The coupled deep neural networks for coupling of the Stokes and Darcy–Forchheimer problems

Jing Yue(岳靖)[1], Jian Li(李剑)[1,†], Wen Zhang(张文)[1], and Zhangxin Chen(陈掌星)[2,3]

[1] *School of Electrical and Control Engineering, School of Mathematics and Data Science, Shaanxi University of Science and Technology, Xi'an 710021, China*
[2] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an* 710049, *China*
[3] *Department of Chemical and Petroleum Engineering, Schulich School of Engineering, University of Calgary,* 2500 *University Drive N.W., Calgary, Alberta T2N* 1*N*4, *Canada*

(Received 7 April 2022; revised manuscript received 18 May 2022; accepted manuscript online 2 June 2022)

We present an efficient deep learning method called coupled deep neural networks (CDNNs) for coupling of the Stokes and Darcy–Forchheimer problems. Our method compiles the interface conditions of the coupled problems into the networks properly and can be served as an efficient alternative to the complex coupled problems. To impose energy conservation constraints, the CDNNs utilize simple fully connected layers and a custom loss function to perform the model training process as well as the physical property of the exact solution. The approach can be beneficial for the following reasons: Firstly, we sample randomly and only input spatial coordinates without being restricted by the nature of samples. Secondly, our method is meshfree, which makes it more efficient than the traditional methods. Finally, the method is parallel and can solve multiple variables independently at the same time. We present the theoretical results to guarantee the convergence of the loss function and the convergence of the neural networks to the exact solution. Some numerical experiments are performed and discussed to demonstrate performance of the proposed method.

Keywords: scientific computing, machine learning, the Stokes equations, Darcy–Forchheimer problems, Beavers–Joseph–Saffman interface condition

PACS: 02.30.Jr, 47.11.–j, 07.05.Tp, 07.05.Mh **DOI:** 10.1088/1674-1056/ac7554

1. Introduction

Fluid flows between porous media and free-flow zones have extensive applications in hydrology, environmental science, and biofluid dynamics. A lot of researchers have derived suitable mathematical and numerical models for fluid movement.[1–3] A system can be viewed as a coupled problem with two physical systems interacting across an interface. The simplest mathematical formulation for the coupled problem is coupling of the Stokes and Darcy flow with proper interface conditions. The most suitable and popular interface conditions are called the Beavers–Joseph–Saffman conditions.[4] However, Darcy's law only provides a linear relationship between the gradient of pressure and velocity in the coupled model, which usually fails for complex physical problems. Forchheimer[5] conducted flow experiments in sand packs and recognized that for moderate Reynolds numbers ($Re > 0.1$ approximately), Darcy's law is not adequate. He found that the pressure gradient and Darcy velocity should satisfy the Darcy–Forchheimer law. Since the great attention has been paid to the coupled model, a large number of traditional mesh-based methods have been devoted to the coupled Stokes and Darcy flows problems.[6–22]

Owing to the enormous potential in approximating complex nonlinear maps,[26–32] deep learning has attracted growing attention in many applications, such as image, speech, text recognition, and scientific computing.[23–25] Many works have arisen based on the function approximation capabilities of the feed-forward fully connected neural network to solve initial/boundary value problems[40–43] in recent decades. The solution to the system of equations can be obtained by minimizing the loss function, which typically consists of the residual error of the governing partial differential equations (PDEs) along with initial/boundary values. Resently, Raissi *et al.* proposed physics informed neural networks (PINNs)[44–46] and they have been widely used.[47–52] Moreover, Sirignano and Spiliopoulos presented the deep learning Galerkin method[54] for solving high dimensional PDEs. Additionally, some recent works have successfully solved the second-order linear elliptic equations and the high dimensional Stokes problems.[33–36] Though several excellent works have been performed in applying deep learning to solve PDEs,[37] the topic for solving complicated coupled physical problems remains to be investigated.

Considering the performance of deep learning for solving PDEs, our contribution is to design the CDNNs as an efficient alternative model for complicated coupled physical problems. We can encode any underlying physical laws naturally as prior information to obey the law of physics. To satisfy the differential operators, boundary conditions and divergence conditions, we train the neural networks on batches of randomly sampled points. The method only inputs random sampling spatial coordinates without considering the nature of samples.

[†]Corresponding author. E-mail: jianli@sust.edu.cn; jiaaanli@gmail.com
© 2023 Chinese Physical Society and IOP Publishing Ltd

Notably, we take the interface conditions as the constraint for the CDNNs. The approach is parallel and solves multiple variables independently at the same time. Specially, the optimal solution can be obtained by using the networks instead of a linear combination of basic functions. Furthermore, we validate the convergence of the loss function under certain conditions and the convergence of the CDNNs to the exact solution. Several numerical experiments are conducted to investigate the performance of the CDNNs.

The article is organized as follows: Section 2 introduces the coupled model and the relation methodology. Section 3 discusses the convergence of the loss function $J(\overline{U})$ and the convergence of the CDNNs to the exact solution. Section 4 reveals some numerical experiments to illustrate the efficiency of the CDNNs. The article ends with conclusion in Section 5.

2. Methodology

Let $\Omega_{\rm S}$ and $\Omega_{\rm D}$ be two bounded and simply connected polygonal domains in $\mathbb{R}^2$, as shown in Fig. 1. Let $\boldsymbol{n}_{\rm S}$ denote the unit normal vector pointing from $\Omega_{\rm S}$ to $\Omega_{\rm D}$, and take $\boldsymbol{n}_{\rm D}$ as the unit normal vector pointing from $\Omega_{\rm D}$ to $\Omega_{\rm S}$, on the interface Γ, then we have $\boldsymbol{n}_{\rm D} = -\boldsymbol{n}_{\rm S}$.

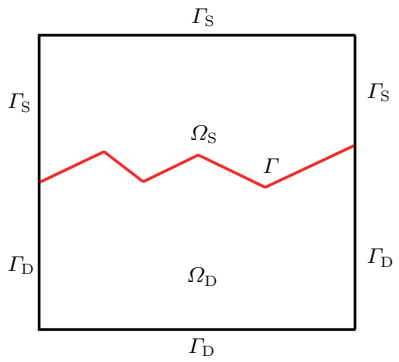


Fig. 1. Coupled domain with interface Γ.

When kinematic effects surpass viscous effects in a porous medium, the Darcy velocity $\boldsymbol{u}_{\rm D}$ and the pressure gradient $\nabla p_{\rm D}$ does not satisfy a linear relation. Instead, a nonlinear approximation, known as the Darcy–Forchheimer model, is considered. When it is imposed on the porous medium $\Omega_{\rm D}$ with homogeneous Dircihlet boundary condition on $\Gamma_{\rm D}$, the equations can be written as

$$\nabla \cdot \boldsymbol{u}_{\rm D} = f_{\rm D}, \quad \text{in } \Omega_{\rm D}, \tag{1}$$

$$\frac{\mu}{\rho}\boldsymbol{K}^{-1}\boldsymbol{u}_{\rm D} + \frac{\beta}{\rho} \mid \boldsymbol{u}_{\rm D} \mid \boldsymbol{u}_{\rm D} + \nabla p_{\rm D} = \boldsymbol{g}_{\rm D}, \quad \text{in } \Omega_{\rm D}, \tag{2}$$

$$p_{\rm D} = 0, \quad \text{on } \Gamma_{\rm D}, \tag{3}$$

where $\boldsymbol{K}$ is the permeability tensor assumed to be uniformly positive definite and bounded, ρ is the density of the fluid, μ is viscosity, and β is a dynamic viscosity, all assumed to be positive constants. In addition, $\boldsymbol{g}_{\rm D}$ and $f_{\rm D}$ are source terms. We remark that in this context we exploit homogeneous Dirichlet boundary condition, in fact, we can also consider homogeneous Neumann boundary condition, i.e., $\boldsymbol{u}_{\rm D} \cdot \boldsymbol{n}_{\rm D} = 0$ on $\Gamma_{\rm D}$ and the arguments used in this paper are still true.

The fluid motion in $\Omega_{\rm S}$ is described by the Stokes equations:

$$-\nu\Delta \boldsymbol{u}_{\rm S} + \nabla p_{\rm S} = \boldsymbol{f}_{\rm S}, \quad \text{in } \Omega_{\rm S}, \tag{4}$$

$$\nabla \cdot \boldsymbol{u}_{\rm S} = 0, \quad \text{in } \Omega_{\rm S}, \tag{5}$$

$$\boldsymbol{u}_{\rm S} = 0, \quad \text{on } \Gamma_{\rm S}, \tag{6}$$

where $\nu > 0$ denotes the viscosity of the fluid.

On the interface, we prescribe the following interface conditions:

$$\boldsymbol{u}_{\rm S} \cdot \boldsymbol{n}_{\rm S} = \boldsymbol{u}_{\rm D} \cdot \boldsymbol{n}_{\rm S}, \quad \text{on } \Gamma, \tag{7}$$

$$p_{\rm S} - \nu \boldsymbol{n}_{\rm S} \frac{\partial \boldsymbol{u}_{\rm S}}{\partial \boldsymbol{n}_{\rm S}} = p_{\rm D}, \quad \text{on } \Gamma, \tag{8}$$

$$-\nu \boldsymbol{t} \frac{\partial \boldsymbol{u}_{\rm S}}{\partial \boldsymbol{n}_{\rm S}} = G\boldsymbol{u}_{\rm S} \cdot \boldsymbol{t}, \quad \text{on } \Gamma. \tag{9}$$

Here, $\boldsymbol{t}$ represents the unit tangential vector along the interface Γ. Condition (7) represents continuity of the fluid velocity's normal components, Eq. (8) represents the balance of forces acting across the interface, and Eq. (9) is the Beavers–Joseph–Saffman condition.[55] The constant $G > 0$ is given and usually obtained from experimental data.

For notational brevity, we set $\overline{\boldsymbol{u}} = (\boldsymbol{u}_{\rm S}, \boldsymbol{u}_{\rm D}, p_{\rm S}, p_{\rm D})$ and recall the classical Sobolev spaces

$$\mathbb{X}_{\rm S}^0 = \{\boldsymbol{v}_{\rm S} \in [H^1(\Omega_{\rm S})]^d : \boldsymbol{v}_{\rm S}|_{\Gamma_{\rm S}} = \boldsymbol{0}\},$$
$$\mathbb{Y}_{\rm D} = \{q_{\rm D} \in [W^{1,3/2}(\Omega_{\rm D})]^2 : q_{\rm D}|_{\Gamma_{\rm D}} = 0\},$$
$$\mathbb{X}_{\rm S} = \{\boldsymbol{v}_{\rm S} \in \mathbb{X}_{\rm S}^0 : \text{div } \boldsymbol{v}_{\rm S} = 0\},$$

where

$$H^k(\Omega) = \left\{ \upsilon \in L^2(\Omega) : D_w^\alpha \upsilon \in L^2(\Omega), \forall \alpha : \mid \alpha \mid \leq k \right\},$$

and their norm

$$\| \upsilon \|_k = \sqrt{(\upsilon, \upsilon)_k} = \left\{ \sum_{|\alpha|=0}^{k} \int_\Omega (D_w^\alpha \upsilon)^2 \mathrm{d}x \right\}^{1/2},$$

$$\| \upsilon \|_{W_{k,p}} = \left\{ \sum_{|\alpha| \leq k} \| \upsilon \|_{L^p}^p \right\}^{1/p}.$$

Particularly, $\|v\|_k = \|v\|_{W^{k,2}}$. Here, $k > 0$ is a positive integer, $\| \upsilon \|_0$ denotes the norm on $L^2(\Omega)$ or $(L^2(\Omega))^2$, and $D_w^\alpha \upsilon$ is the generalized derivative of υ. Moreover, $(\cdot,\cdot)_{\rm D}$ represents the inner product in the domain D and $\langle \cdot, \cdot \rangle$ represents the inner product on the interface Γ.

