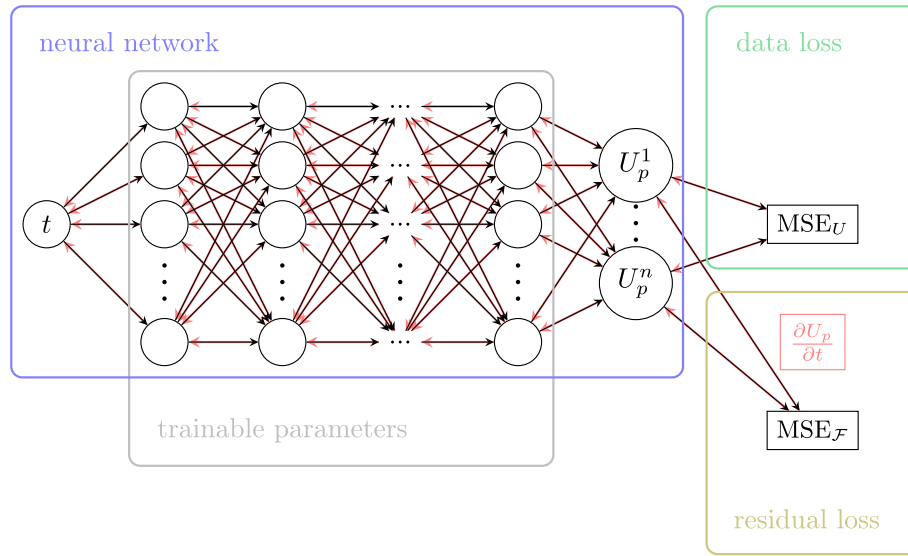
Fig. 6 Graphical representation of a PINN. The derivatives needed for the evaluation of the residual are computed using automatic differentiation in a backward propagation (red). Figure in modified form in ([28], Fig. 3)

where $\mathcal{E}(u)$ is an appropriate energy functional. The basic idea of the Deep Ritz method is now to approximate the function u in (7) by a DNN and to use a numerical quadrature rule where each integration point in Ω is used as a collocation point to approximate the minimizing functional $\mathcal{E}(u)$. This yields the discrete loss function for the Deep Ritz method

$$\widetilde{\mathcal{M}}(W, b) := \widetilde{\mathcal{M}}_{\Omega}(W, b) + \widetilde{\mathcal{M}}_{\partial\Omega}(W, b) \quad \text{with}$$

$$\widetilde{\mathcal{M}}_{\Omega}(W, b) = \frac{1}{\widetilde{N}_f} \sum_{i=1}^{\widetilde{N}_f} h\left(\widetilde{x}_f^i, W, b\right)$$

$$\widetilde{\mathcal{M}}_{\partial\Omega}(W, b) = q \frac{1}{\widetilde{N}_g} \sum_{i=1}^{\widetilde{N}_g} |\mathcal{B}\left(\widetilde{\mathcal{N}}\left(\widetilde{x}_g^i, W, b\right)\right) - g\left(\widetilde{x}_g^i\right)|^2,$$

(8)

where $h(\cdot)$ is the inner function of the energy functional $\mathcal{E}(u)$ approximated by the DNN $\widetilde{\mathcal{N}}$ and q is a Lagrange multiplier. The derivatives of the neural network $\widetilde{\mathcal{N}}$ occuring in (8) are then evaluated using the backward propagation algorithm and automatic differentiation [4]. For a detailed derivation of the loss function (8) for the exemplary case of $\mathcal{L}(u) = \Delta u$, we refer to ([38], Sect. 3.2).

4 Domain decomposition for machine learning

In this section, we review recent works where ideas from domain decomposition are used to accelerate and parallelize the training of different machine learning models, and, in particular, deep learning models. As described in Section 1, the wide range of techniques used in parallel and distributed deep learning can roughly be categorized into *data parallelism*, *model parallelism*, and *pipelining*; see also ([6], Sect. 6). Let us note again that, within this paper, we focus on model parallel training methods and pipelining

rather than on data parallelism; see also Section 1. In particular, the interpretation of model parallelism as a DDM is, from our point of view, more straightforward and also allows for analogous concepts in distributed ML corresponding to a coarse space in DDMs.

In the following, we will distinguish the different methods using DDMs in ML into two classes. First, we consider all approaches using domain decomposition for the acceleration and parallel training of very specific neural networks, that is, physics-informed neural networks, in Section 4.1. For details on PINNs, we refer to Section 3.2. Second, in Section 4.2, we provide an overview of approaches using domain decomposition in other ML algorithms than PINNs, in particular, classical supervised and unsupervised algorithms. Let us note that, in general, the introduction of DDMs in PINNs and other supervised and unsupervised machine learning methods can possibly create an overhead in terms of memory and communication requirements. The size of the overhead depends strongly on the chosen DDM. For example, within the context of PINNs, the overhead depends on the decision whether an overlapping (see, e.g., [87, 105]) or a nonoverlapping (see, e.g., [49, 50]) decomposition of the collocation points is chosen.

4.1 Domain decomposition to accelerate the training of physics-aware neural networks

In general, integrating ideas from domain decomposition into PINNs usually aims at accelerating and parallelizing the training of PINNs, mainly by reducing the complexity of the PINN optimization problem for large problem domains and multi-scale problems. Roughly speaking, this means that instead of solving one large optimization problem, the idea is to use a divide-and-conquer strategy by solving, that is, training smaller local problems in parallel. This helps to parallelize the training and, potentially, can also help to reduce the complexity of the global optimization problem. Analogously to DDMs, the resulting related key question within these parallel training approaches is how to ensure that the different local problems communicate with each other such that the local solutions match across the interface. In the following, we review different parallel training methods for different types of PINNs. We will observe that some approaches deal with the communication between different local problems by integrating additional interface terms in the PINN loss function, whereas other approaches define overlapping subdomains and directly exchange information among the overlapping subdomain areas. In Table 1, a structured overview of all papers reviewed in the following can be found.

4.1.1 Physics-aware neural networks based on the strong form of the PDE

In [50], conservative PINNs (cPINNs) are introduced that make use of a nonoverlapping domain decomposition approach for conservation laws. Here, the spatial computational domain is decomposed into nonoverlapping subdomains and for each subdomain, a separate local PINN, which also enforces certain interface conditions with its neighboring subdomains with respect to the solution and the local PDE residuals, is trained. In [50], the authors present numerical results for different scalar nonlinear and systems of conservation laws.

The idea of cPINNs is further extended in [49] to generic PDEs and arbitrary space-time domains resulting in extended PINNs (XPINNs). Similar as in [50], local sub-PINNs are trained for local subdomains of a domain independently from each other and

Table 1 Summary of the current state of domain decomposition methods to accelerate the training of PINNs as reviewed in Section 4.1. We categorize all papers according to the defined domain decomposition, the considered PINN formulation, and the concrete method of how to integrate interface constraints between neighboring subdomains. We mark the papers according to the following color code: dark green: based on a nonoverlapping domain decomposition, blue: based on an overlapping domain decomposition, light green: an adaptive domain decomposition automatically trained by one or several neural networks

DD for PINNs	DD	PINN formulation	Interface conditions
cPINN [50]	nonoverlapping	conservation law	conservative quantity
XPINN [46, 49]	nonoverlapping	residuals	averaging
PINN for BCM [86]	nonoverlapping	residuals	averaging
GatedPINN [106]	adaptive	residuals	weighted average
MoE-PINN [7]	adaptive	residuals	weighted average
APINN [47]	adaptive	residuals	weighted average
DPINN [18]	nonoverlapping	residuals	flux interface loss
PECANN [3]	nonoverlapping	residuals	generalized Robin-type
Mosaic Flow [24, 113]	overlapping	residuals	alterating Schwarz-like
PINN-PINN, PINN-FOM [105]	overlapping	residuals	loss term/strong enforc.
UFV-PINN [84]	nonoverlapping	residuals	boundary network
Iterative algorithms [118]	overlapping	residuals	weighted average
Partitioned PINN [51]	nonoverlapping	residuals	via Lagrange mult.
FBPINN [15, 87]	overlapping	residuals	Schwarz-like
stacking FBPINN [38]	overlapping	residuals	Schwarz-like
FBKAN [44]	overlapping	residuals	Schwarz-like
NS-DDPINN [31]	nonoverlapping	residuals	flux interface loss
Nonlinear preconditioning [62]	nonoverlapping	residuals	additive/multiplicative
ASM for PINNs [66]	nonoverlapping	residuals	additive
hp-VPINNs [53]	nonoverlapping	var. residuals	via variational loss
TgNN-wf [53]	nonoverlapping	var. residuals	via variational loss
PFNN-2 [100]	overlapping	var. residuals	via boundary network
DFR [110]	overlapping	var. residuals	weighted average

in parallel. In particular, in XPINNs, the domain can be decomposed in both, space and time and no explicit assumptions of a specific PDE type are made. The approximated solutions of the local sub-PINNs are combined by enforcing continuity across the interface by averaging local solutions across neighboring subdomain interfaces as well as by enforcing a residual continuity condition. In ([13], Sect. 3.2), the authors propose to additionally enforce continuity of first-order derivatives between different sub-PINNs to further reduce the training error close to the interface.

