
Convolutional Hierarchical Deep Learning Neural Networks-Tensor Decomposition (C-HiDeNN-TD): a scalable surrogate modeling approach for large-scale physical systems

Jiachen Guo

Theoretical and Applied Mechanics
Northwestern University
Evanston, Illinois, 60201
jiachen.guo@northwestern.edu

Chanwook Park

Department of Mechanical Engineering
Northwestern University
Evanston, Illinois, 60201
chanwook.park@northwestern.edu

Xiaoyu Xie

Department of Mechanical Engineering
Northwestern University
Evanston, Illinois, 60201
xiaoyu.xie@northwestern.edu

Zhongsheng Sang

Theoretical and Applied Mechanics
Northwestern University
Evanston, Illinois, 60201
zhongsheng.sang@northwestern.edu

Gregory J. Wagner

Department of Mechanical Engineering
Northwestern University
Evanston, Illinois, 60201
gregory.wagner@northwestern.edu

Wing Kam Liu

Department of Mechanical Engineering
Northwestern University
Evanston, Illinois, 60201
w-liu@northwestern.edu

Abstract

A common trend in simulation-driven engineering applications is the ever-increasing size and complexity of the problem, where classical numerical methods typically suffer from significant computational time and huge memory cost. Methods based on artificial intelligence have been extensively investigated to accelerate partial differential equations (PDE) solvers using data-driven surrogates. However, most data-driven surrogates require an extremely large amount of training data. In this paper, we propose the Convolutional Hierarchical Deep Learning Neural Network-Tensor Decomposition (C-HiDeNN-TD) method, which can directly obtain surrogate models by solving large-scale space-time PDE without generating any offline training data. We compare the performance of the proposed method against classical numerical methods for extremely large-scale systems.

1 Introduction

Physics-based numerical simulation is the cornerstone in the rapidly evolving field of digital manufacturing. Despite the rapid development of computational hardware and many advances in high-performance computing (HPC) techniques, classical numerical algorithms still face several significant challenges for extremely large-scale problems that involve billions of degrees of freedom (DoF)(1; 2), including long computational time, excessive memory consumption, and costly storage requirements. Deep learning-based numerical solvers, such as physics-informed neural networks (PINN), will also become expensive to solve long-time physical problems(3).

Many data-driven methods have been proposed to address the bottleneck of classical numerical solvers (4; 5; 6; 7; 8). Data-driven methods typically adopt the so-called offline-online approach. During the offline stage, high-fidelity data are generated from experiments or numerical simulations. Subsequently, various data-driven models can be used to approximate functional mapping from input to output spaces as surrogates (5; 4; 6; 9). Despite the success of these methods on various problems, it is still challenging to apply these data-driven methods to large-scale problems where offline training data generation can be very expensive(1).

In order to circumvent the expensive offline data generation stage and improve the scalability of neural network structure for large-scale systems, we propose Convolutional Hierarchical Deep Learning Neural Networks-Tensor Decomposition (C-HiDeNN-TD) as an a priori method to directly obtain surrogates from solving large-scale PDE. In this paper, we will elucidate the key concepts behind C-HiDeNN-TD and showcase its advantages for large-scale simulations in terms of computational speed, memory efficiency, and storage requirements.

2 Method

2.1 C-HiDeNN

Thanks to the universal approximation theorem, multilayer perceptron (MLP) architectures have been successfully applied in deep learning-based solvers as global basis functions(10). However, MLP-based basis functions need special treatment to satisfy the boundary condition(11). In contrast, finite element (FE) basis functions can automatically satisfy the Dirichlet boundary condition due to the Kronecker-Delta property(12). Moreover, numerical integration can be straightforward with Gaussian quadrature. In the previous work, we combined the concept of locally-supported FE basis functions and MLP to develop Convolutional-Hierarchical Deep Learning Neural Networks (C-HiDeNN) basis function. As shown in Eq. 1, N_i is the linear FE basis function at node i ; W is the kernel function that can be represented by a partially connected MLP(13; 14). This partially connected MLP is controlled by 3 hyperparameters: connectivity s , polynomial order p , and dilation parameter a . C-HiDeNN has been shown to be a flexible way to construct a high-order approximation with arbitrary convergence rates and automatic mesh adaptivity based on a linear FE mesh(13). More details of the C-HiDeNN formulation can be found in Appendix A.