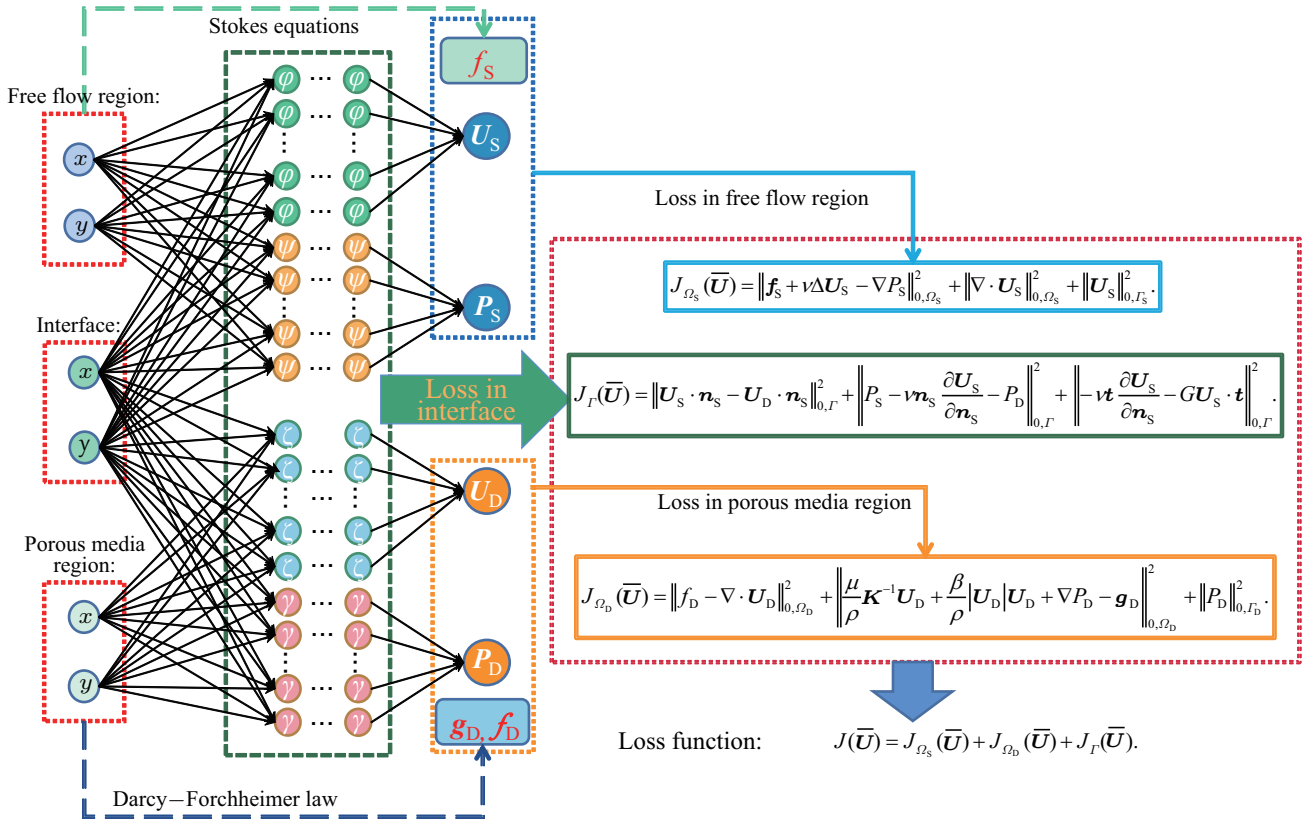


Fig. 2. The structure of the CDNNs.

To solve coupling of the Stokes and Darcy–Forchheimer problems, we propose the CDNNs in Fig. 2, where x, y represent the spatial coordinates in different domains. Furthermore, we give observations of the state variable $\overline{\boldsymbol{U}}(x;\theta) = \big(\boldsymbol{U}_{\rm S}(x;\theta_1),\ \boldsymbol{U}_{\rm D}(x;\theta_2),\ P_{\rm S}(x;\theta_3),\ P_{\rm D}(x;\theta_4)\big)$, which is the neural network solution to the coupled Stokes and Darcy–Forchheimer problems (1)–(9). Here, (θ_1,θ_3) and (θ_2,θ_4) are the stacked parameters of θ for Stokes and Darcy–Forchheimer, respectively. The following constrained optimization procedure aims to reconstruct the parameters θ by minimizing the loss function:

$$J[\overline{\boldsymbol{U}}] = J_{\Omega_{\rm S}}[\overline{\boldsymbol{U}}] + J_{\Omega_{\rm D}}[\overline{\boldsymbol{U}}] + J_{\Gamma}[\overline{\boldsymbol{U}}], \tag{10}$$

where

$$\begin{aligned}
J_{\Omega_{\rm S}}(\overline{\boldsymbol{U}}) =\ & \big\|\boldsymbol{f}_{\rm S} + \nu\Delta\boldsymbol{U}_{\rm S}(x;\theta_1) - \nabla P_{\rm S}(x;\theta_3)\big\|^2_{0,\Omega_{\rm S},\omega_1} \\
& + \big\|\nabla\cdot\boldsymbol{U}_{\rm S}(x;\theta_1)\big\|^2_{0,\Omega_{\rm S},\omega_1} + \big\|\boldsymbol{U}_{\rm S}(x;\theta_1)\big\|^2_{0,\Gamma_{\rm S},\omega_2}, \\
J_{\Omega_{\rm D}}(\overline{\boldsymbol{U}}) =\ & \big\|f_{\rm D} - \nabla\cdot\boldsymbol{U}_{\rm D}(x;\theta_2)\big\|^2_{0,\Omega_{\rm D},\omega_1} \\
& + \Big\|\frac{\mu}{\rho}\boldsymbol{K}^{-1}\boldsymbol{U}_{\rm D}(x;\theta_2) + \frac{\beta}{\rho}\big|\boldsymbol{U}_{\rm D}(x;\theta_2)\big|\boldsymbol{U}_{\rm D}(x;\theta_2) \\
& + \nabla P_{\rm D}(x;\theta_4) - \boldsymbol{g}_{\rm D}\Big\|^2_{0,\Omega_{\rm D},\omega_1} \\
& + \big\|P_{\rm D}(x;\theta_4)\big\|^2_{0,\Gamma_{\rm D},\omega_2}, \\
J_{\Gamma}(\overline{\boldsymbol{U}}) =\ & \big\|\boldsymbol{U}_{\rm S}(x;\theta_1)\cdot\boldsymbol{n}_{\rm S} - \boldsymbol{U}_{\rm D}(x;\theta_2)\cdot\boldsymbol{n}_{\rm S}\big\|^2_{0,\Gamma,\omega_3} \\
& + \Big\|P_{\rm S}(x;\theta_3) - \nu\boldsymbol{n}_{\rm S}\frac{\partial\boldsymbol{U}_{\rm S}(x;\theta_1)}{\partial\boldsymbol{n}_{\rm S}} - P_{\rm D}(x;\theta_4)\Big\|^2_{0,\Gamma,\omega_3} \\
& + \Big\| - \nu\boldsymbol{t}\frac{\partial\boldsymbol{U}_{\rm S}(x;\theta_1)}{\partial\boldsymbol{n}_{\rm S}} - G\boldsymbol{U}_{\rm S}(x;\theta_1)\cdot\boldsymbol{t}\Big\|^2_{0,\Gamma,\omega_3}.
\end{aligned} \tag{11}$$

It should be noted that $J(\overline{\boldsymbol{U}})$ can measure how well the approximate solution satisfies differential operators, divergence conditions, boundary conditions and interface conditions. Furthermore, the norm in Eq. (11) means that $\big\|f(y)\big\|^2_{0,\mathcal{Y},\omega} = \int_{\mathcal{Y}}\big|f(y)\big|^2\omega(y)\,{\rm d}y$, where $\omega(y)$ is the probability density of variable y in domain $\mathcal{Y}$. In this study, the training data are sampled randomly from interior domains, boundary domains and interface by the respective probability densities ω_1, ω_2, and ω_3. Especially, if $J(\overline{\boldsymbol{U}}) = 0$, then $\overline{\boldsymbol{U}}$ is the solution to the coupled Stokes and Darcy–Forchheimer problems (1)–(9). Since it is infeasible to estimate θ by directly minimizing $J(\overline{\boldsymbol{U}})$ when integrated over a higher dimensional region, we apply a sequence of randomly sampled points from domain instead forming mesh grid. The main steps of the CDNNs for the coupled Stokes and Darcy–Forchheimer equations are presented as Algorithm 1. Another noticeable point is that the term $\nabla_\theta G(\rho^{(n)},\theta^n)$ is unbiased estimate of $\nabla_\theta J\big(\overline{\boldsymbol{U}}(\cdot;\theta^n)\big)$ because the population parameters can be estimated by sample mathematical expectations.