First steps regarding a direct comparison of PINNs and XPINNs with respect to generalization properties are presented in [46]. Specifically, the authors derive an a priori generalization bound based on the complexity of the considered PDE problem as well as an a posteriori generalization bound based on the matrix norm of the neural network after optimization and use this theory to derive observations with respect to the generalization properties of XPINNs. On the one hand, the decomposition of the global problem into smaller subproblems in XPINNs decreases the complexity of each subproblem and hence can boost generalization. On the other hand, each subproblem is trained with a smaller amount of training data which tends to increase overfitting of the local models. In [46], different experimental results are presented which validate that the two factors in the derived generalization bounds lead to a tradeoff when comparing the performance of XPINNs to PINNs.

In [101], a unified parallel algorithm is presented for cPINNs and XPINNs which employs domain decomposition in space and space-time, respectively. The distributed

framework in [101] for the specific PINN extensions from [49, 50] is constructed by a hybrid programming model using MPI [30, 104]. Within the unified framework, the authors compare the parallel performance of cPINNs and XPINNs for different forward problems. The experimental results indicate that cPINNs are more effective regarding communication cost whereas XPINNs provide greater flexibility such that they can also handle space-time domain decompositions for any differential equations as well as deal with arbitrarily shaped subdomains.

The algorithmic concept of XPINNs is transferred to a multi-block Cartesian mesh computational fluid dynamics (CFD) framework, the Building Cube Method (BCM) [88], in [86]. Here, the decomposition of the domain into subdomains corresponds to the decomposition of a multi-block domain into several smaller sub-blocks. The information exchange between neighboring subdomains is realized in analogy to XPINN [49], that is, by explicit information exchange of the network parameters along the interface and by enforcing a residual continuity condition. The authors present results for incompressible Navier-Stokes equations on a domain consisting of a sphere of unit diameter and a box as the outer boundary. A schematic representation of different DDMs for the training of different PINNS, such as cPINNs, XPINNs, and DPINNs is given in Fig. 7.

The authors of [106] propose to automatically learn a domain decomposition of a PINN by incorporating conditional computing into the PINN framework in order to learn an arbitrary decomposition of the neural network, which is adaptively tuned during the training process. Here, conditional computing means that only certain neurons or units of a neural network are activated depending on the network input; see also [5, 98] for more details. Based on the idea of conditional computing the authors introduce GatedPINNs which rely on a number of *experts* which decompose the neural network and are each modeled by a simple neural network themselves. The outputs of the different experts are then aggregated by a linear or nonlinear *gate network*. The authors
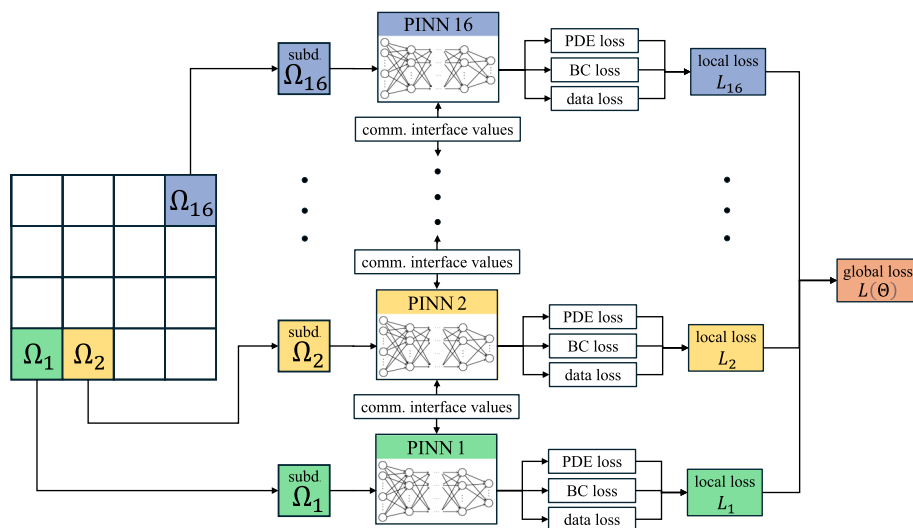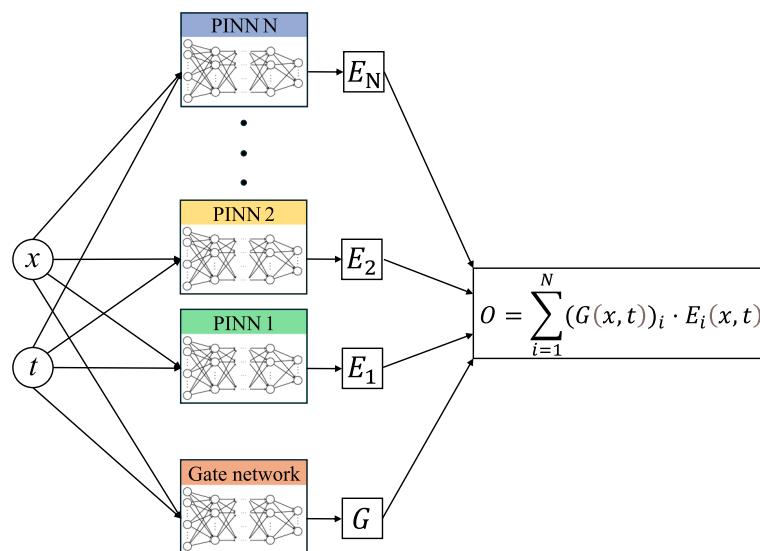

Fig. 7 Schematic representation of different DDMs for the training of PINNs, e.g., cPINNs [50], XPINNs [49], and DPINNs [18]. Note that, for simplicity, here the decomposition into subdomains is nonoverlapping but it could be as well overlapping

provide comparative experimental results for the GatedPINN and standard PINN approach and are able to reduce the training time significantly.

In [7], the concept of GatedPINNs is further extended by adding an additional regularization term to the loss function to define Mixture of Experts Physics-Informed Neural Networks (MoE-PINNs). In particular, the implemented regularization can serve as a data-driven selection of the amounts of different experts, that is, local PINNs. The authors of [7] present experimental results for the Poisson equation on L-shaped domains and observe, for this test problem, improved accuracy in comparison to Gated-PINNs without regularization when a small number of experts is trained.

A related approach was recently introduced in [47], where also a *gate network* is trained to find an augmented domain decomposition and flexibel parameter sharing. Concretely, the gate network does not need any specific interface losses and as output data, provides a weight average of several local subnets. Additionally, it uses a flexible partial parameter sharing of the subnets and initializes the subnets with the hard decomposition obtained from XPINN, resulting in an augmented PINN (APINN). The authors of [47] provide results for different PDEs comparing APINN to XPINN and classic PINNs showing that APINNs result in an improved performance and better generalization properties. A schematic representation of different mixture of expert models for PINNs, such as, APINNs, GatedPINNs, and MoE-PINNs is given in Fig. 8.

In [18], the authors use ideas from finite volume methods where the computational domain is partitioned into multiple cells to define a distributed PINN (DPINN) for the data-efficient solution of partial differential equations. Here, the authors decompose the computational domain into nonoverlapping cells, which can also be interpreted as the subdomains in a DDM, and train a separate, local PINN for each cell with a loss function based exclusively on collocation points from the interior of the cell. Each of these local PINNs is relatively simple and, for the results presented in [18], only two layers deep.

Fig. 8 Schematic representation of different mixture of expert models for PINNs, e.g., APINNs [47], GatedPINNs [106], and MoE-PINNs [7]. Here, $E_i(x, t)$, $i = 1, \ldots, N$ represents the expert output of the i-th PINN and $(G(x, t))_i$, $i = 1, \ldots, N$ the corresponding weighting coefficient obtained from the gate network

In addition to the PINN losses for each cell, a loss term for the interface conditions is introduced which is associated with collocation points located on the boundary of the cells. This interface loss is inspired by the flux conditions of the finite volume method. Then, the DPINN approach is based on training all local PINNs together by minimizing the sum of all local losses plus the interface loss. In particular, the described approach requires the exchange of interface information in each step of the optimization approach used in the training of the network.

A related approach which mostly differs by the treatment of the boundary conditions between neighboring subdomains is presented in [3]. There, the authors use an optimized Schwarz-type nonoverlapping domain decomposition for solving forward and inverse PDE problems using PINNs. Within each subdomain, physics and equality constrained artificial neural networks (PECANNs) [2] are trained in parallel as solvers for each subdomain. Moreover, generalized Robin-type interface transmission conditions are considered as an additional constraint on the solution of each subdomain to ensure continuity. Hence, local information are exchanged at the end of each local training within each outer iteration of the described method. The authors of [3] show that by incorporating the Robin-type interface constraints into the iterative learning process, the effectiveness and reliability of the method can be enhanced.

In [24], an end-to-end parallelization of the physics-informed neural PDE solver Mosaic Flow [113] based on domain decomposition is presented. Roughly speaking, Mosaic Flow is an iterative algorithm inspired by the alternating Schwarz method designed for solving PDEs with arbitrary boundary conditions on diverse domains. It basically consists of two steps. First, local physics-informed PDE solvers (SDNets) are pretrained on small domains with arbitrary boundary conditions, using the discretized boundary function as input data; see also [113] for more mathematical details. Then, in a second step, the Mosaic Flow predictor is trained which decomposes a large domain into overlapping subdomains and uses the pretrained SDNets as subdomain solvers. In [24], the authors combine data parallel training and domain parallelism for inference to develop a distributed domain decomposition algorithm for Mosaic Flow to achieve strong scalability for its parallel training on GPUs. Numerical results for the parallel solution of the Laplace equation on large domains, distributed on 32 GPUs, are presented.

