

Two-level Deep Domain Decomposition Method

Victorita Dolean^[0000-0002-5885-1903], Serge Gratton^[0000-1111-2222-3333],
Alexander Heinlein^[0000-0003-1578-8104], and Valentin Mercier^[0000-1111-2222-3333]

1 Introduction

The successful application of machine learning in image and language processing has extended its reach to solving physical equations. One example for the combination of scientific computing and machine learning is the field of physics-informed machine learning, with physics-informed neural networks becoming its most prominent example. The idea of this method is to integrate the partial differential equation directly into the loss function for training a neural network to approximate the solution of a boundary value problem (BVP). Therefore, let us consider the generic BVP:

$$\begin{aligned}\mathcal{N}[u](x) &= f(x) \quad \forall x \in \Omega \\ \mathcal{B}[u](x) &= g(x) \quad \forall x \in \partial\Omega\end{aligned}$$

with $\Omega \in \mathbf{R}^d$ and $\partial\Omega$ its boundary, \mathcal{N} a differential and \mathcal{B} a boundary operator.

The approximation of this problem using a neural network u_θ parametrized by weights and biases gathered in θ can be found by the following optimization problem:

Victorita Dolean

Eindhoven University of Technology, Department of Mathematics and Computer Science, PO Box 513, 5600 MB Eindhoven, the Netherlands e-mail: v.dolean.maini@tue.nl

Serge Gratton

Université de Toulouse, INP-ENSEEIH, IRIT, ANITI, Toulouse, France. e-mail: serge.gratton@toulouse-inp.fr

Alexander Heinlein

Delft University of Technology, Delft Institute of Applied Mathematics, Mekelweg 4, 2628 CD Delft, Netherlands e-mail: a.heinlein@tudelft.nl

Valentin Mercier

Université de Toulouse, ANITI, CERFACS, IRIT, Toulouse, and BRLi, France. e-mail: valentin.mercier@toulouse-inp.fr

$$\theta^* = \arg \min_{\theta} \mathcal{M}(\theta),$$

where

$$\mathcal{M}(\theta) = \frac{\lambda_{\Omega}}{N_{\Omega}} \sum_{i=1}^{N_{\Omega}} (\mathcal{N}[u_{\theta}](x_i) - f(x_i))^2 + \frac{\lambda_{\partial\Omega}}{N_{\partial\Omega}} \sum_{i=1}^{N_{\partial\Omega}} (\mathcal{B}[u_{\theta}](\hat{x}_i) - g(\hat{x}_i))^2 \quad (1)$$

Here, $\{x_i\}_{i=1}^{i=N_{\Omega}}$ and $\{\hat{x}_i\}_{i=1}^{i=N_{\partial\Omega}}$ are sets of collocation points sampled in Ω respectively on $\partial\Omega$. The back-propagation algorithm allows for both the evaluation of the residual of the partial differential equation (PDE) and the optimization of the loss function eq. (1) with respect to the network parameters θ .

While domain decomposition methods (DDMs) are well-established solvers for PDEs using classical discretizations, the use of neural network-based discretizations, in particular, PINNs has been explored more recently. This concept has been explored in various studies like for example in [11], where the authors discuss non-overlapping DDMs for parallel training. Extensions to Schwarz methods for Deep Ritz networks, which integrate the variational form into the loss function, are presented in [7]. The coupling in the finite basis PINNs (FBPINNs) approach in [9] differs from the aforementioned approaches; the authors introduce an overlapping domain decomposition, and the coupling is performed via a corresponding partition of unity scaling and hard enforcement of boundary conditions. For a broader overview over the combination of domain decomposition and machine learning methods, we refer to the review [6].

Our focus is on employing PINNs as the subdomain solver in a classical Schwarz approach iteration; this is known as the Deep Domain Decomposition Method (Deep-DDM) introduced in [8]. Given that the subdomain problems can be solved sufficiently accurately, the convergence properties are the same as for the classical Schwarz iteration. As a result, the method is not numerically scalable when increasing the number of subdomains. In this paper, we will incorporate a coarse level, in order to retain numerical scalability. A related two-level Schwarz approach, which uses a different coupling between the two levels of a classical Schwarz iteration has been presented in [5]. Other related multilevel approaches are, for instance, multilevel FBPINNs [1, 2] and multilevel optimization methods using frequency-aware networks [4].