Algorithm 1: The CDNNs for coupling of the Stokes and Darcy-Forchheimer problems

Input: $\rho_{\omega_1}^{(n)}=\{x_{\rm S}^n,x_{\rm D}^n\},\rho_{\omega_2}^{(n)}=\{r_{\rm S}^n,r_{\rm D}^n\},\rho_{\omega_3}^{(n)}=\{r_\Gamma^n\}$, max iterations M, learning rate α

Output: θ_{n+1}

1. Randomly generated sample points $\rho^{(n)}=\{\rho_{\omega_1}^{(n)},\rho_{\omega_2}^{(n)},\rho_{\omega_3}^{(n)}\}$ from $(\Omega_{\rm S},\Omega_{\rm D})$, $(\Gamma_{\rm S},\Gamma_{\rm D})$ and Γ by the respective probability densities ω_1, ω_2 and ω_3;
2. Initialize the parameters θ;
3. **while** *iterations* $\le M$ **do**

read current;

$$G(\rho^{(n)},\theta^n)=G_{\rm S}(\rho^{(n)},\theta^n)+G_{\rm D}(\rho^{(n)},\theta^n)+G_\Gamma(\rho^{(n)},\theta^n),$$

where

$$\begin{aligned}G_{\rm S}(\rho^{(n)},\theta^n)&=\Big(\boldsymbol{f}_{\rm S}+\nu\Delta\boldsymbol{U}_{\rm S}(x_{\rm S}^n;\theta_1)-\nabla P_{\rm S}(x_{\rm S}^n;\theta_3)\Big)^2\\&\quad+\Big(\nabla\cdot\boldsymbol{U}_{\rm S}(x_{\rm S}^n;\theta_1)\Big)^2+\Big(\boldsymbol{U}_{\rm S}(r_{\rm S}^n;\theta_1)\Big)^2,\\G_{\rm D}(\rho^{(n)},\theta^n)&=\Big(\frac{\mu}{\rho}K^{-1}\boldsymbol{U}_{\rm D}(x_{\rm D}^n;\theta_2)+\frac{\beta}{\rho}\big|\boldsymbol{U}_{\rm D}(x_{\rm D}^n;\theta_2)\big|\boldsymbol{U}_{\rm D}(x_{\rm D}^n;\theta_2)+\nabla P_{\rm D}(x_{\rm D}^n;\theta_4)-\boldsymbol{g}_{\rm D}\Big)^2\\&\quad+\Big(f_{\rm D}-\nabla\cdot\boldsymbol{U}_{\rm D}(x_{\rm D}^n;\theta_2)\Big)^2+\Big(P_{\rm D}(r_{\rm D}^n;\theta_4)\Big)^2,\end{aligned}$$

and

$$\begin{aligned}G_\Gamma(\rho^{(n)},\theta^n)&=\Big(\boldsymbol{U}_{\rm S}(x;\theta_1)\cdot\boldsymbol{n}_{\rm S}-\boldsymbol{U}_{\rm D}(x;\theta_2)\cdot\boldsymbol{n}_{\rm S}\Big)^2\\&\quad+\Big(P_{\rm S}(x;\theta_3)-\nu\boldsymbol{n}_{\rm S}\frac{\partial\boldsymbol{U}_{\rm S}(x;\theta_1)}{\boldsymbol{n}_{\rm S}}-P_{\rm D}(x;\theta_4)\Big)^2\\&\quad+\Big(-\nu\boldsymbol{t}\frac{\partial\boldsymbol{U}_{\rm S}(x;\theta_1)}{\partial\boldsymbol{n}_{\rm S}}-G\boldsymbol{U}_{\rm S}(x;\theta_1)\cdot\boldsymbol{t}\Big)^2,\end{aligned}$$

and

$$\theta^{n+1}=\theta_n-\alpha\nabla_\theta G(\rho^{(n)},\theta^n);$$

if $\lim_{n\to\infty}\|\nabla_\theta G(\rho^{(n)},\theta_n)\|=0$ **then**

return the parameters θ_{n+1};

else

go back to the beginning of current section.

3. Convergence

According to the definition of loss function $J(\overline{U})$, it can measure how well $\overline{U}$ satisfies Eqs. (1)–(9). Neural networks are a set of algorithms for classification and regression tasks inspired by the biological neural networks in brains. There have various types of neural networks with different neuron connection forms and architectures. According to Ref. [29], if there is only one hidden layer and output, a set of functions implemented by following networks with m_1, m_2, m_3 and m_4 hidden units for coupling of the Stokes and Darcy–Forchheimer problems are

$$[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}^{m_1}(\varphi)]^d=\Big\{\Theta(x):\mathbb{R}^d\mapsto\mathbb{R}^d\Big|\Theta(x)=\sum_{i=1}^{m_1}\beta_i\varphi\Big(\sum_{j=1}^d\sigma_{j,i}x_j+c_i\Big)\Big\},$$

$$[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}^{m_2}(\zeta)]^d=\Big\{\Lambda(x):\mathbb{R}^d\mapsto\mathbb{R}^d\Big|\Lambda(x)=\sum_{i=1}^{m_2}\beta_i'\zeta\Big(\sum_{j=1}^d\sigma_{j,i}'x_j+c_i'\Big)\Big\},$$

$$\mathfrak{C}_{P_{\rm S}}^{m_3}(\psi)=\Big\{\Psi(x):\mathbb{R}^d\mapsto\mathbb{R}\Big|\Psi(x)=\sum_{i=1}^{m_3}\beta_i''\psi\Big(\sum_{j=1}^d\sigma_{j,i}''x_j+c_i''\Big)\Big\},$$

$$\mathfrak{C}_{P_{\rm D}}^{m_4}(\gamma)=\Big\{\Upsilon(t,x):\mathbb{R}^d\mapsto\mathbb{R}\Big|\Upsilon(x)=\sum_{i=1}^{m_4}\beta_i'''\gamma\Big(\sum_{j=1}^d\sigma_{j,i}'''x_j+c_i'''\Big)\Big\},$$

where

$$\begin{aligned}\Theta(x)&=\big(\Theta_1(x),\Theta_2(x),\ldots,\Theta_d(x)\big),\\\Lambda(x)&=\big(\Lambda_1(x),\Lambda_2(x),\ldots,\Lambda_d(x)\big),\end{aligned}$$

φ, ζ, ψ and γ are the shared activation functions of the hidden units in $\mathcal{C}^2(\Omega)$, bounded and non-constant; x_j is input, β_i, β_i', β_i'', β_i''', σ_{ji}, σ_{ji}', σ_{ji}'' and σ_{ji}''' are weights, c_i, c_i', c_i'' and c_i''' are thresholds of the neural networks.

More generally, we use the similar notation

$$[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d\times[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d\times\mathfrak{C}_{P_{\rm S}}(\psi)\times\mathfrak{C}_{P_{\rm D}}(\gamma)$$

for the multi layer neural networks with arbitrarily large number of hidden units m_1, m_2, m_3 and m_4. In particular, the parameters and the activation function in each dimension of $[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}^{m_1}(\varphi)]^d$ or $[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}^{m_2}(\zeta)]^d$ are same as before. For convenient, we set n as the number of the neurons in numerical experiments. Then the parameters of the CDNNs can be formalized

as follows:

$$\begin{aligned}\theta_1^k &= (\beta_1,\ldots,\beta_n,\sigma_{11},\ldots,\sigma_{dn},c_1,\ldots,c_n),\\ \theta_2^k &= (\beta_1',\ldots,\beta_n,\sigma_{11}',\ldots,\sigma_{dn}',c_1',\ldots,c_n'),\\ \theta_3 &= (\beta_1'',\ldots,\beta_n',\sigma_{11}'',\ldots,\sigma_{dn}'',c_1'',\ldots,c_n''),\\ \theta_4 &= (\beta_1''',\ldots,\beta_n'',\sigma_{11}''',\ldots,\sigma_{dn}''',c_1''',\ldots,c_n'''),\end{aligned}$$

where $k=1,2,\ldots,d$, $\theta_1\in\mathbb{R}^{(2+d)nd}$, $\theta_2\in\mathbb{R}^{(2+d)nd}$, $\theta_3\in\mathbb{R}^{(2+d)n}$, and $\theta_4\in\mathbb{R}^{(2+d)n}$.

In the next two subsections, we prove that the neural network $\overline{\boldsymbol{U}}^n$ with n hidden units for $\boldsymbol{U}_{\rm S}^n,\boldsymbol{U}_{\rm D}^n,P_{\rm S}^n$ and $P_{\rm D}^n$ satisfy the differential operators, boundary conditions, divergence conditions and interface conditions arbitrarily well for sufficiently large n. Specifically, we confirm that there exists $\overline{\boldsymbol{U}}^n\in[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d\times[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d\times\mathfrak{C}_{P_{\rm S}}(\psi)\times\mathfrak{C}_{P_{\rm D}}(\gamma)$ satisfying $J(\overline{\boldsymbol{U}}^n)\to 0$ as $n\to\infty$. Moreover, $\overline{\boldsymbol{U}}^n\to\overline{\boldsymbol{u}}$ as $n\to\infty$, where $\overline{\boldsymbol{u}}$ is the exact solution to the coupled equations (1)–(9).

3.1. Convergence of the loss function $J(\overline{\boldsymbol{U}})$

In this subsection, we prove that the CDNNs $\overline{\boldsymbol{U}}$ can make the loss function $J(\overline{\boldsymbol{U}})$ arbitrarily small.

Assumption 1 $\nabla\boldsymbol{v}(x),\Delta\boldsymbol{v}(x)$ and $\nabla q(x)$ are locally Lipschitz with Lipschitz coefficient such that they have at most polynomial growth on $\boldsymbol{v}(x)$ and $q(x)$. Then, for some constants $0\le q_i\le\infty$ $(i=1,2,\ldots,6)$ we have

$$\begin{aligned}&|\Delta\boldsymbol{V}(x;\theta)-\Delta\boldsymbol{v}(x)|\\ &\le(|\nabla\boldsymbol{V}(x;\theta)|^{q_1/2}+|\nabla\boldsymbol{v}(x)|^{q_2/2})|\nabla\boldsymbol{V}(x;\theta)-\nabla\boldsymbol{v}(x)|, \quad (12)\\ &|\nabla Q(x;\theta)-\nabla q(x)|\\ &\le(|Q(x;\theta)|^{q_3/2}+|q(x)|^{q_4/2})|Q(x;\theta)-q(x)|, \quad (13)\\ &|\nabla\boldsymbol{V}(x;\theta)-\nabla\boldsymbol{v}(x)|\\ &\le(|\boldsymbol{V}(x;\theta)|^{q_5/2}+|\boldsymbol{v}(x)|^{q_6/2})|\boldsymbol{V}(x;\theta)-\boldsymbol{v}(x)|. \quad (14)\end{aligned}$$

Theorem 1 Under Assumption 1, there exists a neural network $\overline{\boldsymbol{U}}\in[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d\times[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d\times\mathfrak{C}_{P_{\rm S}}(\psi)\times\mathfrak{C}_{P_{\rm D}}(\gamma)$, satisfying

$$J(\overline{\boldsymbol{U}})\le C\varepsilon^2,\quad\forall\varepsilon>0, \quad (15)$$

where C depends on the data $\{\Omega_{\rm S},\Omega_{\rm D},\Gamma,\mu,\rho,\beta,\boldsymbol{K}^{-1},\omega_1,\omega_2,\omega_3,f_{\rm D},g_{\rm D},\boldsymbol{f}_{\rm S}\}$.