A related approach which is also inspired by the alternating Schwarz method is presented in [105]. In [105], the authors use the alternating Schwarz method and an overlapping domain decomposition to couple local PINNs and full order models, that is, local solutions obtained via the finite element, finite difference, or finite volume method with each other to accelerate the training of PINNs. Furthermore, three different approaches to enforce Dirichlet boundary conditions within the local subdomain PINNs are empirically compared. Concretely, the three different approaches are the following: A weak boundary condition enforcement through the loss function, a strong enforcement through a solution transformation, and a combination of both such that the boundary conditions are enforced strongly on the computational domain's boundary and weakly on the subdomains' boundaries. Numerical examples are provided for the one-dimensional steady state advection-diffusion equation with focus on the advection-dominated regime (high Péclet number). The presented

results show that the convergence of the Schwarz method is strongly linked to the chosen boundary condition implementation within the local, coupled PINNs.

In [84], a new PINN is proposed which unifies ideas from domain decomposition, the finite volume method, and physics-aware machine learning resulting in the unified finite volume PINN (UFV-PINN). In UFV-PINNs, the domain is decomposed into a number of nonoverlapping subdomains. For each of these subdomains, the solution is obtained by a finite volume method, that is, for each finite volume cell by a linear equation solver, as well as by a direct PINN prediction for each finite volume cell. Additionally, a separate neural network is trained to learn the boundary conditions on the boundary of a selected number of subdomains. The overall loss function of the UFV-PINN is then defined as the discrepancy between the neural network prediction and the finite volume solution summed over all subdomains.

A different iterative approach to parallelize the training of PINNs while also reducing the communication cost is presented in [118]. In [118], the authors propose an iteration method which is based on the classical additive Schwarz method. Hence, similarly as in [69, 75], an outer training loop for the global network is defined such that, in each iteration of the outer training loop, decoupled smaller local problems are solved which are then stitched back together by a weighted average. In each outer iteration, the local problems obtain their boundary values from the current global iteration and an overlapping decomposition of the collocation points into subdomains is used. To compute the next iterate of the global problem, a weighted sum of the previous iterate and the partitioned neural network solution using a relaxation parameter is computed. Additionally, to further improve the convergence properties of the described approach, the authors of [118] propose a two-level extension by solving an additional coarse problem with specific structure in each outer iteration. Note that related preliminary work has also been published in [54, 117] by the authors. In [54], first experimental results for the proposed additive Schwarz algorithm have been presented whereas [117] contains first results regarding the convergence analysis of the approach.

In [51], nonoverlapping domain decomposition is introduced for the training of PINNs and combined with Lagrange multipliers. By using the decomposition of the domain into nonoverlapping subdomains, the PINN loss function is reformulated into a sum of localized loss functions. Each of these local loss functions is enriched with an augmented Lagrangian term where the interface and boundary conditions are enforced as constraints with respect to the local solutions by using Lagrange multipliers. Furthermore, similarly as in [118], an iterative training algorithm for the decomposed PINN model is presented to enhance the training efficiency under the parallel computing environment.

In [87], the authors transfer ideas from classical finite element methods where the solution of a PDE or an ODE is expressed as a sum of a finite set of basis functions with a compact support to PINNs. The resulting method uses local window functions to form a global function, that is, the global solution on the entire domain, and is referred to as Finite Basis Physics-Informed Neural Network (FBPINN). The global function is hence expressed as a sum of the localized window functions which are learned by local neural networks. In particular, the local window or basis functions are defined over small, overlapping subdomains such that, in FBPINNs, no additional interface condition is

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 17 of 31

necessary. In [87], the authors present numerical results for both small and large, multi-scale problems.

The idea of FBPINNs is further developed in [15, 16]. In [15], additive, multiplicative, and hybrid iteration methods based on a Schwarz-like domain decomposition method for the training of FBPINNs are introduced. In [16], these ideas are extended even further by adding multiple levels of domain decompositions to the solution ansatz, similar as in classical multilevel Schwarz methods. This results in improved convergence properties and increased accuracies in the considered PDE solutions and, at the same time, mitigates convergence problems related to the spectral bias of neural networks.

In [39], the concepts of FBPINNs is combined with stacking neural networks resulting in stacking FBPINNs. Within the approach of stacked PINNs [45], multiple neural networks are stacked on top of each other with the aim that the networks on top of the stack learn features that are not learned by the previous networks. This concept is combined with FBPINNs where the learning of multiscale features is improved by localization. The authors apply the proposed stacking FBPINNs to different multiscale problems, that is, a pendulum, a two-frequency problem, and the Allen-Cahn equation where the authors observe more accurate solutions than for stacking PINNs without a domain decomposition in time.

In [44], the idea of FBPINNs is transferred to Kolmogorov-Arnold networks (KANs) [73], resulting in finite basis KANs (FBKANs). KANs can be seen as an alternative to classic multilayer perceptrons and use trainable activation functions represented by splines. In FBKANs, in principle, the neural networks used in FBPINNs are replaced by local KANs. Due to the enhanced flexibility of KANs compared to classic multilayer perceptrons, the authors of [44] observe improved accuracy results compared to KANs without a domain decomposition-based localization approach.

A comparison of different domain decomposition approaches for PINNs is provided in [33] for the Navier-Stokes equation (NS-DDPINN). The authors compare different strategies to handle the interface constraints, for example, residual and flux continuity. Besides, in [33], the authors propose two novel strategies for continuity conditions at the interface which are a dynamic weight method and a neural architecture incorporating attention mechanisms.

The performance and convergence properties of L-BFGS (Limited-memory Broy-den-Fletcher-Goldfarb-Shanno) [72] optimizers used for the training of PINNs are enhanced in [62] by using nonlinear preconditioning strategies based on Schwarz-like domain decomposition approaches. More specifically, nonlinear additive and multiplicative preconditioners are constructed by decomposing the network model, that is, the corresponding parameters in a layer-wise manner. This is, in principle, different from the aforementioned approaches where ideas from DDMs are used to decompose the computational domain. The authors present results for both additive and multiplicative nonlinear preconditioners for different benchmark problems, indicating that the proposed layer-wise decomposition of the network can significantly accelerate the convergence of the L-BFGS optimizer.

The ideas of [72] are further extended in [66], where two-level overlapping additive Schwarz preconditioners are used for the training of both, PINNs and operator learning approaches [76]. Similar as in [66], an overlapping layer-wise decomposition of DNNs

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 18 of 31

is introduced where the subdomain corrections are combined using weighted average techniques as in the additive Schwarz method (ASM). Moreover, in [66], an additional coarse level training step is defined which enables global communication between the subdomains and can help to yield more accurate DNN models.

4.1.2 Physics-aware neural networks based on the variational formulation of the PDE

In [53], the authors introduce *hp*-variational PINNs (*hp*-VPINNs) which are based on the variational formulation of the residuals of the considered PDE, similar to the approach considered in [69]. In particular, the authors use piecewise polynomial test functions within the variational formulation of the considered PDE such that each test function v_j is always a polynomial of a chosen order over the subdomain j and zero otherwise. Let us note that with respect to the implementation of this approach, the authors employ a single DNN to approximate the solution over the whole computational domain despite virtually decomposing the domain into several subdomains; see also ([53], Remark 2.1). Due to the use of a single DNN no subdomain interface conditions have to be considered in the loss function. However, this makes it more difficult to parallelize *hp*-VPINNs. Different numerical examples of function approximation for continuous and discontinuous functions as well as for solving the Poisson equation are presented.

An approach which is closely related to [53] and which is also based on the weak, that is, the variational form of a given system of PDEs is presented in [116]. The idea of the presented neural networks is to integrate the weak formulation of the PDE in the loss function as well as data constraints and initial or boundary conditions. Due to the theory guided structure of the networks based on the weak formulation of the considered PDE the authors refer to their approach by the acronym TgNN-wf. In particular, similar to [53], the authors use locally defined test functions to perform a domain decomposition of the computational domain to accelerate the training process. As a further novelty, the authors formulate the original loss minimization problem into a Lagrangian duality problem in order to optimize the penalty term parameters of the different components of the loss function within the training process. The authors provide comparative results for the strong form TgNN and TgNN-wf for an unsteady-state 2D single-phase flow problem and a 1D two-phase flow problem in terms of accuracy, training time, and robustness with respect to noise in the training data.

In [100], the authors use an overlapping domain decomposition approach for the training process of penalty-free neural networks (PFNN), originally introduced in [99]. Applying the ideas introduced in [99] for PFNN, the resulting new PFNN-2 adopts two neural networks, with one network learning the initial condition and essential boundary conditions of a given PDE problem, and a second network approximating the solution of the PDE in its variational form on the remaining part of the domain. To speedup the learning process, the authors introduce an overlapping decomposition of the domain into subdomains and train two local networks, that is, one for the initial and boundary conditions and a second one for the local interior solution, for each subdomain. This approach requires direct exchange of information between neighboring subdomains to enforce continuity via the locally trained networks which learn the boundary conditions of each subdomain.

In [110], the authors use ideas from overlapping domain decomposition and adaptive mesh refinements for the Deep Fourier Residual (DFR) method, a specific type of PINNs based on the variational formulation. In particular, in [110], the test space rather than the trial space is decomposed by partitioning the domain Ω as an overlapping union of n-rectangles where for each of these local n-rectangles, local refinements are used based on local error estimators. Due to the decomposition of the test space, in the DFR method, no iteration between subdomains or transmission conditions are necessary. The presented approach is also related to [53] but also differs by the implementation of spectral test functions.