2 The Deep Domain Decomposition Method

Let the computational domain Ω be decomposed into S overlapping subdomains $\Omega_1, \dots, \Omega_S$. Then, we consider the classical Schwarz iteration [10], which involves an independent problem on each subdomain Ω_s : find u_s

$$\begin{cases} \mathcal{N}(u_s) = f & \text{in } \Omega_s, \\ \mathcal{B}(u_s) = g & \text{on } \partial\Omega_s \setminus \Gamma_s, \\ \mathcal{D}(u_s) = \mathcal{D}(u_r) & \text{on } \Gamma_s, \end{cases} \quad (2)$$

where \mathcal{D} is an operator for the transmission conditions (e.g., Dirichlet, Neumann, or Robin) on the subdomain boundary, and Γ_s is the interface between the subdomain Ω_s and the neighboring subdomains Ω_r , with $\Omega_r \cap \Omega_s \neq \emptyset$. For convenience, we define the multiple overlapping subdomains as a single subdomain Ω_r with a corresponding solution network u_r . In order to train a PINN model to solve eq. (2), we incorporate an additional term accounting for the transmission conditions into the loss function eq. (1). In particular, the loss term is computed by sampling points on the interface denoted $\{\tilde{x}_i\}_{i=1}^{N_\Gamma} \subset \partial\Omega_r$. A visualization of all sampling points is shown in Fig. 1 (2). For the transmission to Ω_r , with $\Omega_r \cap \Omega_s \neq \emptyset$, we incorporate the loss term:

$$\mathcal{M}_\Gamma(\theta) = \frac{1}{N_\Gamma} \sum_{i=1}^{N_\Gamma} |\mathcal{D}(u_s(\tilde{x}_k)) - W_i|^2 \quad (3)$$

In this term, we minimize the difference between the trained network u_s and the neighboring networks u_r on the interface with respect to the transfer operator \mathcal{D} . Here,

$$W_i = \mathcal{D}(u_r(\tilde{x}_k)) \quad (4)$$

is defined based on the network u_r from the previous outer Schwarz iteration. Once all the local subnetworks have been trained up to a certain stopping criterion (e.g., number of iterations or tolerance), the interface values W_i are being updated, and we proceed to the next outer iteration.

The one-level Deep DDM algorithm, without the **bold** parts, is shown in Algorithm 1. Initially, two stopping criteria were used to assess convergence: differences in the network solution between subsequent iterations, both in the interior and at the boundary. In addition, a loss criterion variation was used to terminate the training of each network, alongside a maximum epoch limit. This method made comparisons between runs difficult as the number of epochs varied. Here we simplify by using a fixed number of epochs per training and fixed outer iterations. For further details on the original criteria, refer to [8].

Algorithm 1 Two-level DeepDDM; **coarse level**

- 1: Sampling the fine **and the coarse** collocation points
 - 2: Initialization of the network parameters θ_s^0 **and** θ_c
 - 3: Initialization of interface values $W = [W_1, \dots, W_s]$
 - 4: **Initialization of weights** λ_f **and** λ_c
 - 5: while Iteration limits not reached **do**
 - 6: Local network training
 - 7: **Coarse network training**
 - 8: **Compute** $\sum_{s=1}^S E_s(\chi_s u_s(x_{i,coarse}))$ **for each coarse points**
 - 9: Update of W_k values at interfaces with eq. (3) **or eq. (5)**
 - 10: **Update** λ_f **and** λ_c
-

As mentioned in the introduction, the convergence of the one-level method does not scale with the number subdomains. We consider two types of scalability:

- *Strong scalability*: Strong scalability is defined as how the solution time varies with the number of cores for a fixed total problem size. Ideally, the elapsed time is inversely proportional to the number of processing units.
- *Weak scalability*: Weak scalability is defined as how the solution time varies with the number of cores for a fixed problem size per core. Ideally, the elapsed time is constant for a fixed ratio between the size of the problem and the number of processing units.

Since these terms are not clearly defined for of deep learning, we will assume that the size of our problem is the number of points sampled to solve the problem, and the number of processing units is the number of PINNs models used to solve the problem (and therefore the number of subdomains).