Proof From Theorem 3 of Ref. [29], we obtain that there exists $\overline{\boldsymbol{U}}\in[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d\times[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d\times\mathfrak{C}_{P_{\rm S}}(\psi)\times\mathfrak{C}_{P_{\rm D}}(\gamma)$, which are uniformly 2-dense on compacts of $\mathcal{C}^2(\bar\Omega_{\rm S})\times\mathcal{C}^2(\bar\Omega_{\rm D})\times\mathcal{C}^1(\bar\Omega_{\rm S})\times\mathcal{C}^1(\bar\Omega_{\rm D})$. It means that for $\overline{\boldsymbol{u}}(x)\in\mathcal{C}^2(\bar\Omega_{\rm S})\times\mathcal{C}^2(\bar\Omega_{\rm D})\times\mathcal{C}^1(\bar\Omega_{\rm S})\times\mathcal{C}^1(\bar\Omega_{\rm D}),\forall\varepsilon>0$, we confirm that

$$\begin{aligned}&\max_{a\le2}\sup_{x\in\Omega_{\rm S}}|\partial_x^a\boldsymbol{U}_{\rm S}(x;\theta_1)-\partial_x^a\boldsymbol{u}_{\rm S}(x)|<\varepsilon, \quad (16)\\ &\max_{a\le2}\sup_{x\in\Omega_{\rm D}}|\partial_x^a\boldsymbol{U}_{\rm D}(x;\theta_2)-\partial_x^a\boldsymbol{u}_{\rm D}(x)|<\varepsilon, \quad (17)\\ &\sup_{x\in\Omega_{\rm S}}|P_{\rm S}(x;\theta_3)-p_{\rm S}(x)|<\varepsilon, \quad (18)\\ &\sup_{x\in\Omega_{\rm D}}|P_{\rm D}(x;\theta_4)-p_{\rm D}(x)|<\varepsilon. \quad (19)\end{aligned}$$

Firstly, we recall the form and discuss the convergence of $J_{\Omega_{\rm S}\setminus\Gamma}(\overline{\boldsymbol{U}})$,

$$\begin{aligned}J_{\Omega_{\rm S}}(\overline{\boldsymbol{U}})=&\big\|\boldsymbol{f}_{\rm S}+\nu\Delta\boldsymbol{U}_{\rm S}(x;\theta_1)-\nabla P_{\rm S}(x;\theta_3)\big\|^2_{0,\Omega_{\rm S},\omega_1}\\ &+\big\|\nabla\cdot\boldsymbol{U}_{\rm S}(x;\theta_1)\big\|^2_{0,\Omega_{\rm S},\omega_1}+\big\|\boldsymbol{U}_{\rm S}(x;\theta_1)\big\|^2_{0,\Gamma_{\rm S},\omega_2}. \quad (20)\end{aligned}$$

According to Assumption 1, by using the Hölder inequality and Young inequality, setting conjugate numbers r_1 and r_2 such that $\frac{1}{r_1}+\frac{1}{r_2}=1$, it follows that

$$\begin{aligned}&\int_{\Omega_{\rm S}}|\Delta\boldsymbol{U}(x;\theta_1)-\Delta\boldsymbol{u}(x)|^2\,{\rm d}\omega_1(x)\\ &\le\int_{\Omega_{\rm S}}\big(|\nabla\boldsymbol{U}(x;\theta_1)|^{l_1}+|\nabla\boldsymbol{u}(x)|^{l_2}\big)\\ &\quad\times\big(\nabla\boldsymbol{U}(x;\theta_1)-\nabla\boldsymbol{u}(x)\big)^2\,{\rm d}\omega_1(x)\\ &\le\Big[\int_{\Omega_{\rm S}}\big(|\nabla\boldsymbol{U}(x;\theta_1)|^{l_1}+|\nabla\boldsymbol{u}(x)|^{l_2}\big)^{r_1}\,{\rm d}\omega_1(x)\Big]^{1/r_1}\\ &\quad\times\Big[\int_{\Omega_{\rm S}}\big(\nabla\boldsymbol{U}(x;\theta_1)-\nabla\boldsymbol{u}(x)\big)^{2r_2}\,{\rm d}\omega_1(x)\Big]^{1/r_2}\\ &\le\Big[\int_{\Omega_{\rm S}}\big(|\nabla\boldsymbol{U}(x;\theta_1)-\nabla\boldsymbol{u}(x)|^{l_1}+|\nabla\boldsymbol{u}(x)|^{l_1\vee l_2}\big)^{r_1}\,{\rm d}\omega_1(x)\Big]^{1/r_1}\\ &\quad\times\Big[\int_{\Omega_{\rm S}}\big(\nabla\boldsymbol{U}(x;\theta_1)-\nabla\boldsymbol{u}(x)\big)^{2r_2}\,{\rm d}\omega_1(x)\Big]^{1/r_2}\\ &\le C\big(\varepsilon^{l_1}+\sup_{\Omega_{\rm S}}|\nabla\boldsymbol{u}(x)|^{l_1\vee l_2}\big)\big)\varepsilon^2. \quad (21)\end{aligned}$$

Here we set $l_1\vee l_2=\max\{l_1,l_2\}$. In the same way,

$$\begin{aligned}&\int_{\Omega_{\rm S}}\Big|\nabla P_{\rm S}(x;\theta_3)-\nabla p_{\rm S}(x)\Big|^2\,{\rm d}\omega_1(x)\\ &\le\int_{\Omega_{\rm S}}\big(|P_{\rm S}(x;\theta_3)|^{l_3}+|p_{\rm S}(x)|^{l_4}\big)\\ &\quad\times\big(P_{\rm S}(x;\theta_3)-p_{\rm S}(x)\big)^2\,{\rm d}\omega_1(x)\\ &\le\Big[\int_{\Omega_{\rm S}}\big(|P_{\rm S}(x;\theta_3)|^{l_3}+|p_{\rm S}(x)|^{l_4}\big)^{r_3}\,{\rm d}\omega_1(x)\Big]^{1/r_3}\\ &\quad\times\Big[\int_{\Omega_{\rm S}}\big(P_{\rm S}(x;\theta_3)-p_{\rm S}(x)\big)^{2r_4}\,{\rm d}\omega_1(x)\Big]^{1/r_4}\\ &\le\Big[\int_{\Omega_{\rm S}}\big(|P_{\rm S}(x;\theta_3)-p_{\rm S}(x)|^{l_3}+|p_{\rm S}(x)|^{l_3\vee l_4}\big)^{r_3}\,{\rm d}\omega_1(x)\Big]^{1/r_3}\\ &\quad\times\Big[\int_{\Omega_{\rm S}}\big(P_{\rm S}(x;\theta_3)-p_{\rm S}(x)\big)^{2r_4}\,{\rm d}\omega_1(x)\Big]^{1/r_4}\\ &\le C\big(\varepsilon^{l_3}+\sup_{\Omega_{\rm S}}|p_{\rm S}(x)|^{l_3\vee l_4}\big)\big)\varepsilon^2, \quad (22)\end{aligned}$$

where $\frac{1}{r_3}+\frac{1}{r_4}=1$ and $l_3\vee l_4=\max\{l_3,l_4\}$.

For the boundary condition, we have

$$\textstyle\int_{\Gamma_{\rm S}}|\boldsymbol{U}_{\rm S}(x;\theta_1)-\boldsymbol{u}_{\rm S}(x)|^2\,{\rm d}\omega_2(x)\le C\varepsilon^2. \quad (23)$$

Owing to Eqs. (21)–(23), we conclude that

$$\begin{aligned}J_{\Omega_{\rm S}}(\overline{\boldsymbol{U}})=&\big\|\boldsymbol{f}_{\rm S}+\nu\Delta\boldsymbol{U}_{\rm S}(x;\theta_1)-\nabla P_{\rm S}(x;\theta_3)\big\|^2_{0,\Omega_{\rm S},\omega_1}\\ &+\big\|\nabla\cdot\boldsymbol{U}_{\rm S}(x;\theta_1)\big\|^2_{0,\Omega_{\rm S},\omega_1}+\big\|\boldsymbol{U}_{\rm S}(x;\theta_1)\big\|^2_{0,\Gamma_{\rm S},\omega_2}\\ \le&\ \nu\big\|\Delta\boldsymbol{U}_{\rm S}(x;\theta_1)-\Delta\boldsymbol{u}_{\rm S}(x)\big\|^2_{0,\Omega_{\rm S},\omega_1}\end{aligned}$$

$$
\begin{aligned}
&+\left\|\nabla P_{\mathrm{S}}(x;\theta_3)-\nabla p_{\mathrm{S}}(x)\right\|^2_{0,\Omega_{\mathrm{S}},\omega_1}\\
&+\left\|\nabla\cdot\boldsymbol{U}_{\mathrm{S}}(x;\theta_1)\right\|^2_{0,\Omega_{\mathrm{S}},\omega_1}+\left\|\boldsymbol{U}(x;\theta_1)\right\|^2_{0,\Gamma_{\mathrm{S}},\omega_2}\\
\le\ & \nu\int_{\Omega_{\mathrm{S}}}|\Delta\boldsymbol{U}_{\mathrm{S}}(x;\theta_1)-\Delta\boldsymbol{u}_{\mathrm{S}}(x)|^2\,\mathrm{d}\omega_1(x)\\
&+\int_{\Omega_{\mathrm{S}}}|\nabla P_{\mathrm{S}}(x;\theta_3)-\nabla p_{\mathrm{S}}(x)|^2\,\mathrm{d}\omega_1(x)\\
&+\int_{\Omega_{\mathrm{S}}}\left|\nabla\cdot\big(\boldsymbol{U}_{\mathrm{S}}(x;\theta_1)-\boldsymbol{u}_{\mathrm{S}}(x)\big)\right|^2\mathrm{d}\omega_1(x)\\
&+\int_{\Gamma_{\mathrm{S}}}|\boldsymbol{U}_{\mathrm{S}}(x;\theta_1)-\boldsymbol{u}_{\mathrm{S}}(x)|^2\,\mathrm{d}\omega_2(x)\le C\varepsilon^2. \qquad (24)
\end{aligned}
$$

Next, we remain to prove the convergence of $J_{\Omega_{\mathrm{D}}}(\overline{U})$ and $J_{\Gamma}(\overline{U})$. We know that

$$
\begin{aligned}
J_{\Omega_{\mathrm{D}}}(\overline{U})\ =\ & \left\|f_{\mathrm{D}}-\nabla\cdot\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\right\|^2_{0,\Omega_{\mathrm{D}},\omega_1}+\left\|P_{\mathrm{D}}(x;\theta_4)\right\|^2_{0,\Gamma_{\mathrm{D}},\omega_2}\\
&+\Big\|\frac{\mu}{\rho}K^{-1}\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)+\frac{\beta}{\rho}\left|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\right|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\\
&+\nabla P_{\mathrm{D}}(x;\theta_4)-\boldsymbol{g}_{\mathrm{D}}\Big\|^2_{0,\Omega_{\mathrm{D}},\omega_1}. \qquad (25)
\end{aligned}
$$