4.2 Domain decomposition for classical supervised and unsupervised machine learning algorithms

Apart from the wide application area of DDMs for PINNs, the idea of domain decomposition has also been transferred to other machine learning algorithms of which we try to give a broad overview in this section. Note that the idea of DDMs is, on the one hand, applied to the training of different supervised machine learning models, for example, convolutional or residual neural networks, and, on the other hand, also to unsupervised algorithms, as, for example, principle component analysis (PCA); see Fig. 9 for a schematic overview.

4.2.1 Domain decomposition for different types of neural networks

In [31, 34], the idea of DDMs is transferred to the training of CNNs for different image recognition problems. The authors propose to decompose a global CNN along its width into a finite number of smaller, local subnetworks which can be trained independently from each other. In particular, the local CNNs are trained in parallel without any
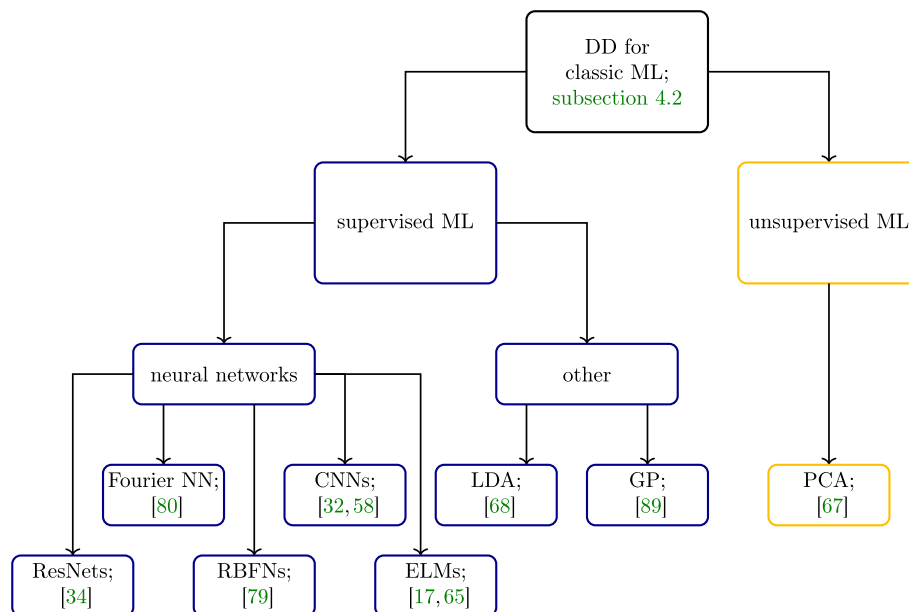
Fig. 9 Schematic representation of the articles reviewed in Section 4.2 with respect to domain decomposition methods in machine learning

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 20 of 31

communication between the different processes. The obtained weights of the local subnetworks are then used as an initialization for the subsequently trained global network by using a transfer learning strategy. The approach can be interpreted as an analogy to nonlinear preconditioning [77] and the decomposition of the inputs of the global CNN along its width corresponds to a decomposition of the input images into smaller partial images. The authors present results for different two-dimensional image data indicating that the suggested idea can significantly accelerate the training of the global network on a GPU cluster.

A slightly similar but different parallel training approach also for CNNs in image recognition has been proposed in [58]. In [58], the authors decompose two- or three-dimensional image data into nonoverlapping or overlapping subsets and train smaller CNNs operating exclusively on parts of the decomposed input data. Based on the decomposed image data, all local CNNs provide a probability distribution with respect to the different classes of the underlying image classification problem. Then, in a second step, a global coarse DNN is trained which evaluates the local decisions into a final, global decision. In particular, this approach does not require the training of a global, large CNN operating on the entire images and the smaller, local CNNs can be trained in parallel on different GPU clusters. The results provided for two- and three-dimensional image data show a significant acceleration of the training and additionally, provide at least the same classification accuracies compared to the training of a single, global CNN for almost all tested datasets.

Another model parallel training approach based on domain decomposition, that is, a layer-parallel training approach has been suggested in [32] for ResNets [43]. In this work, the authors make use of the fact that the forward propagation through a ResNet can be interpreted as a forward Euler discretization of a time-dependent ordinary differential equation (ODE) [19] where the time-dependent control variables represent the weights of the neural network. Hence, the time domain in the time-dependent ODE corresponds to the layer domain of a ResNet. Based on this interpretation, the authors employ a multigrid reduction in time approach [23] to replace the classic forward and backward propagation in the training progress by a parallel nonlinear multigrid iteration applied to the layer domain such that chunks of layers can be processed in parallel. In [32], scalability and speedup results for this method are presented for three different image recognition problems.

In [80], domain decomposition techniques are used to define a parallel training process of trigonometric neural networks which use trigonometric activation functions. Based on a nonoverlapping domain decomposition, three local neural networks are trained for each subdomain. The first neural network, denoted as primary neural network is completely local, the second one, denoted as boundary neural network, learns for each subdomain the interface conditions with its neighbors, and the third one, denoted as modifier neural network, then learns the update for the primary network including the interface conditions. Finally, the local solution is defined as the sum of the primary and modifier network. This process is repeated for each subdomain until convergence.

Another work on the combination of DD and a special type of neural networks is presented in [79]. The authors combine different types of radial basis function networks (RBFNs) [96], which typically use only one hidden layer, with the concept of domain

decomposition to approximate nonlinear functions or to solve Poisson's equation on a rectangular domain. The computational domain is decomposed into nonoverlapping subdomains and each of the subdomains is discretized by a shallow RBFN. Based on this decomposition, the authors propose an iterative solution algorithm where, in each iteration, all local subproblems are solved using the RBFNs. Then, in each iteration of the outer training loop, the interface condition between the subdomains is estimated and updated using boundary integral equations.

In [17], the idea of extreme learning machines (ELMs) [48] is combined with domain decomposition and local neural networks. Basically, in order to compute the solution for a given PDE, the authors of [17] divide the domain into nonoverlapping subdomains and represent the solution on each of these subdomains by a local feedforward neural network. To obtain a global solution, continuity constraints are imposed along the subdomain boundaries. Originating from the idea of ELMs, each local neural network consists of a small number of hidden layers, while its last hidden layer is typically wide. As in ELMs, the weights and bias coefficients in all hidden layers of the local neural networks are pre-set to random values and, during training, only the weight coefficients of the output layers of the local neural networks are training parameters. Moreover, also in analogy to ELMs, the overall neural network is trained by a linear or nonlinear least squares computation instead of using the backpropagation algorithm. The authors compare the presented method with the deep Galerkin method (DGM), the PINN method, and classical FEM.

A related approach using domain decomposition for ELMs is presented in [65]. In [65], the authors also define a decomposition of the domain into nonoverlapping subdomains and train local ELMs for each of the subdomains. However, in contrast to [17], an explicit system for the auxiliary variable u_Γ on the interface and the parameters of the output layers of the local ELMs is constructed. This system is then reduced to a Schur complement system by eliminating the parameters of the local ELMs such that a parallel computation of the reduced local systems is enabled.

4.2.2 Domain decomposition for Gaussian Processes and linear discriminant analysis

In [89], the idea of domain decomposition is not transferred to a special type of neural network but, instead, to Gaussian Process (GP) regression [115]. Given that the computational complexity $O(N^3)$, where N is the number of training samples, is often a major limitation in GP regression, the main idea of the authors is to decompose the optimization problem related to the GP regression into smaller local optimization problems that provide local predictions. This is in analogy of decomposing a domain into smaller subdomains. To deal with the possible mismatch problem of the local predictions along the boundaries of neighboring subdomains, the authors of [89] impose continuity constraints, analogously to interface conditions in DDMs.

In [68], DDMs are applied to linear discriminant analysis (LDA) [25], a supervised classification method based on a low-dimensional projection of the data with maximum separability. The authors of [68] apply LDA to different image classification problems (CIFAR-10/CIFAR-100) and a numeric gene expression dataset for cancer patients. The LDA approach is combined with a DDM such that the original dataset is split into subsets with a smaller number of classes and a local LDA is applied separately for each of

theses subsets in the training phase. In the test phase, the test data is projected into multiple subspaces instead of into one lower dimensional subspace as for the classic LDA. The final classification for each test data point is then obtained by the computation of a scaled distance vector computed for all subspaces and choosing the class with the minimum distance vector. Additionally, in [68], the authors provide an optimality theory stating that the decomposed LDA can provide better classification accuracy compared to the classic LDA under a technical assumption that all subspaces have the same desired reduced dimension.