$$u^h(\xi) = \sum_{i \in A^e} N_i(\xi) \sum_{j \in A_s^i} \mathcal{W}_{s,a,p,j}^i(x^h(\xi)) \mathbf{u}_j = \sum_{k \in A_s^e} \widetilde{N}_k(\xi) \mathbf{u}_k \quad (1)$$

2.2 C-HiDeNN Tensor decomposition

Tensor decomposition (TD) is a popular method to decompose higher-order tensors into 1D vectors(15). Likewise, a continuous multivariate function can be decomposed into univariate functions using TD(16). In C-HiDeNN-TD, we decompose a multivariate function using the products of univariate functions approximated by the C-HiDeNN basis function. For an n -variate function, the C-HiDeNN-TD approximation can be written as:

$$u(\mathbf{p}) = u(p_1, p_2, \dots, p_I) \approx \sum_{m=1}^M \left[\widetilde{N}_{p_1}(p_1) \mathbf{u}_{p_1}^{(m)} \right] \cdot \left[\widetilde{N}_{p_2}(p_2) \mathbf{u}_{p_2}^{(m)} \right] \cdot \dots \cdot \left[\widetilde{N}_{p_I}(p_I) \mathbf{u}_{p_I}^{(m)} \right] \quad (2)$$

where m is defined as the decomposition rank; M is the total rank number; \widetilde{N}_{p_i} is the 1D C-HiDeNN basis function for the i -th dimension; $\mathbf{u}_{p_i}^{(m)}$ is the solution vector of i -th dimension for rank m ; \cdot is the vector inner product; Eq. 2 can also be written in a compact vector form:

$$u(\mathbf{p}) \approx [\widetilde{N}_{p_1}(p_1) \mathbf{U}_{p_1}] \odot [\widetilde{N}_{p_2}(p_2) \mathbf{U}_{p_2}] \odot \dots \odot [\widetilde{N}_{p_n}(p_n) \mathbf{U}_{p_n}] \quad (3)$$

where \odot is the Hadamard product; \mathbf{U}_{p_i} is the solution vector for all ranks: $\mathbf{U}_{p_i} = [\mathbf{u}_{p_i}^{(1)}, \mathbf{u}_{p_i}^{(2)}, \dots, \mathbf{u}_{p_i}^{(M)}]$.

Data-driven TD has been extensively investigated as a data compression method for higher-order tensors(15; 17). Unlike data-driven TD, a priori C-HiDeNN-TD can compute the surrogates (solution vectors \mathbf{U}_{p_i}) directly from PDE by circumventing the exorbitant offline high-resolution data generation process. Moreover, for high-dimensional problems, C-HiDeNN-TD will only have linear growth

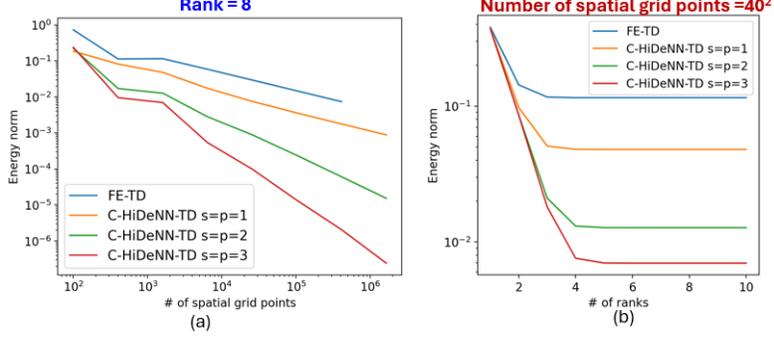


Figure 1: Convergence study of Poisson equation

of the number of grid points compared to exponential growth in classical numerical algorithms. As a result, C-HiDeNN-TD can directly obtain surrogate models for exa-scale problems (number of grid points $> 10^{18}$) with significant speed-up, low memory cost, and efficient disk storage. More details about the C-HiDeNN-TD algorithm can be found in Appendix B.