From Eq. (19), we have

$$
\begin{aligned}
&\int_{\Omega_{\mathrm{D}}}\left|\nabla P_{\mathrm{D}}(x;\theta_4)-\nabla p_{\mathrm{D}}(x)\right|^2\mathrm{d}\omega_1(x)\\
\le\ & \int_{\Omega_{\mathrm{D}}}\left(|P_{\mathrm{D}}(x;\theta_4)|^{l_5}+|p_{\mathrm{D}}(x)|^{l_6}\right)\\
&\times\big(P_{\mathrm{D}}(x;\theta_4)-p_{\mathrm{D}}(x)\big)^2\,\mathrm{d}\omega_1(x), \qquad (26)
\end{aligned}
$$

which can be updated by using the Hölder inequality and Young inequality, thus we have

$$
\begin{aligned}
&\Big[\int_{\Omega_{\mathrm{D}}}\left(|P_{\mathrm{D}}(x;\theta_4)|^{l_5}+|p_{\mathrm{D}}|^{l_6}\right)^{r_5}\mathrm{d}\omega_1(x)\Big]^{1/r_5}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}(P_{\mathrm{D}}(x;\theta_4)-p_{\mathrm{D}}(x))|^{2r_6}\,\mathrm{d}\omega_1(x)\Big]^{1/r_6}\\
\le\ & \Big[\int_{\Omega_{\mathrm{D}}}\left(|P_{\mathrm{D}}(x;\theta_4)-p_{\mathrm{D}}(x)|^{l_5}+|p_{\mathrm{D}}(x)|^{l_5\vee l_6}\right)^{r_5}\mathrm{d}\omega_1(x)\Big]^{1/r_5}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\big(P_{\mathrm{D}}(x;\theta_4)-p_{\mathrm{D}}(x)\big)^{2r_6}\,\mathrm{d}\omega_1(x)\Big]^{1/r_6}\\
\le\ & C\big(\varepsilon^{l_5}+\sup_{\Omega_{\mathrm{D}}}|p_{\mathrm{D}}(x)|^{l_5\vee l_6}\big)\varepsilon^2, \qquad (27)
\end{aligned}
$$

where $\frac{1}{r_5}+\frac{1}{r_6}=1$ and $l_5\vee l_6=\max\{l_5,l_6\}$.

Next, we prove the boundedness of term $|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)$,

$$
\begin{aligned}
&\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-|\boldsymbol{u}_{\mathrm{D}}(x)|\boldsymbol{u}_{\mathrm{D}}(x)\right)^2\mathrm{d}\omega_1(x)\\
=\ & \int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|-|\boldsymbol{u}_{\mathrm{D}}(x)|\right)\\
&+|\boldsymbol{u}_{\mathrm{D}}(x)|\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)\big)^2\,\mathrm{d}\omega_1(x)\\
=\ & \int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|-|\boldsymbol{u}_{\mathrm{D}}(x)|\right)\big)^2\,\mathrm{d}\omega_1(x)\\
&+\int_{\Omega_{\mathrm{D}}}\big(|\boldsymbol{u}_{\mathrm{D}}(x)|\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)\big)^2\,\mathrm{d}\omega_1(x)\\
&+2\int_{\Omega_{\mathrm{D}}}\big(|\boldsymbol{u}_{\mathrm{D}}(x)|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|-|\boldsymbol{u}_{\mathrm{D}}(x)|\right)\\
&\times\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)\,\mathrm{d}\omega_1(x), \qquad (28)
\end{aligned}
$$

where

$$
\begin{aligned}
&\int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|-|\boldsymbol{u}_{\mathrm{D}}(x)|\right)\big)^2\,\mathrm{d}\omega_1(x)\\
\le\ & \Big[\int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\big)^{2r_7}\,\mathrm{d}\omega_1(x)\Big]^{1/r_7}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|+|\boldsymbol{u}_{\mathrm{D}}(x)|\right)^{2r_8}\mathrm{d}\omega_1(x)\Big]^{1/r_8}\\
\le\ & \Big[\int_{\Omega_{\mathrm{D}}}\big(\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)+\boldsymbol{u}_{\mathrm{D}}(x)\big)^{2r_7}\,\mathrm{d}\omega_1(x)\Big]^{1/r_7}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)|+|\boldsymbol{u}_{\mathrm{D}}(x)|\right)^{2r_8}\mathrm{d}\omega_1(x)\Big]^{1/r_8},
\end{aligned} \qquad (29)
$$

by using the Hölder inequality and Young inequality, setting conjugate numbers r_7 and r_8 such that $\frac{1}{r_7}+\frac{1}{r_8}=1$.

Similarly, we can obtain

$$
\begin{aligned}
&\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{u}_{\mathrm{D}}(x)|\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)\right)^2\mathrm{d}\omega_1(x)\\
\le\ & \Big[\int_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|^{r_9}\,\mathrm{d}\omega_1(x)\Big]^{1/r_9}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)^{2r_{10}}\,\mathrm{d}\omega_1(x)\Big]^{1/r_{10}}, \qquad (30)
\end{aligned}
$$

where $\frac{1}{r_9}+\frac{1}{r_{10}}=1$. Furthermore, we set $\frac{1}{r_{11}}+\frac{1}{r_{12}}=1$ and $\frac{1}{r_{13}}+\frac{1}{r_{14}}=1$,

$$
\begin{aligned}
&\int_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|-|\boldsymbol{u}_{\mathrm{D}}(x)|\right)\\
&\times\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)\,\mathrm{d}\omega_1(x)\\
\le\ & \Big[\int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|-|\boldsymbol{u}_{\mathrm{D}}(x)|\right)\big)^{r_{11}}\,\mathrm{d}\omega_1(x)\Big]^{1/r_{11}}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{u}_{\mathrm{D}}(x)|\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)\right)^{r_{12}}\mathrm{d}\omega_1(x)\Big]^{1/r_{12}}\\
\le\ & \Big[\int_{\Omega_{\mathrm{D}}}\big(\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)+\boldsymbol{u}_{\mathrm{D}}(x)\big)^{r_{11}r_{13}}\,\mathrm{d}\omega_1(x)\Big]^{1/r_{11}r_{13}}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)|+|\boldsymbol{u}_{\mathrm{D}}(x)|\right)^{r_{11}r_{13}}\mathrm{d}\omega_1(x)\Big]^{1/r_{11}r_{13}}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|^{r_{12}r_{14}}\,\mathrm{d}\omega_1(x)\Big]^{1/r_{12}r_{14}}\\
&\times\Big[\int_{\Omega_{\mathrm{D}}}\big(\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)\big)^{r_{12}}\,\mathrm{d}\omega_1(x)\Big]^{1/r_{12}}. \qquad (31)
\end{aligned}
$$

According to the inequalities (28)–(31), we can conclude

$$
\begin{aligned}
&\int_{\Omega_{\mathrm{D}}}\left(|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-|\boldsymbol{u}_{\mathrm{D}}(x)|\boldsymbol{u}_{\mathrm{D}}(x)\right)^2\mathrm{d}\omega_1(x)\\
\le\ & (\varepsilon^2+\sup_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|^2)^2+\sup_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|\varepsilon^2\\
&+2\varepsilon\sup_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|(\varepsilon+\sup_{\Omega_{\mathrm{D}}}|\boldsymbol{u}_{\mathrm{D}}(x)|)^2. \qquad (32)
\end{aligned}
$$

For the boundary condition, we know

$$
\int_{\Gamma_{\mathrm{D}}}|\boldsymbol{U}_{\mathrm{D}}(x;\theta_2)-\boldsymbol{u}_{\mathrm{D}}(x)|^2\,\mathrm{d}\omega_2(x)\le C\varepsilon^2. \qquad (33)
$$

Combining Eqs. (26)–(33), we obtain

$$\begin{aligned}
J_{\Omega_{\rm D}}(\overline{U}) =& \left\|f_{\rm D}-\nabla\cdot\boldsymbol{U}_{\rm D}(x;\theta_2)\right\|^2_{0,\Omega_{\rm D},\omega_1}+\left\|P_{\rm D}(x;\theta_4)\right\|^2_{0,\Gamma_{\rm D},\omega_2}\\
&+\left\|\frac{\mu}{\rho}K^{-1}\boldsymbol{U}_{\rm D}(x;\theta_2)+\frac{\beta}{\rho}\left|\boldsymbol{U}_{\rm D}(x;\theta_2)\right|\boldsymbol{U}_{\rm D}(x;\theta_2)\right.\\
&\left.+\nabla P_{\rm D}(x;\theta_4)-\boldsymbol{g}_{\rm D}\right\|^2_{0,\Omega_{\rm D},\omega_1}\\
\le&\int_{\Omega_{\rm D}}\left|\nabla\cdot\left(\boldsymbol{U}_{\rm D}(x;\theta_2)-\boldsymbol{u}_{\rm D}(x)\right)\right|^2{\rm d}\omega_1(x)\\
&+\int_{\Gamma_{\rm D}}\left(P_{\rm D}(x;\theta_4)-p_{\rm D}(x)\right)^2{\rm d}\omega_2(x)\\
&+\frac{\beta}{\rho}\int_{\Omega_{\rm D}}\left(\left|\boldsymbol{U}_{\rm D}(x;\theta_2)\right|\boldsymbol{U}_{\rm D}(x;\theta_2)\right.\\
&\left.-\left|\boldsymbol{u}_{\rm D}(x)\right|\boldsymbol{u}_{\rm D}(x)\right){\rm d}\omega_1(x)\\
&+\frac{\mu}{\rho}K^{-1}\int_{\Omega_{\rm D}}\left(\boldsymbol{U}_{\rm D}(x;\theta_2)-\boldsymbol{u}_{\rm D}(x)\right)^2{\rm d}\omega_1(x)\\
&+\int_{\Omega_{\rm D}}\left(\nabla P_{\rm D}(x;\theta_4)-\nabla p_{\rm D}(x)\right)^2{\rm d}\omega_1(x)\\
\le& C\varepsilon^2. \qquad (34)
\end{aligned}$$

The loss in interface is referred in Eq. (11). According to Assumption 3.1, we know

$$\begin{aligned}
&\int_{\Gamma}\left|\nabla\boldsymbol{U}_{\rm D}(x;\theta_2)-\nabla\boldsymbol{u}_{\rm D}(x)\right|^2{\rm d}\omega_3(x)\\
\le&\int_{\Gamma}\left(|\boldsymbol{U}_{\rm D}(x;\theta_2)|^{l_7}+|\boldsymbol{u}_{\rm D}(x)|^{l_8}\right)\\
&\times\left(\boldsymbol{U}_{\rm D}(x;\theta_2)-\boldsymbol{u}_{\rm D}(x)\right)^2{\rm d}\omega_3(x), \qquad (35)
\end{aligned}$$

which can be updated by using the Hölder inequality and Young inequality, thus we have

$$\begin{aligned}
&\left[\int_{\Gamma}\left(|\boldsymbol{U}_{\rm D}(x;\theta_3)|^{l_7}+|\boldsymbol{u}_{\rm D}|^{l_8}\right)^{r_{15}}{\rm d}\omega_3(x)\right]^{1/r_{15}}\\
&\times\left[\int_{\Gamma}(\boldsymbol{U}_{\rm D}(x;\theta_3)-\boldsymbol{u}_{\rm D}(x))|^{2r_{16}}{\rm d}\omega_3(x)\right]^{1/r_{16}}\\
\le&\left[\int_{\Gamma}\left(|\boldsymbol{U}_{\rm D}(x;\theta_3)-\boldsymbol{u}_{\rm D}(x)|^{l_7}+|\boldsymbol{u}_{\rm D}(x)|^{l_7\vee l_8}\right)^{r_{15}}{\rm d}\omega_3(x)\right]^{1/r_{15}}\\
&\times\left[\int_{\Gamma}\left(\boldsymbol{U}_{\rm D}(x;\theta_3)-\boldsymbol{u}_{\rm D}(x)\right)^{2r_{16}}{\rm d}\omega_3(x)\right]^{1/r_{16}}\\
\le& C\big(\varepsilon^{l_7}+\sup_{\Gamma}|\boldsymbol{u}_{\rm D}(x)|^{l_7\vee l_8})\big)\varepsilon^2, \qquad (36)
\end{aligned}$$

where $\frac{1}{r_{15}}+\frac{1}{r_{16}}=1$ and $l_7\vee l_8=\max\{l_7,l_8\}$.