4.2.3 Domain decomposition for unsupervised machine learning

In [67], the authors propose a local PCA approach which also relies on the idea of DDMs. The PCA is applied to different images which are decomposed into nonoverlapping smaller subimages. Here, instead of computing the principal components for the entire images, only principal components for the different subimages are determined. Then, in a second step, the local principal components are used to compute low-dimensional projections of the different subimages which are composed again to a lower dimensional representation of the original global image. The authors show empirically that the decomposed PCA requires a smaller number of principal components to provide an accurate low-dimensional representation of the original image compared to the global PCA. Moreover, the authors provide a mathematical proof that the local PCA reduces the summation pollution effect of the global PCA and thus, results in a lower or equal reconstruction error compared to the global PCA. In other words, the local PCA is never worse than the global PCA; see [67] for more details.
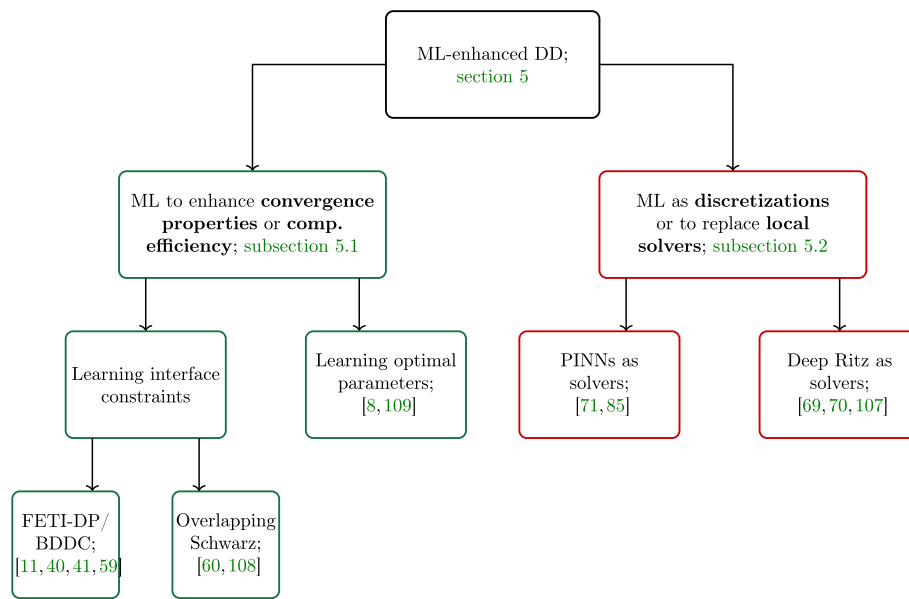
5 Machine learning-enhanced domain decomposition methods

In this section, we review recent work which uses different, mainly supervised, machine learning algorithms to enhance different DDMs. In general, one can divide the approaches using ML to enhance DDMs into two classes; see also [38]. The first class consists of methods where ML is used to improve the convergence properties or the computational efficiency within classical DDMs. Typically, this is done by learning or approximating optimal interface conditions or by learning other optimal parameters. Existing approaches of this class are considered in Section 5.1. In the second class, different types of neural networks are used as discretization methods and replace classical local subdomain or coarse solvers based on finite elements or finite differences. Existing methods of this class are summarized in Section 5.2. For a schematic overview of the existing literature, see Fig. 10.

5.1 Machine learning to enhance convergence properties or computational efficiency of domain decomposition methods

As explained in Section 2, introducing appropriate global communication and interface conditions between subdomains in a DDM is a crucial task to ensure the correct and globally continuous solution as well as a satisfactory convergence behavior. Hence, a number of different attempts have been made to automatically construct such interface conditions, for example, by using regression neural networks.

Fig. 10 Schematic representation of the articles reviewed in Section 5 with respect to machine learning-enhanced domain decomposition methods

5.1.1 Machine learning for learning or predicting interface constraints

In [11], the authors propose an approach to learn an adaptive coarse space for BDDC (Balancing Domain Decomposition by Constraints) algorithms for stochastic elliptic PDEs. The authors focus on a specific adaptive coarse space [55, 56] where local generalized eigenvalue problems have to be solved for parts of the interface between neighboring subdomains. In particular, these eigenvalue problems depend on the (stochastic) coefficients of the PDE. Hence, the authors train a dense feedforward neural network to approximate the relation between the stochastic PDE coefficients and the adaptive coarse space. As an input for the neural networks, a truncation of the Karhunen-Loève expansion [120] is used. Numerical results are presented for different oscillatory and high contrast coefficients, a fixed discretization of the domain $\Omega \subset \mathbb{R}^2$, and a neural network with a single hidden layer. Note that, so far, this method is not independent of the discretization and thus has to be retrained for different mesh resolutions.

As a preparatory step to automatically construct an adaptive coarse space for DDMs, in [37, 41], the authors present a supervised classification approach to identify parts of the interface where adaptive coarse constraints are necessary. Then, adaptive constraints resulting from the local generalized eigenvalue problem in [81, 82] are integrated into a FETI-DP (Finite Element Tearing and Interconnecting - Dual Primal) coarse space exclusively for edges or faces, respectively, which are categorized as critical by a pretrained neural network. In particular, the decision of the neural network is based on a sampling procedure of the coefficient function of the underlying elliptic PDE and is independent of the finite element discretization. Hence, the trained neural network can be evaluated for different finite element meshes as used in the training procedure and can be applied to both, two- and three-dimensional problems. Additionally, in [42], the described approach is also analogously applied to the AGDSW (Adaptive Generalized Dryja-Smith-Widlund) method [40] in two spatial dimensions.

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 24 of 31

In [57], the work described above has been extended to directly learn the adaptive constraints to setup a robust FETI-DP coarse space in two spatial dimensions. Here, the authors train a fixed number of regression neural networks to directly predict an approximation of adaptive edge constraints. Again, the input data used for the neural networks rely on a mesh-independent sampling strategy such that the trained neural networks can be applied to different finite element mesh resolutions. Numerical results are provided for different elliptic PDE problems and realistic coefficient distributions orginating from a microsection of a dual-phase steel. In [59], these results are further extended to irregular decompositions obtained by the graph partitioning software METIS [52].

In order to provide a robust and efficient iterative solution method for the time-harmonic Maxwell's equations, in [60], the authors use a feedforward neural network to approximate the interface operator in an optimized Schwarz method [14]. The concrete formulation of the optimized Schwarz method for the time-harmonic Maxwell's equations is based on [21]. In [60], preliminary numerical results for a two-dimensional domain decomposed into two subdomains are provided. Additionally, comparative results for different network architectures and different activation functions are presented and compared with classical DDMs.

A related approach to learn optimal interface conditions also for an optimized Schwarz method [26] using Graph Convolutional Neural Networks (GNNs) is suggested in [109]. There, the authors aim to learn the subdomain interface matrices for both, structured grids where the optimal matrix values can be predicted by Fourier analysis and for unstructured grids where the optimal Schwarz parameters are generally not known. Hence, the GNNs are trained using as input the degrees of freedom of a discretized problem, its decomposition, and sparsity constraints on the subdomain interface matrices. As output, the networks predict approximated optimized values for these matrices. Additionally, the authors consider an improved loss function which is a relaxation of the ideal Frobenius norm minimization.

5.1.2 Machine learning to learn optimal parameters

In [108], a new multigrid GNN (MG-GNN) architecture to learn optimized parameters in two-level DDMs is introduced. In particular, the new MG-GNN model is used to learn the Robin-type subdomain boundary conditions in a two-level optimized Schwarz method and the overall coarse-to-fine interpolator. In a way, the work in [108] can thus be seen as an extension of the method in [109], where also GNNs are used to learn optimal interface conditions in a one-level optimized Schwarz method.

A slightly different approach, where not adaptive constraints themselves but interface conditions in form of varying overlap between neighboring subdomains are learned is presented in [8]. More precisely, a method is suggested to optimize the width δ of the overlap in Schwarz methods such that the number of expected floating point operations until convergence is minimized. The authors compare four different regression algorithms for two-dimensional diffusion problems with jumps in the coefficient function. As input for the different regression algorithms, the authors use certain features collected for each subdomain as well as its neighborhood, as, for example, the maximal or minimal coefficient within certain sets of rows and columns of degrees of freedom in the surrounding of the boundary of the overlapping subdomain.

5.2 Replacing the subdomain solvers in domain decomposition methods by neural networks

In addition to learning optimal interface conditions to improve the robustness of DDMs, a different approach that has attracted much attention very recently is to replace the subdomain solvers by machine learning algorithms as, for example, different types of neural networks. As a starting point for this section, we consider again boundary value problem (1) and, in the following, review some works which aim to solve the local subdomain problems by training a neural network $\mathcal{N}$ with certain collocation points and the loss function given in (5) or (8), respectively.

In [69, 75], the subdomain solvers in a parallel overlapping Schwarz method are replaced by PINNs and the Deep Ritz method, respectively, such that each of these local networks solves a single subdomain problem. The latter method is named D3M in [69] and the former method DeepDDM in [75]. As explained above, in any DDM, it is crucial to exchange information between the local subdomains in order to obtain a continuous, global solution. In DeepDDM [75] as well as in D3M [69], which are both based on a parallel overlapping Schwarz fixed point iteration, the exchange of information is enforced by additional boundary conditions, which change in each fixed-point iteration until convergence; see [71] or [111] for further details on the parallel overlapping Schwarz method. These additional boundary conditions result in a third loss term of the local subdomain networks with additional collocation points for the exchange of information between neighboring subdomains.

In [85], the authors extend the DeepDDM algorithm [75] described above by a coarse space correction, similarly to what is done in traditional two-level DDMs [111]. The aim is to improve the convergence properties of the resulting algorithm for an increasing number of subdomains due to the enhanced exchange of global information. To define a coarse problem, the authors of [85] introduce an additional coarse sampling of the complete domain defined by a set of coarse collocation points. These coarse collocation points are then used as input data for an additional fully connected network that is trained subsequently after the local networks in each outer training iteration. The authors present numerical results for different elliptic boundary value problems indicating the advanced numerical scalability for an increasing number of subdomains compared to the original DeepDDM approach.

A different extension of the D3M approach [69] is presented in [74]. Here, the aim is to mitigate the spectral bias [78, 92, 94] observed with the local networks in the original D3M approach by introducing new multi Fourier feature networks (MFFNet) in each local subdomain. This results in a new Fourier feature based deep DDM (F-D3M) which is more successful in approximating the high frequency modes of a PDE solution. In particular, in order to obtain satisfactory approximations for both, low and high frequency modes, the MFFNet consists of different subnetworks with different random frequencies of the Fourier features and whose outputs are concatenated with a linear layer to compute the final output of the MFFNet.