3 Results

3.1 Convergence study: Poisson equation

In this section, we first study the convergence of C-HiDeNN-TD for the Poisson equation. The 2D Poisson equation is shown below.

$$\begin{cases} \Delta u(x, y) + b(x, y) = 0 & \text{in } \Omega_{(x,y)} \subset \mathbb{R}^2, \\ u|_{\partial\Omega} = 0. \end{cases} \quad (4)$$

where $\Omega_{(x,y)} = [0, 10]^2$ and homogeneous boundary condition is considered in the current example. We assume a concentrated load $b(x, y) = e^{[-10(x-5)^2 - 10(y-5)^2]}$. With C-HiDeNN-TD, the solution to Eq. 4 can be approximated as:

$$u^h(x, y) \approx [\tilde{N}_x(x)U_x] \cdot [\tilde{N}_y(y)U_y] \quad (5)$$

To measure the accuracy of C-HiDeNN-TD, we define the energy norm error.

$$error = \frac{\|u^h - u^{ref}\|_E}{\|u^{ref}\|_E} = \frac{\sqrt{\int_{\Omega} |\nabla(u^h - u^{ref})|^2 dx}}{\sqrt{\int_{\Omega} |\nabla u^{ref}|^2 dx}} \quad (6)$$

where the reference solution u^{ref} is obtained using the linear finite element analysis with 8000×8000 elements. The convergence plots are shown in Fig. 1, where number of spatial grid points is defined as the size of the original spatial mesh. In Fig. 1(a), it can be seen that the C-HiDeNN-TD has better accuracy compared to the FE-TD (FE basis function is used). Furthermore, C-HiDeNN-TD has a higher accuracy as connectivity (s) and polynomial order (p) of the neural network increase. This is because larger s and p will lead to higher-order smoothness of the basis function. As can be seen from Fig. 1(b), when the number of ranks increases, the energy norm error will decrease quickly for all cases until it reaches a plateau.

3.2 Large scale transient diffusion analysis

In this section, we investigate the performance of C-HiDeNN-TD for the large-scale transient diffusion problem. The governing equation is shown below, where the solution u is a function of space and time.

$$\begin{cases} \rho c_p \dot{u} + k \Delta u + b(x, y, z) = 0 & \text{in } \Omega_{(x,y,z)} \otimes \Omega_t, \\ u|_{\partial\Omega} = 0, \\ u|_{t=0} = 0 \end{cases} \quad (7)$$

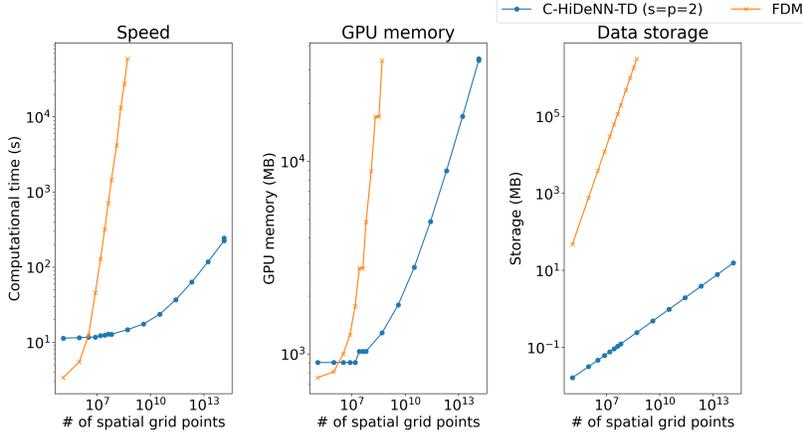


Figure 2: Comparison of C-HiDeNN-TD and finite difference method (FDM)

where $\Omega_{(x,y,z)} = [0, 0.01]^3$; $\Omega_t = (0, 0.5]$; density $\rho = 1000$; heat capacity $c_p = 367.8$; heat conductivity $k = 5.836$; forcing term $b(x, y, z) = 2.55 \times 10^{11} \times H(z - 9.5 \times 10^{-3}) [e^{-2[\frac{(x-2 \times 10^{-3})^2 + (y+2 \times 10^{-3})^2}{(5 \times 10^{-4})^2}]}] + e^{-2[\frac{(x-2 \times 10^{-3})^2 + (y-2 \times 10^{-3})^2}{(5 \times 10^{-4})^2}]}] + e^{-2[\frac{(x+2 \times 10^{-3})^2 + (y-2 \times 10^{-3})^2}{(5 \times 10^{-4})^2}]}]$.