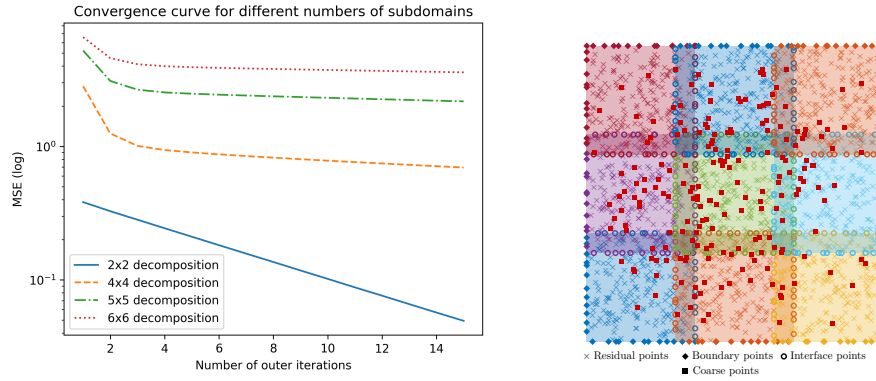


Fig. 1 (1) Strong scalability test on the Deep-DDM method (2) Sampling points for the two-level Deep-DDM method

We consider a Poisson equation with Dirichlet boundary conditions:

$$\begin{cases} \Delta u = r(x) & \text{in } \Omega = [0, 1] \times [0, 1], \\ u = g(x) & \text{on } \partial\Omega. \end{cases}$$

We choose r and g such that the exact solution is $u(x) = \sin(\omega_1 \pi x_1) \sin(\omega_1 \pi x_2) + \sin(\omega_2 \pi x_1) \sin(\omega_2 \pi x_2)$, in this paper we will perform tests for several values of ω_1 and ω_2 . To investigate the scalability of the one-level method we choose $\omega_1 = \omega_2 = 1$ and we perform 1500 epochs per training (other settings are the same as in 4). Here, we test the strong scalability on regular rectangle domain decomposition. A regular rectangle domain decomposition with overlap divides $\Omega = [L_{x_0}, L_{x_1}] \times [L_{y_0}, L_{y_1}]$ into $N_x \times N_y$ subdomains, each overlapping by $\alpha_x \Delta x$ and $\alpha_y \Delta y$. The total number of points sampled remains the same in each experiment, corresponding to our problem size, while we increase the number of subdomains and thus the number of processing

units. In Fig. 1 (1), the mean squared error over outer Schwarz iterations for several numbers of subdomains are plotted. We observe that the convergence deteriorates when increasing the number of subdomains. Therefore, in the next section, we will introduce a coarse level for the Deep-DDM, which facilitates fast global transport of information, to retain scalability.

3 Extension via a Coarse Network

Our coarse level corresponds to training a neural network acting on the entire domain, which we will denote as the *coarse network*. The convergence of the two-level method will then depend on the coarse network solution as well as the exchange of information with the local networks on the first level of the method. In particular, as the coarse network, we train a classical PINN model on the global domain and add an additional loss term incorporating the local subdomain networks. Conversely, after training the coarse network, we will incorporate the coarse network into the loss function for the training of the local networks.

Extension operators and a partition of unity [3]: Let us define an extension operator

$$E_s(w_s) = \begin{cases} w_s & \text{in } \Omega_s \\ 0 & \text{otherwise.} \end{cases}$$

Here, w_s is a function defined on Ω_s . Moreover, we define partition of unity functions χ_s with $\chi_s \geq 0$, $\text{supp}(\chi_s) \subset \Omega_s$, $\chi_s(x) = 0$ for $x \in \partial\Omega_s \setminus \partial\Omega$, and

$$w = \sum_{s=1}^S E_s(\chi_s w|_{\Omega_s})$$

for any function w defined on Ω . We then add the term

$$\mathcal{M}_{fine}(\theta_c) = \frac{1}{N_{\Omega_c}} \sum_{k=1}^{N_{\Omega_c}} \left| u_c(x_k^c) - \sum_{s=1}^S E_k(\chi_s u_s(x_k^c)) \right|^2,$$

to the loss function eq. (1) for the training of the coarse network. Here, $\{x_k^c\}_{k=1}^{N_{\Omega_c}}$ are the sampling points for the coarse network Ω related loss (example in Fig. 1 (2)), and u_c and θ_c are the coarse network and its network parameters, respectively. This term which transfers information from the fine solution to the coarse network is weighted with λ_f . As we aim to train the coarse network concurrently with the fine networks, making our algorithm fully parallelizable, we use the u_s from the previous outer iteration. This information is unavailable for the first iteration, so in this case, we set λ_f to 0.

In order to transfer information from the coarse to the local networks, we introduce a loss term into the M_Γ loss of the local networks via the interface term eq. (4):

$$W_i = \lambda_c \mathcal{D}(u_r(\tilde{x}_i)) + (1 - \lambda_c) \mathcal{D}(u_c(\tilde{x}_i)), \quad (5)$$

where $\lambda_c \in [0, 1]$ is a weight balancing the impact of the coarse network on the local networks.