Above all, we can obtain

$$\begin{aligned}
J_{\Gamma}(\overline{U}) =& \left\|\boldsymbol{U}_{\rm S}(x;\theta_1)\cdot\boldsymbol{n}_{\rm S}-\boldsymbol{U}_{\rm D}(x;\theta_2)\cdot\boldsymbol{n}_{\rm S}\right\|^2_{0,\Gamma,\omega_3}\\
&+\left\|P_{\rm S}(x;\theta_3)-\nu\boldsymbol{n}_{\rm S}\frac{\partial\boldsymbol{U}_{\rm S}(x;\theta_1)}{\partial\boldsymbol{n}_{\rm S}}-P_{\rm D}(x;\theta_4)\right\|^2_{0,\Gamma,\omega_3}\\
&+\left\|-\nu\boldsymbol{t}\frac{\partial\boldsymbol{U}_{\rm S}(x;\theta_1)}{\partial\boldsymbol{n}_{\rm S}}-G\boldsymbol{U}_{\rm S}(x;\theta_1)\cdot\boldsymbol{t}\right\|^2_{0,\Gamma,\omega_3}\\
\le&\left\|\boldsymbol{U}_{\rm S}(x;\theta_2)-\boldsymbol{u}_{\rm S}(x)\right\|^2_{0,\Gamma,\omega_3}\left\|\boldsymbol{n}_{\rm S}\right\|^2_{0,\Gamma,\omega_3}\\
&+\left\|\boldsymbol{U}_{\rm D}(x;\theta_2)-\boldsymbol{u}_{\rm D}(x)\right\|^2_{0,\Gamma,\omega_3}\left\|\boldsymbol{n}_{\rm S}\right\|^2_{0,\Gamma,\omega_3}\\
&+\left\|P_{\rm S}(x;\theta_3)-p_{\rm S}(x)\right\|^2_{0,\Gamma,\omega_3}\\
&+\left\|P_{\rm D}(x;\theta_4)-p_{\rm D}(x)\right\|^2_{0,\Gamma,\omega_3}\\
&+\nu\left\|\boldsymbol{n}_{\rm S}\right\|^4_{0,\Gamma,\omega_3}\left\|\nabla\boldsymbol{U}_{\rm S}(x;\theta_1)-\nabla\boldsymbol{u}_{\rm S}(x;\theta_1)\right\|^2_{0,\Gamma,\omega_3}\\
&+G\left\|\boldsymbol{t}\right\|^2_{0,\Gamma,\omega_3}\left\|\boldsymbol{U}_{\rm S}(x;\theta_1)-\boldsymbol{u}_{\rm S}(x)\right\|^2_{0,\Gamma,\omega_3}\\
&+\nu\left\|\boldsymbol{t}\right\|^2_{0,\Gamma,\omega_3}\left\|\nabla\boldsymbol{U}_{\rm S}(x;\theta_1)\right.\\
&\left.-\nabla\boldsymbol{u}_{\rm S}(x;\theta_1)\right\|^2_{0,\Gamma,\omega_3}\left\|\boldsymbol{n}_{\rm S}\right\|^2_{0,\Gamma,\omega_3}\le C\varepsilon^2, \qquad (37)
\end{aligned}$$

which completes the proof.

3.2. Convergence of the CDNNs to the exact solution

In the above subsection, we have proved the convergence of the loss function. In this subsection, we remain to discuss the convergence of the CDNNs to the exact solution. According to the Galerkin method, the neural networks satisfy

$$\begin{aligned}
\nabla\cdot\boldsymbol{U}^n_{\rm D}-f_{\rm D}&=0, \quad \text{in } \Omega_{\rm D}, \qquad (38)\\
\frac{\mu}{\rho}K^{-1}\boldsymbol{U}^n_{\rm D}+\frac{\beta}{\rho}\mid\boldsymbol{U}^n_{\rm D}\mid\boldsymbol{U}^n_{\rm D}+\nabla P^n_{\rm D}-\boldsymbol{g}_{\rm D}&=0, \quad \text{in } \Omega_{\rm D}, \qquad (39)\\
P^n_{\rm D}&=0, \quad \text{on } \Gamma_{\rm D}, \qquad (40)\\
-\nu\Delta\boldsymbol{U}^n_{\rm S}+\nabla P^n_{\rm S}-\boldsymbol{f}_{\rm S}&=0, \quad \text{in } \Omega_{\rm S}, \qquad (41)\\
\nabla\cdot\boldsymbol{U}^n_{\rm S}&=0, \quad \text{in } \Omega_{\rm S}, \qquad (42)\\
\boldsymbol{U}^n_{\rm S}&=0, \quad \text{on } \Gamma_{\rm S}, \qquad (43)\\
\boldsymbol{U}^n_{\rm S}\cdot\boldsymbol{n}_{\rm S}-\boldsymbol{U}^n_{\rm D}\cdot\boldsymbol{n}_{\rm S}&=0, \quad \text{on } \Gamma, \qquad (44)\\
P^n_{\rm S}-\nu\boldsymbol{n}_{\rm S}\frac{\partial\boldsymbol{U}^n_{\rm S}}{\partial\boldsymbol{n}_{\rm S}}-P^n_{\rm D}&=0, \quad \text{on } \Gamma, \qquad (45)\\
-\nu\boldsymbol{t}\frac{\partial\boldsymbol{U}^n_{\rm S}}{\partial\boldsymbol{n}_{\rm S}}-G\boldsymbol{U}^n_{\rm S}\cdot\boldsymbol{t}&=0, \quad \text{on } \Gamma. \qquad (46)
\end{aligned}$$

Based on the above system of equations, we give the following assumption and theorem to guarantee the convergency of the CDNNs to the exact solution.

Assumption 2 We assume $(\boldsymbol{u}_{\rm S},\boldsymbol{u}_{\rm D})\in C^{\xi}(\bar{\Omega}_{\rm S})\times C^{\xi}(\bar{\Omega}_{\rm D})$, where $\xi>2$ with itself and its first derivative bounded in $\bar{\Omega}_{\rm S}\times\bar{\Omega}_{\rm D}$. Moreover, for every $n\in N$, $\boldsymbol{U}^n_{\rm S}\times\boldsymbol{U}^n_{\rm D}\in C^{1,2}(\bar{\Omega}_{\rm S})\cap\mathbb{X}_{\rm S}\times C^{1,2}(\bar{\Omega}_{\rm D})\cap L^3(\Omega_{\rm D})^2$, we assume that the subspaces $\mathbb{X}^n_{\rm S}\times L^3(\Omega_{\rm D})^2\times L^2(\Omega_{\rm S})\times\mathbb{Y}^n_{\rm D}\subset[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d\times[\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d\times\mathfrak{C}_{P_{\rm S}}(\psi)\times\mathfrak{C}_{P_{\rm D}}(\gamma)$ satisfy the discrete *inf-sup* conditions.

Theorem 2 Under Assumptions 1–2 and Theorem 1, the neural network $\boldsymbol{U}^n_{\rm S}$ can converge strongly to $\boldsymbol{u}_{\rm S}$ in L^2; $P^n_{\rm S}$, $\boldsymbol{U}^n_{\rm D}$ and $P^n_{\rm D}$ can converge strongly to $p_{\rm S},\boldsymbol{u}_{\rm D}$ and $p_{\rm D}$ in H^{-1}. In addition, if the sequences $\{\boldsymbol{U}^n_{\rm S}\}_{n\in\mathbb{N}},\{P^n_{\rm S}\}_{n\in\mathbb{N}},\{\boldsymbol{U}^n_{\rm D}\}_{n\in\mathbb{N}}$ and $\{P^n_{\rm D}\}_{n\in\mathbb{N}}$ are uniformly bounded and equicontinuous in $\Omega_{\rm S}$ and $\Omega_{\rm D}$, they can converge uniformly to $\boldsymbol{u}_{\rm S}$, $p_{\rm S},\boldsymbol{u}_{\rm D}$ and $p_{\rm D}$, respectively.