Unlike the classical semi-discretization approach in FE, where the spatial solution is approximated using FE basis function and the temporal problem is solved with time-stepping method, we directly use the C-HiDeNN-TD approximation for both space and time and solve the problem with a space-time approach.

$$u^h(x, y, z, t) \approx [\widetilde{\mathcal{N}}_x(x)U_x] \odot [\widetilde{\mathcal{N}}_y(y)U_y] \odot [\widetilde{\mathcal{N}}_z(z)U_z] \cdot [\widetilde{\mathcal{N}}_t(t)U_t] \quad (8)$$

The performance of C-HiDeNN-TD is compared with explicit finite difference method (FDM) in terms of computational time, GPU memory and disk storage. For fair comparison, both methods are coded in JAX and accelerated with just-in-time (JIT) compilation(18). Different discretizations have been used to study the scalability of the method. C-HiDeNN-TD uses the same discretization for all dimensions, whereas FDM uses the critical time increment for each case. All cases were tested on NVIDIA A6000 GPU with 48GB VRAM.

The computational time is compared in Fig. 2(b), where the x -axis is the total number of points for the spatial grids. For example, for the $100 \times 100 \times 100$ spatial grid, the total number of spatial grid points is 10^6 . As can be seen, the computational time of C-HiDeNN-TD is much less than FDM for large-scale problems. In terms of GPU memory usage, C-HiDeNN-TD requires much less memory as it decomposes the original 4D problem into a series of 1D problems. The largest space-time grids that C-HiDeNN-TD can resolve is $51,200^4 = 6.87 \times 10^{18}$ (51,200 points along x, y, z and t , respectively). Moreover, C-HiDeNN-TD also has a significant storage gain compared to FDM. Classical numerical methods like FDM store the spatial-temporal simulation data as a fourth-order tensor. However, C-HiDeNN-TD directly solves for 1D vectors, which are much cheaper to store. For the $51,200^4$ space-time grids, C-HiDeNN-TD only requires 15.6 MB to store the solution.

4 Conclusion and future work

In this paper, we proposed a new type of neural network-based solver, C-HiDeNN-TD, to tackle the challenge of the ever-growing size of computational problems in real engineering applications. C-HiDeNN-TD can obtain surrogates without the expensive high-fidelity simulation data generation process. Moreover, C-HiDeNN-TD can degenerate high-dimensional problems into 1D subproblems. We have demonstrated the performance of C-HiDeNN-TD for steady-state/transient diffusion equations. C-HiDeNN-TD can solve for surrogate models for large-scale problems without generating expensive data. The current method can be easily extended to a more general space-time-parameter (S-T-P) formulation, where the parameters of PDE can also be included as extra-coordinates(16). Moreover, C-HiDeNN-TD will be applied to solve nonlinear PDE such as Navier-Stokes equation.