In [107], the subdomain solvers within a DDM are also replaced by a Deep Ritz approach but, in contrast to [69], the computational domain is decomposed into non-overlapping subdomains. In contrast to replacing the subdomain solvers by neural networks within an overlapping Schwarz method as done in D3M or DeepDDM, in its

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 26 of 31

nonoverlapping counterpart a direct flux exchange across subdomain interfaces is required. In particular, simply replacing the subdomain solvers in nonoverlapping subdomains by the classic Deep Ritz approach often results in a local minimizer where the given Dirichlet boundary conditions are satisfied satisfactory but with inaccurate Neumann boundary conditions; see [107]. In order to address the issue of the low accuracy of the related flux transmission between neighboring nonoverlapping subdomains, the authors of [107] propose a compensated Deep Ritz method as subdomain solvers that enable reliable flux transmission in the presence of erroneous interface conditions. The compensated Deep Ritz method defines a weighted sum of the Ritz form of the loss function and the PINN residual loss given that PINNs are empirically found to provide more accurate estimates of partial derivatives. The authors provide numerical examples for elliptic problems and two and four subdomains.

6 Discussion and future work

In this paper, we have explored how established techniques and ideas from DDMs can effectively be combined with machine learning algorithms to develop major advances in SciML methods. In particular, we have categorized all reviewed work in the two main classes of DDMs for machine learning and machine learning-enhanced DDMs. By taking all cited work into account, we observe that, so far, higher efforts exist to implement ideas from DDMs into machine learning models than vice versa. However, in both areas, several ongoing challenges and opportunities for further developments and research exist.

In the field of DDMs for machine learning, a very large group of publications exists that use DDMs within physics-aware neural networks. Here, we have observed a strong focus of using DDMs as model or data parallelism to efficiently accelerate and distribute the training of physics-aware neural networks, or, in particular, PINNs. Furthermore, decomposing the training of one large global neural network into training several smaller neural networks in parallel can also have important impacts on the generalization properties of the underlying machine learning model. As investigated analytically in [46], decomposing the global optimization problem into smaller subproblems decreases the complexity of each subproblem which can improve the generalization of the resulting global model. On the other hand, each subproblem is trained with a smaller amount of training data which tends to increase overfitting of the local models. Thus, a tradeoff between both effects has to be considered. Additionally, training decoupled local PINN models in a preliminary phase which are subsequently aggregated into a global model can also help to mitigate the spectral bias often observed for deep neural networks.

Besides, concepts from DDMs have also been integrated into classical machine learning models, for example, CNNs for image recognition, ResNets, Gaussian Processing, or PCA. Here, the primary objective is also to distribute the training of the respective machine learning model to different GPUs in order to accelerate the required training time. As a secondary effect, also for DDMs in classical machine learning models, in some cases, an improvement of the accuracy of the trained model can be observed. For some models, a mathematical comparison of the original global model versus the decomposed model already exists, for example, [31, 67], whereas for other cases, a mathematical proof of a corresponding correlation is still an open research question.

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 27 of 31

Within the second reviewed field of machine-learning enhanced DDMs, a range of work exists on learning interface conditions or restriction operators by regression neural networks. So far, all of these approaches are developed for concrete domain decomposition methods and test problems, for example, for FETI-DP and BDDC or optimized Schwarz methods for stationary diffusion problems, stochastic PDEs, or Maxwell's equations, respectively. Here, a possible priority for future work could be the aim to develop more universal approaches which can be applied to various problems and DDMs or are independent of the underlying finite element discretization. Currently, only some of the cited approaches partly fulfill such requirements. Additionally, some work on learning optimal parameters in DDMs exist, as, for example, learning the optimal width of overlap for overlapping subdomains, optimal stopping criterions, or optimal restriction operators. These efforts, considering the current state, are also restricted to very concrete cases where specific information with respect to the considered DDM and test problem are used as input data for different regression models. Thus, in this area, more generally valid approaches are also a topic of potential future research.

Finally, there are several efforts to use PINNs or Deep Ritz methods to replace the subdomain solvers or the discretization within classical DDMs. These approaches make use of the universal approximation capabilities of neural networks to represent, under certain assumptions on the activation function, etc., solutions of PDEs. However, there are some challenges within these approaches in order to be competitive with classical numerical methods for the solution of PDE problems. Most notably, most existing approaches require to be retrained for varying boundary and initial conditions which makes the training of such approaches computationally expensive and limits the applicability of the resulting models. Hence, practically relevant applications for such models have to be carefully thought of, for example, large-scale multi-physics problems, and more flexible surrogate models will much likely be an important topic of future research.

Neural PDE solvers or, more general, neural operators [76] can offer a remedy for the described drawback that PINNs or Deep Ritz usually need to be retrained for varying boundary conditions. Neural operators are a type of model that is trained to approximate the PDE solution operator and to be evaluated for various instances of a boundary value problem with different boundary conditions. Practically, this means that the neural operator is trained with a representative set of discretized boundary functions as input data. With respect to the aim of applying neural operators to large, complex domains with arbitrary boundary conditions, some efforts have been made to combine neural operators with ideas from DDMs to enable the efficient solution of complex boundary value problems and a scalable training of these models; see, for example, [61, 70, 83, 119]. The different avenues of ongoing research, so far, still face several challenges, for example, with respect to applicability to irregular meshes, complex PDE solutions, and computationally expensive training. Hence, combining DDMs and other promising concepts as, for example, transformers with neural operators [35] to address these challenges is likely to be a future research topic of rising interest and importance.

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 28 of 31

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

References

1. Baker N, Alexander F, Bremer T, Hagberg A, Kevrekidis Y, Najm H, et al. Brochure on basic research needs for scientific machine learning: core technologies for artificial intelligence. United States: USDOE Office of Science (SC); 2018. https://doi.org/10.2172/1484362. https://www.osti.gov/biblio/1484362.
2. Basir S, Senocak I. Physics and equality constrained artificial neural networks: Application to forward and inverse problems with multi-fidelity data fusion. J Comput Phys. 2022;463:111301.
3. Basir S, Senocak I. A Generalized Schwarz-type Non-overlapping Domain Decomposition Method using Physics-constrained Neural Networks. 2023. arXiv:2307.12435.
4. Baydin AG, Pearlmutter BA, Radul AA, Siskind JM. Automatic differentiation in machine learning: a survey. J Mach Learn Res. 2017;18(1):5595–637.
5. Bengio E, Bacon PL, Pineau J, Precup D. Conditional computation in neural networks for faster models. 2015. arXiv:1511.06297.
6. Ben-Nun T, Hoefler T. Demystifying Parallel and Distributed Deep Learning: An In-depth Concurrency Analysis. ACM Comput Surv. 2020;52(4):1–43.
7. Bischof R, Kraus MA. Mixture-of-experts-ensemble meta-learning for physics-informed neural networks. In: Proceedings of 33. Forum Bauinformatik. 2022. https://mediatum.ub.tum.de/doc/1688403/uic8b0xn1c845e7rac1or092o.Bischof%20et%20Al.%202022.pdf.
8. Burrows S, Frochte J, Völske M, Torres ABM, Stein B. Learning Overlap Optimization for Domain Decomposition Methods. In: Advances in Knowledge Discovery and Data Mining. Springer; 2013. pp. 438–449.
9. Chen Y, Lu L, Karniadakis GE, Dal Negro L. Physics-informed neural networks for inverse problems in nano-optics and metamaterials. Opt Express. 2020;28(8):11618–33.
10. Chollet F. Deep learning with Python. Simon and Schuster; 2017.
11. Chung E, Kim HH, Lam MF, Zhao L. Learning Adaptive Coarse Spaces of BDDC Algorithms for Stochastic Elliptic Problems with Oscillatory and High Contrast Coefficients. Math Comput Appl. 2021;26(2):4. https://www.mdpi.com/2297-8747/26/2/44.
12. Cuomo S, Di Cola VS, Giampaolo F, Rozza G, Raissi M, Piccialli F. Scientific machine learning through physics-informed neural networks: Where we are and what's next. J Sci Comput. 2022;92(3):88.
13. De Ryck T, Jagtap AD, Mishra S. Error estimates for physics-informed neural networks approximating the Navier–Stokes equations. IMA J Numer Anal. 2023;44(1):83–119.
14. Dolean V, Gander MJ, Gerardo-Giorda L. Optimized Schwarz methods for Maxwell's equations. SIAM J Sci Comput. 2009;31(3):2193–213.
15. Dolean V, Heinlein A, Mishra S, Moseley B. Finite basis physics-informed neural networks as a Schwarz domain decomposition method. 2023. arXiv:2211.05560.
16. Dolean V, Heinlein A, Mishra S, Moseley B. Multilevel domain decomposition-based architectures for physics-informed neural networks. 2023. arXiv:2306.05486.
17. Dong S, Li Z. Local Extreme Learning Machines and Domain Decomposition for Solving Linear and Nonlinear Partial Differential Equations. Comput Methods Appl Mech Eng. 2021;387:114129.
18. Dwivedi V, Parashar N, Srinivasan B. Distributed physics informed neural network for data-efficient solution to partial differential equations. 2019. arXiv:1907.08967.
19. E W. A proposal on machine learning via dynamical systems. Commun Math Stat. 2017;1(5):1–11.
20. Eichinger M, Heinlein A, Klawonn A. Surrogate convolutional neural network models for steady computational fluid dynamics simulations. Electron Trans Numer Anal. 2022;56:235–55.
21. El Bouajaji M, Thierry B, Antoine X, Geuzaine C. A quasi-optimal domain decomposition algorithm for the time-harmonic Maxwell's equations. J Comput Phys. 2015;294:38–57.
22. E W, Yu B. The Deep Ritz method: a Deep Learning-Based Numerical Algorithm for Solving Variational Problems. Commun Math Stat. 2018;6(1):1–12.
23. Falgout RD, Friedhoff S, Kolev TV, MacLachlan SP, Schroder JB. Parallel time integration with multigrid. SIAM J Sci Comput. 2014;36(6):C635–61.
24. Feeney A, Li Z, Bostanabad R, Chandramowlishwaran A. Breaking Boundaries: Distributed Domain Decomposition with Scalable Physics-Informed Neural PDE Solvers. 2023. arXiv:2308.14258.
25. Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugenics. 1936;7(2):179–88.
26. Gander MJ, Halpern L, Nataf F, et al. Optimized Schwarz methods. In: Twelfth International Conference on Domain Decomposition Methods. Chiba: Citeseer; 2001. pp. 15–28.
27. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press Cambridge; 2016.
28. Grimm V, Heinlein A, Klawonn A, Lanser M, Weber J. Estimating the time-dependent contact rate of SIR and SEIR models in mathematical epidemiology using physics-informed neural networks. Electron Trans Numer Anal. 2022;56:1–27.