Proof Firstly, we give the weak formulation for Eqs. (38)–(46). Multiplying Eq. (41) by $\boldsymbol{V}^n_{\rm S}\in\mathbb{X}_{\rm S}\bigcap[\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d$ and Eq. (42) by $Q^n_{\rm S}\in\mathfrak{C}_{P_{\rm S}}(\psi)$ and integration by parts yields

$$\begin{aligned}
&\nu(\nabla\boldsymbol{U}^n_{\rm S},\nabla\boldsymbol{V}^n_{\rm S})_{\Omega_{\rm S}}-\nu\langle\nabla\boldsymbol{U}^n_{\rm S}\cdot\boldsymbol{n}_{\rm S},\boldsymbol{V}^n_{\rm S}\rangle_{\Gamma}\\
&-(P^n_{\rm S},\nabla\cdot\boldsymbol{V}^n_{\rm S})_{\Omega_{\rm S}}+\langle P^n_{\rm S},\boldsymbol{V}^n_{\rm S}\cdot\boldsymbol{n}_{\rm S}\rangle_{\Gamma}=(\boldsymbol{f}_{\rm S},\boldsymbol{V}^n_{\rm S})_{\Omega_{\rm S}}, \qquad (47)\\
&(\nabla\cdot\boldsymbol{U}^n_{\rm S},Q^n_{\rm S})_{\Omega_{\rm S}}=0. \qquad (48)
\end{aligned}$$

Multiplying Eq. (38) by $Q_{\rm D}^n \in \mathbb{Y}_{\rm D} \bigcap \mathfrak{C}_{P_{\rm D}}(\gamma)$ and Eq. (39) by $\boldsymbol{V}_{\rm D}^n \in [\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d$, it then follows from integration by parts that

$$-(\boldsymbol{U}_{\rm D}^n, \nabla \boldsymbol{Q}_{\rm D}^n)_{\Omega_{\rm D}} + \langle \boldsymbol{U}_{\rm D}^n \cdot \boldsymbol{n}_{\rm D}, \boldsymbol{Q}_{\rm D}^n \rangle_\Gamma = (f_{\rm D}, \boldsymbol{Q}_{\rm D}^n)_{\Omega_{\rm D}}, \tag{49}$$

$$\frac{\mu}{\rho}(K^{-1}\boldsymbol{U}_{\rm D}^n, \boldsymbol{V}_{\rm D}^n)_{\Omega_{\rm D}} + \frac{\beta}{\rho}(|\boldsymbol{U}_{\rm D}^n|\boldsymbol{U}_{\rm D}^n, \boldsymbol{V}_{\rm D}^n)_{\Omega_{\rm D}} - (P_{\rm D}^n, \nabla \cdot \boldsymbol{V}_{\rm D}^n)_{\Omega_{\rm D}} + \langle P_{\rm D}^n, \boldsymbol{V}_{\rm D}^n \cdot \boldsymbol{n}_{\rm D} \rangle_\Gamma = (g_{\rm D}, \boldsymbol{V}_{\rm D}^n)_{\Omega_{\rm D}}. \tag{50}$$

Considering the interface conditions, simple algebraic calculation yields

$$\begin{aligned}\langle \nabla \boldsymbol{U}_{\rm S}^n \boldsymbol{n}_{\rm S}, \boldsymbol{V}_{\rm S}^n \rangle_\Gamma \;=\; & \langle \boldsymbol{n}_{\rm S} \nabla \boldsymbol{U}_{\rm S}^n \cdot \boldsymbol{n}_{\rm S}, \boldsymbol{V}_{\rm S}^n \cdot \boldsymbol{n}_{\rm S} \rangle_\Gamma \\ & + \langle \boldsymbol{n}_{\rm S} \nabla \boldsymbol{U}_{\rm S}^n \cdot \boldsymbol{t}, \boldsymbol{V}_{\rm S}^n \cdot \boldsymbol{t} \rangle_\Gamma,\end{aligned}$$

which, by employing interface conditions (45) and (46), gives

$$\begin{aligned}& \langle (P_{\rm S}^n I - \nu \nabla \boldsymbol{U}_{\rm S}^n) \boldsymbol{n}_{\rm S}, \boldsymbol{V}_{\rm S}^n \rangle_\Gamma \\ =\; & \langle P_{\rm D}^n, \boldsymbol{V}_{\rm S}^n \cdot \boldsymbol{n}_{\rm S} \rangle_\Gamma + G \langle \boldsymbol{U}_{\rm S}^n \cdot \boldsymbol{t}, \boldsymbol{V}_{\rm S}^n \cdot \boldsymbol{t} \rangle_\Gamma.\end{aligned} \tag{51}$$

For convenience of presentation, we introduce the nonlinear operator $A : L^3(\Omega_{\rm D})^2 \to L^{3/2}(\Omega_{\rm D})^2$ defined by

$$A(\boldsymbol{V}) = \frac{\mu K^{-1}}{\rho} \boldsymbol{V} + \frac{\beta}{\rho} |\boldsymbol{V}| \boldsymbol{V}. \tag{52}$$

The definition of Eq. (52) gives

$$(A(\boldsymbol{U}_{\rm D}^n), \boldsymbol{U}_{\rm D}^n)_{\Omega_{\rm D}} \geq C(\|\boldsymbol{U}_{\rm D}^n\|_0^2 + \|\boldsymbol{U}_{\rm D}^n\|_{L^3}^3).$$

According to Assumption 2 and Eq. (50),

$$\begin{aligned}\|P_{\rm D}^n\|_0 \;\leq\; & C \sup_{\boldsymbol{V}_{\rm D}^n \in L^3(\Omega_{\rm D})^2 \bigcap [\mathfrak{C}_{\boldsymbol{U}_{\rm D}}(\zeta)]^d} \\ & \frac{(A(\boldsymbol{U}_{\rm D}^n), \boldsymbol{V}_{\rm D}^n)_{\Omega_{\rm D}} - (g_{\rm D}, \boldsymbol{V}_{\rm D}^n)_{\Omega_{\rm D}}}{\|\boldsymbol{V}_{\rm D}^n\|_{L^3(\Omega_{\rm D})}} \\ \leq\; & C(\|\boldsymbol{U}_{\rm D}^n\|_0 + \|\boldsymbol{U}_{\rm D}^n\|_{L^3}^2 + \|\boldsymbol{g}_{\rm D}\|_0).\end{aligned}$$

Taking $\boldsymbol{V}_{\rm S}^n = \boldsymbol{U}_{\rm S}^n, Q_{\rm S}^n = P_{\rm S}^n, \boldsymbol{V}_{\rm D}^n = \boldsymbol{U}_{\rm D}^n$ and $Q_{\rm D}^n = P_{\rm D}^n$ in Eqs. (47)–(50) and adding the resulting equation (51) yields

$$\begin{aligned}& \nu(\nabla \boldsymbol{U}_{\rm S}^n, \nabla \boldsymbol{U}_{\rm S}^n)_{\Omega_{\rm S}} + G\langle \boldsymbol{U}_{\rm S}^n \cdot \boldsymbol{t}, \boldsymbol{U}_{\rm S}^n \cdot \boldsymbol{t} \rangle_\Gamma + \frac{1}{2}(A(\boldsymbol{U}_{\rm D}^n), \boldsymbol{U}_{\rm D}^n)_{\Omega_{\rm D}} \\ =\; & (\boldsymbol{f}_{\rm S}, \boldsymbol{U}_{\rm S}^n)_{\Omega_{\rm S}} + (f_{\rm D}, P_{\rm D}^n)_{\Omega_{\rm D}} + \frac{1}{2}(\boldsymbol{g}_{\rm D}, \boldsymbol{U}_{\rm D}^n)_{\Omega_{\rm D}}.\end{aligned}$$

According to the definition of the H^1 norm, the Young inequality and the Poincaré inequality, we can obtain

$$\begin{aligned}& \|\boldsymbol{U}_{\rm S}^n\|_1^2 + \|\boldsymbol{U}_{\rm D}^n\|_0^2 + \|\boldsymbol{U}_{\rm D}^n\|_{L^3}^3 \\ \leq\; & C(\|\boldsymbol{f}_{\rm S}\|_0 \|\boldsymbol{U}_{\rm S}^n\|_0 + \|f_{\rm D}\|_0 \|P_{\rm D}^n\|_0 + \|\boldsymbol{g}_{\rm D}\|_0 \|\boldsymbol{U}_{\rm D}^n\|_0) \\ \leq\; & C(\|\boldsymbol{f}_{\rm S}\|_0 \|\boldsymbol{U}_{\rm S}^n\|_0 + \|f_{\rm D}\|_0 (\|\boldsymbol{U}_{\rm D}^n\|_0 \\ & + \|\boldsymbol{U}_{\rm D}^n\|_{L^3}^2 + \|\boldsymbol{g}_{\rm D}\|_0) + \|\boldsymbol{g}_{\rm D}\|_0 \|\boldsymbol{U}_{\rm D}^n\|_0) \\ \leq\; & C(\|f_{\rm D}\|_0^2 + \|f_{\rm D}\|_0^3 + \|\boldsymbol{f}_{\rm S}\|_0^2 + \|\boldsymbol{g}_{\rm D}\|_0^2).\end{aligned}$$

Thus we have

$$\begin{aligned}\|P_{\rm D}^n\|_0 \;\leq\; & C(\|\boldsymbol{U}_{\rm D}^n\|_0 + \|\boldsymbol{U}_{\rm D}\|_{L^3}^2 + \|\boldsymbol{g}_{\rm D}\|_0) \\ \leq\; & C(\|f_{\rm D}\|_0 + \|f_{\rm D}\|_0^2 + \|\boldsymbol{f}_{\rm S}\|_0 \\ & + \|\boldsymbol{f}_{\rm S}\|_0^2 + \|\boldsymbol{g}_{\rm D}\|_0 + \|\boldsymbol{g}_{\rm D}\|_0^2).\end{aligned}$$

According to Assumption 2, Eqs. (47)–(48) and Eq. (51), we have

$$\begin{aligned}\|P_{\rm S}^n\|_0 \;\leq\; & C \sup_{\boldsymbol{V}_{\rm S}^n \in [\mathbb{X}_{\rm S}]^2 \bigcap [\mathfrak{C}_{\boldsymbol{U}_{\rm S}}(\varphi)]^d} \frac{(\boldsymbol{f}_{\rm S}, \boldsymbol{V}_{\rm S}^n) - (\nabla \boldsymbol{U}_{\rm S}^n, \nabla \boldsymbol{V}_{\rm S}^n) - \langle P_{\rm D}^n, \boldsymbol{V}_{\rm S}^n \cdot \boldsymbol{n}_{\rm S} \rangle_\Gamma - G\langle \boldsymbol{U}_{\rm S}^n \cdot \boldsymbol{t}, \boldsymbol{V}_{\rm S}^n \cdot \boldsymbol{t} \rangle_\Gamma}{\|\nabla \boldsymbol{V}_{\rm S}^n\|_0} \\ \leq\; & C(\|\boldsymbol{f}_{\rm S}\|_0 + \|\nabla \boldsymbol{U}_{\rm S}^n\|_0 + \|P_{\rm D}^n\|_0 + \|\boldsymbol{U}_{\rm S}^n\|_0) \leq C(\|f_{\rm D}\|_0 + \|f_{\rm D}\|_0^2 + \|\boldsymbol{f}_{\rm S}\|_0 + \|\boldsymbol{f}_{\rm S}\|_0^2 + \|\boldsymbol{g}_{\rm D}\|_0 + \|\boldsymbol{g}_{\rm D}\|_0^2).\end{aligned}$$

Up to now, we obtain:

$$\{\boldsymbol{U}_{\rm S}^n\}_{n \in N} \text{ is uniformly bounded in } H^1(\Omega_{\rm S}). \tag{53}$$

By using the uniformly boundedness of $\boldsymbol{U}_{\rm S}^n$, we can extract a subsequence $\{\boldsymbol{U}_{\rm S}^n\}_{n \in \mathbb{N}}$ of $\boldsymbol{U}_{\rm S}^n$, which converge weakly in $H^1(\Omega_{\rm S})$. Due to the compact embedding $H^1(\Omega_{\rm S}) \hookrightarrow L^2(\Omega_{\rm S})$, we have $\lim\limits_{n \to \infty} \| \boldsymbol{U}_{\rm S}^n - \boldsymbol{u}_{\rm S} \|_{0,\Omega_{\rm S}} = 0$.