References

- [1] Yangfan Li, Satyajit Mojumder, Ye Lu, Abdullah Al Amin, Jiachen Guo, Xiaoyu Xie, Wei Chen, Gregory J Wagner, Jian Cao, and Wing Kam Liu. Statistical parameterized physics-based machine learning digital shadow models for laser powder bed fusion process. *Additive Manufacturing*, 87:104214, 2024.
- [2] Hengyang Li, Stefan Knapik, Yangfan Li, Chanwook Park, Jiachen Guo, Satyajit Mojumder, Ye Lu, Wei Chen, Daniel W Apley, and Wing Kam Liu. Convolution hierarchical deep-learning neural network tensor decomposition (c-hidenn-td) for high-resolution topology optimization. *Computational Mechanics*, 72(2):363–382, 2023.
- [3] Xuhui Meng, Zhen Li, Dongkun Zhang, and George Em Karniadakis. Ppinn: Parareal physics-informed neural network for time-dependent pdes. *Computer Methods in Applied Mechanics and Engineering*, 370:113250, 2020.
- [4] Owen Huang, Sourav Saha, Jiachen Guo, and Wing Kam Liu. An introduction to kernel and operator learning methods for homogenization by self-consistent clustering analysis. *Computational Mechanics*, 72(1):195–219, 2023.
- [5] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.
- [6] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deepnet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.
- [7] Sourav Saha, Chanwook Park, Stefan Knapik, Jiachen Guo, Owen Huang, and Wing Kam Liu. Deep learning discrete calculus (dlcd): a family of discrete numerical methods by universal approximation for stem education to frontier research. *Computational mechanics*, 72(2):311–331, 2023.
- [8] Xiaoyu Xie, Arash Samaei, Jiachen Guo, Wing Kam Liu, and Zhengtao Gan. Data-driven discovery of dimensionless numbers and governing laws from scarce measurements. *Nature communications*, 13(1):7562, 2022.
- [9] Chanwook Park, Sourav Saha, Jiachen Guo, Xiaoyu Xie, Satyajit Mojumder, Miguel A Bessa, Dong Qian, Wei Chen, Gregory J Wagner, Jian Cao, et al. Engineering software 2.0 by interpolating neural networks: unifying training, solving, and calibration. *arXiv preprint arXiv:2404.10296*, 2024.
- [10] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.
- [11] Stefano Berrone, Claudio Canuto, Moreno Pintore, and Natarajan Sukumar. Enforcing dirichlet boundary conditions in physics-informed neural networks and variational physics-informed neural networks. *Heliyon*, 9(8), 2023.
- [12] Junuthula Narasimha Reddy. An introduction to the finite element method. *New York*, 27:14, 1993.
- [13] Ye Lu, Hengyang Li, Lei Zhang, Chanwook Park, Satyajit Mojumder, Stefan Knapik, Zhongsheng Sang, Shaoqiang Tang, Daniel W Apley, Gregory J Wagner, et al. Convolution hierarchical deep-learning neural networks (c-hidenn): finite elements, isogeometric analysis, tensor decomposition, and beyond. *Computational Mechanics*, 72(2):333–362, 2023.
- [14] Chanwook Park, Ye Lu, Sourav Saha, Tianju Xue, Jiachen Guo, Satyajit Mojumder, Daniel W Apley, Gregory J Wagner, and Wing Kam Liu. Convolution hierarchical deep-learning neural network (c-hidenn) with graphics processing unit (gpu) acceleration. *Computational Mechanics*, 72(2):383–409, 2023.

- [15] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
- [16] Ye Lu, Satyajit Mojumder, Jiachen Guo, Yangfan Li, and Wing Kam Liu. Extended tensor decomposition model reduction methods: Training, prediction, and design under uncertainty. *Computer Methods in Applied Mechanics and Engineering*, 418:116550, 2024.
- [17] Y Lu, N Blal, and A Gravouil. Datadriven hopgd based computational vademecum for welding parameter identification. *Computational mechanics*, 64(1):47–62, 2019.
- [18] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/google/jax>.
- [19] Lei Zhang, Lin Cheng, Hengyang Li, Jiaying Gao, Cheng Yu, Reno Domel, Yang Yang, Shaoqiang Tang, and Wing Kam Liu. Hierarchical deep-learning neural networks: finite elements and beyond. *Computational Mechanics*, 67:207–230, 2021.

Appendix

A. C-HiDeNN theory

The C-HiDeNN basis function exploits the merits of the finite element basis function and the flexibility of multilayer perceptron (MLP). The first part of C-HiDeNN is the Lagrange polynomial in finite element methods, which can exactly satisfy the Dirichlet boundary condition. As shown in the previous work(19), 1st-order Lagrange polynomial is equivalent to a partially connected 2-layer MLP. To further enhance the smoothness, we add an additional third layer with nonlinear activation with p reproducing polynomial order. The full mathematical form of C-HiDeNN basis function can be written as:

$$u^{h,e}(\xi) = \sum_{i \in A^e} N_i(\xi) \sum_{j \in A_s^i} \mathcal{W}_{s,a,p,j}^i(x^{h,e}(\xi)) \mathbf{u}_j = \sum_{k \in A_s^e} \widetilde{N}_k(\xi) \mathbf{u}_k \quad (9)$$

where $N_i(\xi)$ is polynomial interpolant at node i , $W_{s,a,p,j}^i(\xi)$ is the convolution patch function at node j defined on the nodal patch domain A_s^i with three hyper-parameters (s : patch size, a : dilation parameter, p : polynomial order), and $x^{h,e}(\xi)$ is the FEM mapping from the parent coordinate ξ to the physical coordinate x in element e . Note that the input of the neural network is changed from the physical coordinate x to the parent coordinate ξ in Fig. 3(a) to facilitate the numerical integration of the weak form of PDE. The convolution patch functions are precomputed from the meshfree theory and used as weights between the second and third hidden layers in Figure 3(b). More details on how the weights are calculated can be found in (14).