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 29 of 31

29. Grimm V, Heinlein A, Klawonn A. Learning the solution operator of two-dimensional incompressible Navier-Stokes equations using physics-aware convolutional neural networks. 2023. arXiv preprint arXiv:2308.02137.

30. Gropp WD, Huss-Lederman S, Lumsdaine A. MPI-The Complete Reference, Volume 2: The MPI Extensions. Cambridge: MIT Press; 1998.

31. Gu L, Zhang W, Liu J, Cai XC. Decomposition and composition of deep convolutional neural networks and training acceleration via sub-network transfer learning. Electron Trans Numer Anal. 2022;56:157–86.

32. Gunther S, Ruthotto L, Schroder JB, Cyr EC, Gauger NR. Layer-parallel training of deep residual neural networks. SIAM J Math Data Sci. 2020;2(1):1–23.

33. Gu L, Qin S, Xu L, Chen R. Physics-informed neural networks with domain decomposition for the incompressible Navier–Stokes equations. Phys Fluids. 2024;36(2).

34. Gu L, Zhang W, Liu J, Cai XC. Decomposition and Preconditioning of Deep Convolutional Neural Networks for Training Acceleration. In: Domain Decomposition Methods in Science and Engineering XXVI, vol. 145. Cham: Springer International Publishing; 2022. pp. 153–160. Lecture Notes in Computational Science and Engineering.

35. Hao Z, Wang Z, Su H, Ying C, Dong Y, Liu S, et al. GNOT: A general neural operator transformer for operator learning. In: International Conference on Machine Learning. PMLR; 2023. pp. 12556–12569.

36. Haykin S. Neural networks and learning machines. 3rd ed. Pearson Education India; 2010.

37. Heinlein A, Klawonn A, Lanser M, Weber J. Machine Learning in Adaptive Domain Decomposition Methods - Predicting the Geometric Location of Constraints. SIAM J Sci Comput. 2019;41(6):A3887–912.

38. Heinlein A, Klawonn A, Lanser M, Weber J. Combining Machine Learning and Domain Decomposition Methods for the Solution of Partial Differential Equations-A Review. GAMM-Mitt. 2021;44(1):e202100001. https://doi.org/10.1002/gamm.202100001.

39. Heinlein A, Howard AA, Beecroft D, Stinis P. Multifidelity domain decomposition-based physics-informed neural networks for time-dependent problems. 2024. arXiv preprint arXiv:2401.07888.

40. Heinlein A, Klawonn A, Knepper J, Rheinbach O. An Adaptive GDSW Coarse Space for Two-Level Overlapping Schwarz Methods in Two Dimensions. In: Domain Decomposition Methods in Science and Engineering XXIV, vol. 125 of LNCSE. Cham: Springer 2018. pp. 373–382.

41. Heinlein A, Klawonn A, Lanser M, Weber J. Combining Machine Learning and Adaptive Coarse Spaces - A Hybrid Approach for Robust FETI-DP Methods in Three Dimensions. SIAM J Sci Comput. 2021;43(5):S816–S838. Special Section Copper Mountain 2020.

42. Heinlein A, Klawonn A, Lanser M, Weber J. Predicting the geometric location of critical edges in adaptive GDSW overlapping domain decomposition methods using deep learning. In: Domain Decomposition Methods in Science and Engineering XXVI. Springer; 2023. pp. 307–315.

43. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: 2016. p. 770–8. https://doi.org/10.1109/CVPR.2016.90.

44. Howard AA, Jacob B, Murphy SH, Heinlein A, Stinis P. Finite basis Kolmogorov-Arnold networks: domain decomposition for data-driven and physics-informed problems. 2024. arXiv preprint arXiv:2406.19662.

45. Howard AA, Murphy SH, Ahmed SE, Stinis P. Stacked networks improve physics-informed training: applications to neural networks and deep operator networks. 2023. arXiv preprint arXiv:2311.06483.

46. Hu Z, Jagtap AD, Karniadakis GE, Kawaguchi K. When Do Extended Physics-Informed Neural Networks (XPINNs) Improve Generalization? SIAM J Sci Comput. 2022;44(5):A3158–82.

47. Hu Z, Jagtap AD, Karniadakis GE, Kawaguchi K. Augmented Physics-Informed Neural Networks (APINNs): A gating network-based soft domain decomposition methodology. Eng Appl Artif Intell. 2023;126:107183.

48. Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing. 2006;70(1–3):489–501.

49. Jagtap AD, Karniadakis GE. Extended Physics-Informed Neural Networks (XPINNs): A Generalized Space-Time Domain Decomposition Based Deep Learning Framework for Nonlinear Partial Differential Equations. Commun Comput Phys. 2020;28(5):2002–41.

50. Jagtap AD, Kharazmi E, Karniadakis GE. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. Comput Methods Appl Mech Eng. 2020;365:113028.

51. Jang DK, Kim K, Kim HH. Partitioned neural network approximation for partial differential equations enhanced with Lagrange multipliers and localized loss functions. 2023. arXiv preprint arXiv:2312.14370.

52. Karypis G, Kumar V. METIS: Unstructured Graph Partitioning and Sparse Matrix Ordering System, Version 4.0. 2009. http://www.cs.umn.edu/~metis.

53. Kharazmi E, Zhang Z, Karniadakis GE. hp-VPINNs: Variational Physics-Informed Neural Networks With Domain Decomposition. Comput Methods Appl Mech Eng. 2021;374:113547.

54. Kim HH, Yang HJ. Domain Decomposition Algorithms for Physics-Informed Neural Networks. In: Domain Decomposition Methods in Science and Engineering XXVI, vol. 145. Cham: Springer International Publishing; 2022. pp. 697–704. Lecture Notes in Computational Science and Engineering.

55. Kim HH, Chung ET. A BDDC algorithm with enriched coarse spaces for two-dimensional elliptic problems with oscillatory and high contrast coefficients. Multiscale Model Simul. 2015;13(2):571–93.

56. Kim HH, Chung E, Wang J. BDDC and FETI-DP preconditioners with adaptive coarse spaces for three-dimensional elliptic problems with oscillatory and high contrast coefficients. J Comput Phys. 2017;349:191–214.

57. Klawonn A, Lanser M, Weber J. Learning adaptive coarse basis functions of FETI-DP. J Comput Phys. 2024;496:112587.

58. Klawonn A, Lanser M, Weber J. A Domain Decomposition-Based CNN-DNN Architecture for Model Parallel Training Applied to Image Recognition Problems. 2023. arXiv:2302.06564.

59. Klawonn A, Lanser M, Weber J. Learning Adaptive FETI-DP Constraints for Irregular Domain Decompositions. Accepted for publication in the proceedings of the International Conference on Domain Decomposition Methods 27. Springer LNCSE; 2022. https://kups.ub.uni-koeln.de/64122/. TR series, Center for Data and Simulation Science, University of Cologne, Germany, Vol. 2022-3.

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 30 of 31

60. Knoke T, Kinnewig S, Beuchler S, Demircan A, Morgner U, Wick T. Domain Decomposition with Neural Network Interface Approximations for time-harmonic Maxwell's equations with different wave numbers. 2023. arXiv:2303.02590.

61. Kopanicáková A, Karniadakis GE. DeepOnet based preconditioning strategies for solving parametric linear systems of equations. 2024. arXiv preprint arXiv:2401.02016.

62. Kopanicáková A, Kothari H, Karniadakis GE, Krause R. Enhancing training of physics-informed neural networks using domain-decomposition based preconditioning strategies. 2023. arXiv:2306.17648.

63. Kovacs A, Exl L, Kornell A, Fischbacher J, Hovorka M, Gusenbauer M, et al. Conditional physics informed neural networks. Commun Nonlinear Sci Numer Simul. 2022;104:106041.

64. LeCun Y, et al. Generalization and network design strategies. Connectionism Perspect. 1989;19(143–155):18.

65. Lee CO, Lee Y, Ryoo B. A Nonoverlapping Domain Decomposition Method for Extreme Learning Machines: Elliptic Problems. 2024. arXiv preprint arXiv:2406.15959.

66. Lee Y, Kopaničková A, Karniadakis GE. Two-level overlapping additive Schwarz preconditioner for training scientific machine learning applications. 2024. arXiv preprint arXiv:2406.10997.