Note that, since the accuracy of the local networks is often poor in the first iterations, we adjust the corresponding weight in the loss function λ_f during the Schwarz iteration. The final two-level algorithm is given in Algorithm 1, including the **bold** parts.

4 Numerical Results

In this section, we will conduct tests on the previously defined Poisson problem using the following settings. The collocation points are sampled using Latin hypercube sampling in Ω and on $\partial\Omega$ and Γ . For our strong scaling tests, we fix $N_\Omega = 30\,000$ and $N_{\partial\Gamma} = N_\Gamma = 16\,000$ for the whole problem while increasing the number of subdomains. All networks are trained using the Adam optimizer with an initial learning rate of 2×10^{-4} and an exponential decay of 0.999 every 100 epochs. Each neural network is composed of two hidden layers with 30 neurons. The overlap is set to 30% of the subdomain larger side. In each Schwarz iteration, each local and coarse network is trained for 2 500 epochs. The weight controlling the impact of the coarse network on the local networks is set to $\lambda_c = 1 \times 0.9^I$ where I is the index of the Schwarz iteration. This increases the impact of the coarse network during the course of the outer Schwarz iteration. The coefficient λ_f controlling the impact of the fine networks on the coarse network is set to a fixed value of 0.5. All weights λ_* have been optimized using a rough grid search to obtain good performance.

Our implementation uses TensorFlow2 (version 24.02) and runs numerical experiments on a single Nvidia A100-80 GPU. Both the fine and coarse network training processes run concurrently on the GPU using multiprocessing. Although this setup is suboptimal due to the unaddressed GPU load and the complex interaction between TensorFlow and multiprocessing, it allows us to compare the wall time between one-level and two-level methods. The results presented are the median of three independent training runs, each initialized with different seeds.

We focus on investigating the impact of our two-level domain decomposition on the *spectral bias* or *f-principle* [13, 12] of neural networks; this refers to the observation low-frequency components of the target functions are learned much faster than the high-frequency components. The Deep-DDM method tackles this issue by splitting the global problem into smaller subproblems, allowing for better approximation of high frequency components, while the coarse network is supposed to learn the low frequency components.

In order to investigate this, we consider two different pairs of coefficients $(\omega_1, \omega_2) \in \{(1, 3); (1, 6)\}$. For the lower frequency pair of coefficients, that is, $(\omega_1, \omega_2) = (1, 3)$, we observe a noticeable improvement in convergence with the addition of a coarse level; cf. Fig. 2. In particular, the convergence seems to be inde-

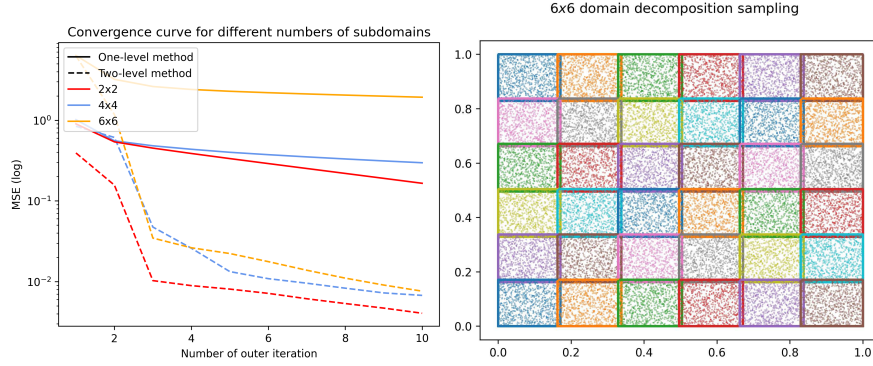


Fig. 2 (1) Strong scalability test for test problem with $\omega_1 = 1$, $\omega_2 = 3$: Test with 2 500 epoch (2) an example of sampling with a 6×6 decomposition