B. A priori C-HiDeNN-TD

In this section, the algorithm of C-HiDeNN-TD is explained for the 2D Poisson equation. C-HiDeNN-TD solves the weak form of the Poisson equation using the Galerkin formulation. The weak form of 2D Poisson equation is:

$$\int_{\Omega} \nabla \delta u \cdot \nabla u - \int_{\partial\Omega} \delta u \cdot (\nabla u \cdot \mathbf{n}) - \int_{\Omega} \delta u b = 0 \quad (10)$$

where u is the trial function, δu is the test function. In Galerkin formulation, we use the same C-HiDeNN-TD basis function for the trial and test functions. Note that we have 2 test functions for x and y dimension, respectively.

$$u(x, y) \approx [\widetilde{N}_x(x) \mathbf{U}_x] \cdot [\widetilde{N}_y(y) \mathbf{U}_y] \quad (11)$$

$$\delta u_x(x, y) \approx [\widetilde{N}_x(x) \delta \mathbf{U}_x] \cdot [\widetilde{N}_y(y) \mathbf{U}_y] \quad (12a)$$

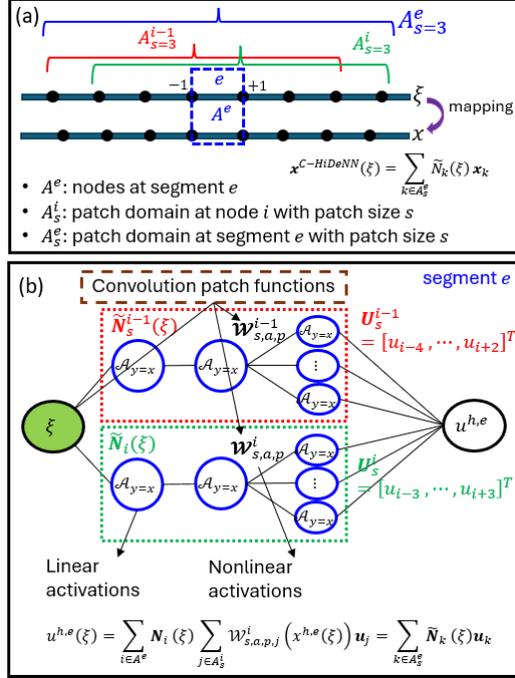


Figure 3: C-HiDeNN structure. (a) 1D C-HiDeNN basis function is constructed by aggregating the information from all the segments patch domain into the current segment (b) Neural network structure of C-HiDeNN basis function: the first 2 hidden layers are mapping the input to piece-wise continuous function N , which satisfies the Kronecker delta property; the 3rd layer is the nonlinear part that acts as the convolution patch function W , which satisfies the reproducing condition; the output field variable $u^{h,e}$ has higher-order smoothness and automatically satisfy the Dirichlet boundary condition thanks to the Kronecker delta property.

$$\delta u_y(x, y) \approx [\tilde{N}_x(x) U_x] \cdot [\tilde{N}_y(y) \delta U_y] \quad (12b)$$

Plugging Eq. 11-12 into Eq. 10 and simplifying the equations, The original 2D problem is decomposed into two 1D subproblems as shown below.

$$A_x U_x = Q_x \quad (13a)$$

$$A_y U_y = Q_y \quad (13b)$$

where $A_i \in \mathbb{R}^{ML_i \times ML_i}$ is the coefficient matrix for dimension i (M refers to number of ranks; L_i is the number of nodes in the i -th dimension); $U_i \in \mathbb{R}^{ML_i}$ is the solution vector. Eq. 13a and Eq. 13b are solved iteratively until the variation of U_x and U_x is within the tolerance range.