67. Li J, Cai X. Summation pollution of principal component analysis and an improved algorithm for location sensitive data. Numer Linear Algebra Appl. 2021;28(5):e2370.

68. Li J, Cai XC. Domain decomposed classification algorithms based on linear discriminant analysis: An optimality theory and applications. Neurocomputing. 2024;575:127261.

69. Li K, Tang K, Wu T, Liao Q. D3M: A deep domain decomposition method for partial differential equations. IEEE Access. 2019;8:5283–94.

70. Li Z, Huang DZ, Liu B, Anandkumar A. Fourier neural operator with learned deformations for PDEs on general geometries. J Mach Learn Res. 2023;24(388):1–26.

71. Lions PL. On the Schwarz alternating method. I. In: First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Paris, 1987). Philadelphia: SIAM; 1988. pp. 1–42.

72. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Program. 1989;45(1–3):503–28.

73. Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, et al. Kan: Kolmogorov-arnold networks. 2024. arXiv preprint arXiv:2404.19756.

74. Li S, Xia Y, Liu Y, Liao Q. A deep domain decomposition method based on Fourier features. 2022. arXiv:2205.01884.

75. Li W, Xiang X, Xu Y. Deep domain decomposition method: Elliptic problems. In: Mathematical and Scientific Machine Learning. PMLR; 2020. pp. 269–286.

76. Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nat Mach Intel. 2021;3(3):218–29.

77. Luo L, Shiu WS, Chen R, Cai XC. A nonlinear elimination preconditioned inexact Newton method for blood flow problems in human artery with stenosis. J Comput Phys. 2019;399:108926.

78. Luo T, Ma Z, Xu ZQJ, Zhang Y. Theory of the frequency principle for general deep neural networks. 2019. arXiv:1906.09235.

79. Mai-Duy N, Tran-Cong T. Mesh-free radial basis function network methods with domain decomposition for approximation of functions and numerical solution of Poisson's equations. Eng Anal Bound Elem. 2002;26(2):133–56.

80. Malek A, Emami Kerdabadi A. Solving Differential Equations by Artificial Neural Networks and Domain Decomposition. Iran J Sci. 2023;47(4):1233–44.

81. Mandel J, Sousedík B. Adaptive selection of face coarse degrees of freedom in the BDDC and the FETI-DP iterative substructuring methods. Comput Methods Appl Mech Engrg. 2007;196(8):1389–99.

82. Mandel J, Sousedík B, Šístek J. Adaptive BDDC in three dimensions. Math Comput Simulation. 2012;82(10):1812–31.

83. Mao C, Lupoiu R, Dai T, Chen M, Fan JA. Towards General Neural Surrogate Solvers with Specialized Neural Accelerators. 2024. arXiv preprint arXiv:2405.02351.

84. Mei D, Zhou K, Liu CH. Unified finite-volume physics informed neural networks to solve the heterogeneous partial differential equations. Knowl-Based Syst. 2024;295:111831.

85. Mercier V, Gratton S, Boudier P. A coarse space acceleration of deep-DDM. 2021. arXiv:2112.03732.

86. Misaka T, Mizuno Y, Nakasumi S, Furukawa Y. Domain-Decomposed Physics-Informed Neural Network Prediction on Cartesian CFD Framework, 2024. Proceedings of the International Conference on Scientific Computing and Machine Learning. 2024. https://scml.jp/2024/paper/21/CameraReady/scml2024.pdf.

87. Moseley B, Markham A, Nissen-Meyer T. Finite basis physics-informed neural networks (FBPINNs): a scalable domain decomposition approach for solving differential equations. Adv Comput Math. 2023;49(4):62.

88. Nakahashi K. High-density mesh flow computations with pre-/post-data compressions. In: 17th AIAA Computational Fluid Dynamics Conference 2005. Toronto: 2005. p. 4876. https://doi.org/10.2514/6.2005-4876.

89. Park C, Huang JZ, Ding Y. Domain Decomposition Approach for Fast Gaussian Process Regression of Large Spatial Data Sets. J Mach Learn Res. 2011;12(47):1697–728.

90. Peteiro-Barral D, Guijarro-Berdiñas B. A survey of methods for distributed machine learning. Prog Artif Intell. 2013;2:1–11.

91. Quarteroni A, Valli A. Numerical Approximation of Partial Differential Equations. 2nd ed. Springer Series in Computational Mathematics, Springer Berlin Heidelberg; 2008.

92. Rahaman N, Baratin A, Arpit D, Draxler F, Lin M, Hamprecht F, et al. On the spectral bias of neural networks. In: International Conference on Machine Learning. PMLR; 2019. pp. 5301–5310.

93. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys. 2019;378:686–707.

94. Ronen B, Jacobs D, Kasten Y, Kritchman S. The convergence rate of neural networks for learned functions of different frequencies. Adv Neural Inf Process Syst. 2019;32.

Klawonn *et al. Computational Science and Engineering* (2024) 1:2

Page 31 of 31

95. Schwarz HA. Ueber einen Grenzübergang durch alternirendes Verfahren. In: Gesammelte Mathematische Abhandlungen, vol. 2. Berlin: Springer; 1890. pp. 133–143. First published in Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich, vol. 15, 1870, pp.272-286.

96. Schwenker F, Kestler HA, Palm G. Three learning phases for radial-basis-function networks. Neural Netw. 2001;14(4–5):439–58.

97. Shalev-Shwartz S, Ben-David S. Understanding Machine Learning. Cambridge University Press; 2014.

98. Shazeer N, Mirhoseini A, Maziarz K, Davis A, Le Q, Hinton G, et al. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. 2017. arXiv:1701.06538.

99. Sheng H, Yang C. PFNN: A penalty-free neural network method for solving a class of second-order boundary-value problems on complex geometries. J Comput Phys. 2021;428:110085.

100. Sheng H, Yang C. PFNN-2: A Domain Decomposed Penalty-Free Neural Network Method for Solving Partial Differential Equations. Commun Comput Phys. 2022;32(4):980–1006.

101. Shukla K, Jagtap AD, Karniadakis GE. Parallel physics-informed neural networks via domain decomposition. J Comput Phys. 2021;447:110683.

102. Smith JD, Azizzadenesheli K, Ross ZE. EikoNet: Solving the Eikonal Equation with Deep Neural Networks. IEEE Trans Geosci Remote Sens. 2020;59(12):10685–96.

103. Smith B, Bjørstad P, Gropp W. Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations. Cambridge University Press; 2004.

104. Snir M, Otto S, Huss-Lederman S, Walker D, Dongarra J. MPI-The Complete Reference, Volume 1: The MPI Core. 2nd ed. Cambridge: MIT Press; 1998.

105. Snyder W, Tezaur I, Wentland C. Domain decomposition-based coupling of physics-informed neural networks via the Schwarz alternating method. 2023. arXiv:2311.00224.

106. Stiller P, Bethke F, Böhme M, Pausch R, Torge S, Debus A, et al. Large-scale Neural Solvers for Partial Differential Equations. 2020. arXiv:2009.03730.

107. Sun Q, Xu X, Yi H. Domain Decomposition Learning Methods for Solving Elliptic Problems. 2022. arXiv:2207.10358.

108. Taghibakhshi A, Nytko N, Zaman TU, MacLachlan S, Olson L, West M. MG-GNN: Multigrid Graph Neural Networks for Learning Multilevel Domain Decomposition Methods. 2023. arXiv:2301.11378.

109. Taghibakhshi A, Nytko N, Zaman T, MacLachlan S, Olson L, West M. Learning Interface Conditions in Domain Decomposition Solvers. 2022. arXiv:2205.09833.

110. Taylor JM, Bastidas M, Calo VM, Pardo D. Adaptive Deep Fourier Residual method via overlapping domain decomposition. Comput Methods Appl Mech Eng. 2024;427:116997.

111. Toselli A, Widlund O. Domain Decomposition Methods—Algorithms and Theory. vol. 34 of Springer Series in Computational Mathematics. Berlin: Springer-Verlag; 2005.

112. Verbraeken J, Wolting M, Katzy J, Kloppenburg J, Verbelen T, Rellermeyer JS. A survey on distributed machine learning. ACM Comput Surv. 2020;53(2):1–33.

113. Wang H, Planas R, Chandramowlishwaran A, Bostanabad R. Mosaic flows: A transferable deep learning framework for solving PDEs on unseen domains. Comput Methods Appl Mech Eng. 2022;389:114424.

114. Weber J. Efficient and robust FETI-DP and BDDC methods – approximate coarse spaces and deep learning-based adaptive coarse space. Universität zu Köln; 2022. http://kups.ub.uni-koeln.de/id/eprint/55179.

115. Williams CK, Rasmussen CE. Gaussian processes for machine learning, vol. 2. MA: MIT Press Cambridge; 2006.

116. Xu R, Zhang D, Rong M, Wang N. Weak form theory-guided neural network (TgNN-wf) for deep learning of subsurface single- and two-phase flow. J Comput Phys. 2021;436:110318.

117. Yang HJ, Kim HH. Additive Schwarz algorithms for neural network approximate solutions. 2022. arXiv:2211.00225.

118. Yang HJ, Kim HH. Iterative algorithms for partitioned neural network approximation to partial differential equations. 2023. arXiv:2211.00225.

119. Yin M, Zhang E, Yu Y, Karniadakis GE. Interfacing finite elements with deep neural operators for fast multiscale modeling of mechanics problems. Comput Methods Appl Mech Eng. 2022;402:115027.

120. Zhang D, Lu Z. An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loeve and polynomial expansions. J Comput Phys. 2004;194(2):773–94.

Publisher's Note