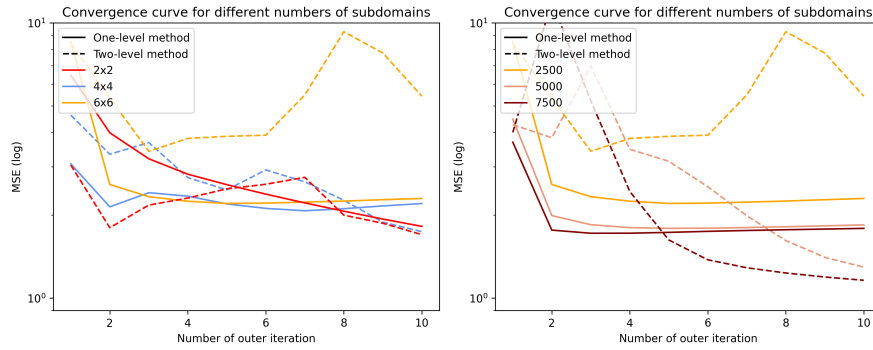


Fig. 3 Strong scalability test for test problem with $\omega_1 = 1$, $\omega_2 = 6$: (1) Test with 2 500 epoch (2) Test on 6×6 decomposition with variation in the number of epochs

| | 2500 epochs | 5000 epochs | 7500 epochs |
|------------------|-------------|-------------|-------------|
| One-level method | 66 min | 148 min | 190 min |
| Two-level method | 71 min | 136 min | 201 min |

Table 1 Median of the wall time for the experiments of the figure 3 (2)

pendent of the number of subdomains. However, when we increase the frequency of the solution, that is, $(\omega_1, \omega_2) = (1, 6)$, we notice that it takes longer for the two-level method before it converges; cf. Fig. 3 (1). We observed that this problem can be easily solved by improving hyperparameter settings. In particular, for 6×6 subdomains, we observe that we can significantly improve the scalability by increasing the number of epochs for each subproblem; cf. Fig. 3 (3). Notably, with 5 000 and 7 500 epochs per subproblem, the two-level methods clearly outperforms the one-level Deep-DDM.

5 Conclusion

We have presented a two-level approach to improve the convergence of the one-level Deep-DDM method. The additional coarse networks facilitates faster global transport of information and enhances the scalability of the Deep-DDM method to larger numbers of subdomains. The cost of training the coarse network is relatively low compared with the total cost of the method and the method is well-suited for a parallel implementation with only small differences in wall time (1) with a non optimized parallelization.

References

1. Dolean, V., Heinlein, A., Mishra, S., Moseley, B.: Finite Basis Physics-Informed Neural Networks as a Schwarz Domain Decomposition Method. In: Z. Dostál, T. Kozubek, A. Klawonn, U. Langer, L.F. Pavarino, J. Šístek, O.B. Widlund (eds.) *Domain Decomposition Methods in Science and Engineering XXVII*, pp. 165–172. Springer Nature Switzerland, Cham (2024)
2. Dolean, V., Heinlein, A., Mishra, S., Moseley, B.: Multilevel domain decomposition-based architectures for physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering* **429**, 117116 (2024)
3. Dolean, V., Jolivet, P., Nataf, F.: *An Introduction to Domain Decomposition Methods*. Society for Industrial and Applied Mathematics (2015)
4. Gratton, S., Mercier, V., Riccietti, E., Toint, P.L.: A block-coordinate approach of multi-level optimization with an application to physics-informed neural networks. *Computational Optimization and Applications* (2024)
5. Jang, D.K., Kim, K., Kim, H.H.: Partitioned neural network approximation for partial differential equations enhanced with lagrange multipliers and localized loss functions. *Computer Methods in Applied Mechanics and Engineering* **429**, 117168 (2024)
6. Klawonn, A., Lanser, M., Weber, J.: Machine learning and domain decomposition methods - a survey. *Computational Science and Engineering* **1**(1), 2 (2024)
7. Li, K., Tang, K., Wu, T., Liao, Q.: D3m: A deep domain decomposition method for partial differential equations. *IEEE Access* **8**, 5283–5294 (2020)
8. Li, W., Xiang, X., Xu, Y.: *Deep Domain Decomposition Method: Elliptic Problems* (2020). ArXiv: 2004.04884
9. Moseley, B., Markham, A., Nissen-Meyer, T.: Finite basis physics-informed neural networks (FBPINNs): a scalable domain decomposition approach for solving differential equations. *Advances in Computational Mathematics* **49**(4), 62 (2023)
10. Schwarz, H.: Über einen Grenzübergang durch alternierendes Verfahren. *Gesammelte Mathematische Abhandlungen* **2**, 133–143 (1890)
11. Shukla, K., Jagtap, A.D., Karniadakis, G.E.: *Parallel Physics-Informed Neural Networks via Domain Decomposition* (2021). ArXiv: 2104.10013
12. Wang, S., Wang, H., Perdikaris, P.: On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering* **384**, 113938 (2021)
13. Xu, Z.Q.J.: Frequency principle: Fourier analysis sheds light on deep neural networks