Similarly, we know $L^3(\Omega_{\rm D}) \subset L^2(\Omega_{\rm D})$, which implies that

$$\{\boldsymbol{U}_{\rm D}^n\}_{n \in N} \text{ is uniformly bounded in } L^2(\Omega_{\rm D}). \tag{54}$$

Due to the compact embedding $L^2(\Omega_{\rm D}) \hookrightarrow H^{-1}(\Omega_{\rm D})$, we have $\lim\limits_{n \to \infty} \| \boldsymbol{U}_{\rm D}^n - \boldsymbol{u}_{\rm D} \|_{-1,\Omega_{\rm D}} = 0$.

The convergence of $P_{\rm S}^n$ and $P_{\rm D}^n$ is given as follows:

$$\lim_{n \to \infty} \| P_{\rm S}^n - P_{\rm S} \|_{-1,\Omega_{\rm S}} = 0, \quad \lim_{n \to \infty} \| P_{\rm D}^n - p_{\rm D} \|_{-1,\Omega_{\rm D}} = 0.$$

For all these reasons, $\{\boldsymbol{U}_{\rm S}^n\}_{n \in \mathbb{N}}$ converges strongly to $\boldsymbol{u}_{\rm S}$ in L^2, $\{P_{\rm S}^n\}_{n \in \mathbb{N}}$, $\{U_{\rm D}^n\}_{n \in \mathbb{N}}$ and $\{P_{\rm D}^n\}_{n \in \mathbb{N}}$ converge strongly to $p_{\rm S}$, $\boldsymbol{u}_{\rm D}$ and $p_{\rm D}$ in H^{-1}. More generally, by the well known Arzelà–Ascoli theorem we conclude that $\{\boldsymbol{U}_{\rm S}^n\}_{n \in \mathbb{N}}$, $\{P_{\rm S}^n\}_{n \in \mathbb{N}}$, $\{\boldsymbol{U}_{\rm D}^n\}_{n \in \mathbb{N}}$ and $\{P_{\rm D}^n\}_{n \in \mathbb{N}}$ converge uniformly to $\boldsymbol{u}_{\rm S}$, $p_{\rm S}$, $\boldsymbol{u}_{\rm D}$ and $p_{\rm D}$, respectively.

4. Numerical experiments

The section presents several numerical tests to confirm the proposed theoretical results. We start with three examples with known exact solution to test the efficiency of the proposed method, where the permeability for the third example is highly oscillatory. Then, the fourth example with no exact solution shows the application of the proposed method to high contrast permeability problem. This section concludes with a physical

flow. The numerical examples presented could violate the interface conditions (8) and (9),[56] that is, Eqs. (8) and (9) are replaced by Eqs. (55) and (56) to deal with this case,

$$p_{\mathrm{S}} - \nu \boldsymbol{n}_{\mathrm{S}} \frac{\partial \boldsymbol{u}_{\mathrm{S}}}{\partial \boldsymbol{n}_{\mathrm{S}}} = p_{\mathrm{D}} + g_1, \quad \text{on } \Gamma, \tag{55}$$

$$-\nu \boldsymbol{t} \frac{\partial \boldsymbol{u}_{\mathrm{S}}}{\partial \boldsymbol{n}_{\mathrm{S}}} = G\boldsymbol{u}_{\mathrm{S}} \cdot \boldsymbol{t} + g_2, \quad \text{on } \Gamma, \tag{56}$$

the variational formulation has only a small change: Eq. (51) now includes the two terms $-\langle g_1, \boldsymbol{V}_{\mathrm{S}} \cdot \boldsymbol{n}_{\mathrm{S}} \rangle_\Gamma - \langle g_2, \boldsymbol{V}_{\mathrm{S}} \cdot \boldsymbol{t} \rangle_\Gamma$ on the right side. In addition, we utilize 16 neurons in each hidden layer and apply the relative L^1 error ($\mathrm{err}\, L^1 : \frac{\|\boldsymbol{r}-\boldsymbol{R}\|_{L^1}}{\|\boldsymbol{r}\|_{L^1}}$) and relative L^2 error ($\mathrm{err}\, L^2 : \frac{\|\boldsymbol{r}-\boldsymbol{R}\|_0}{\|\boldsymbol{r}\|_0}$) to reflect the accuracy between the results of the CDNNs and the exact solution ($\boldsymbol{r}$: the exact solution; $\boldsymbol{R}$: the neural network).

4.1. Test 1

In this subsection we study the performance of the CDNNs for the benchmark problem presented in Ref. [56]. This problem is defined for $\Omega_{\mathrm{S}} = (0,1)^2$, $\Omega_{\mathrm{D}} = (0,1) \times (1,2)$ and the interface $\Gamma = \{0 < x < 1,\ y = 1\}$ as

$$\boldsymbol{u}_{\mathrm{S}} = \begin{pmatrix} x^2 \pi \sin(2\pi y)(x-1)^2 \\ -2x \sin(y\pi)^2 (2x-1)(x-1) \end{pmatrix},$$
$$p_{\mathrm{S}} = (\cos(1) - 1)\sin(1) + \cos(y)\sin(x),$$

$$\boldsymbol{u}_{\mathrm{D}} = \begin{pmatrix} \sin(\pi x)\sin(\pi y) \\ -2x \sin(y\pi)^2 (2x-1) \end{pmatrix}, \quad p_{\mathrm{D}} = \sin(\pi x)\cos(\pi y).$$

Similar to Ref. [56], we fix $\boldsymbol{K}$ to be the identity tensor in $\mathbb{R}^{2\times 2}$, $\mu = \rho = \beta = \nu = 1$. Due to the fact that the interface conditions (8) and (9) are violated, we exploit the interface conditions (55) and (56), where g_1 and g_2 can be computed by the exact solution. Specifically, the errors converge as the

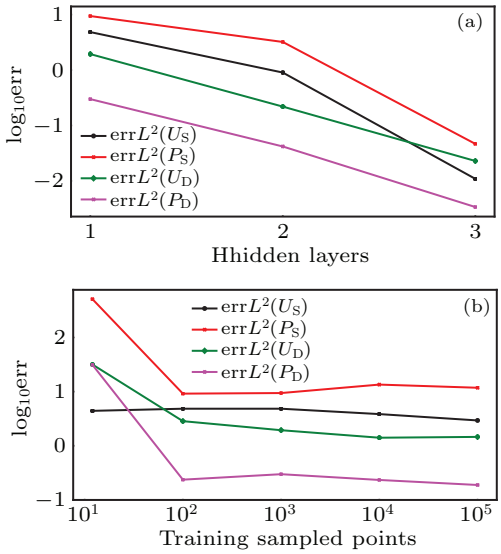


Fig. 3. The influence of different hidden layers and different training data on err L^2 (Test 1): (a) 400 training points, (b) one hidden layer.

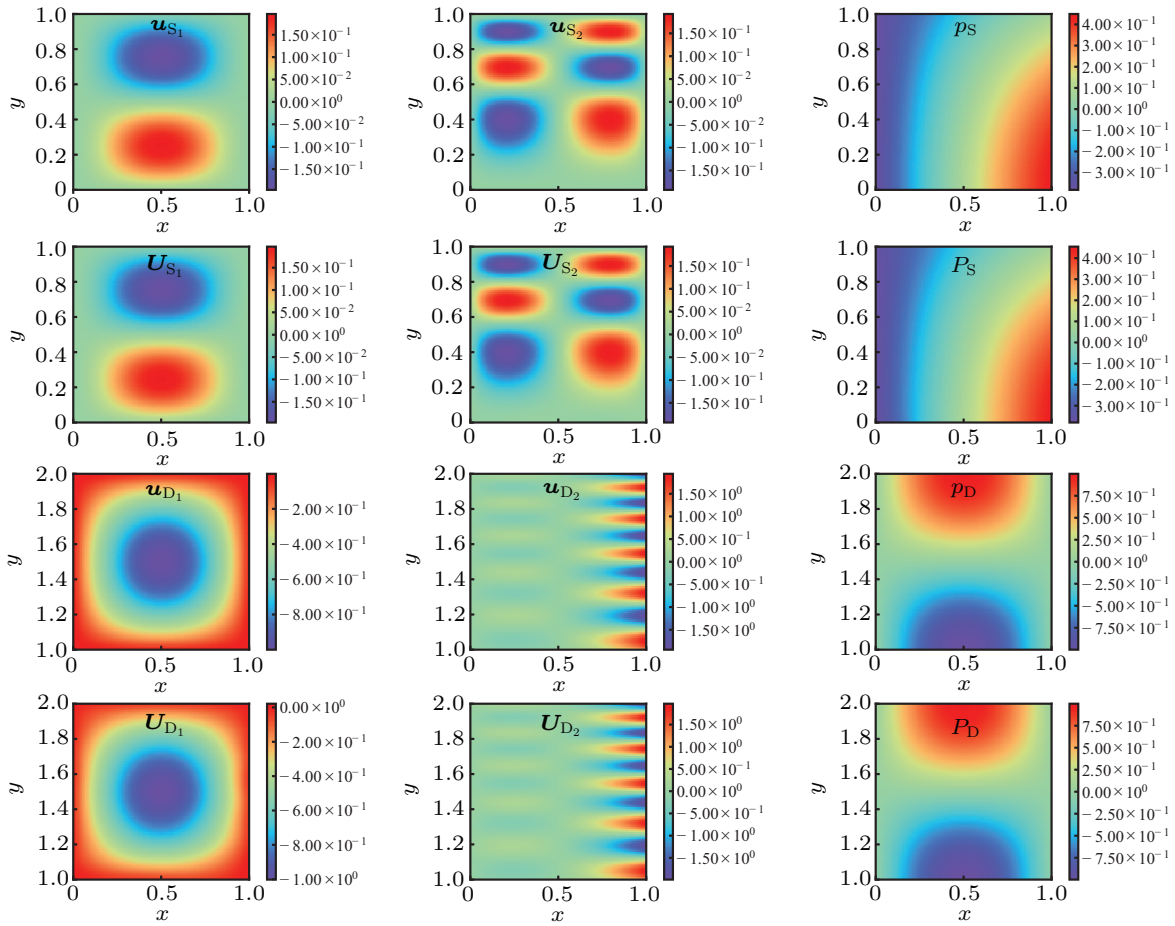


Fig. 4. The contrast of the exact solution and the CDNNs (Test 1).

hidden layer increases in Fig. 3(a). Figure 3(b) reveals that the change of data has no significant influence on errors once the size is larger than 10^2. In particular, Fig. 4 displays the exact solution of the couples Stokes Darcy–Forchheimer problems and the results of the CDNNs, one can observe that both are approximately identical. The point-wise errors are depicted in Fig. 5. As can be seen, the absolute errors are almost all 0, which also indicates the closeness between the exact solution and the approximation solution. Table 1 exhibits the details of the results, which are consistent with our theory.

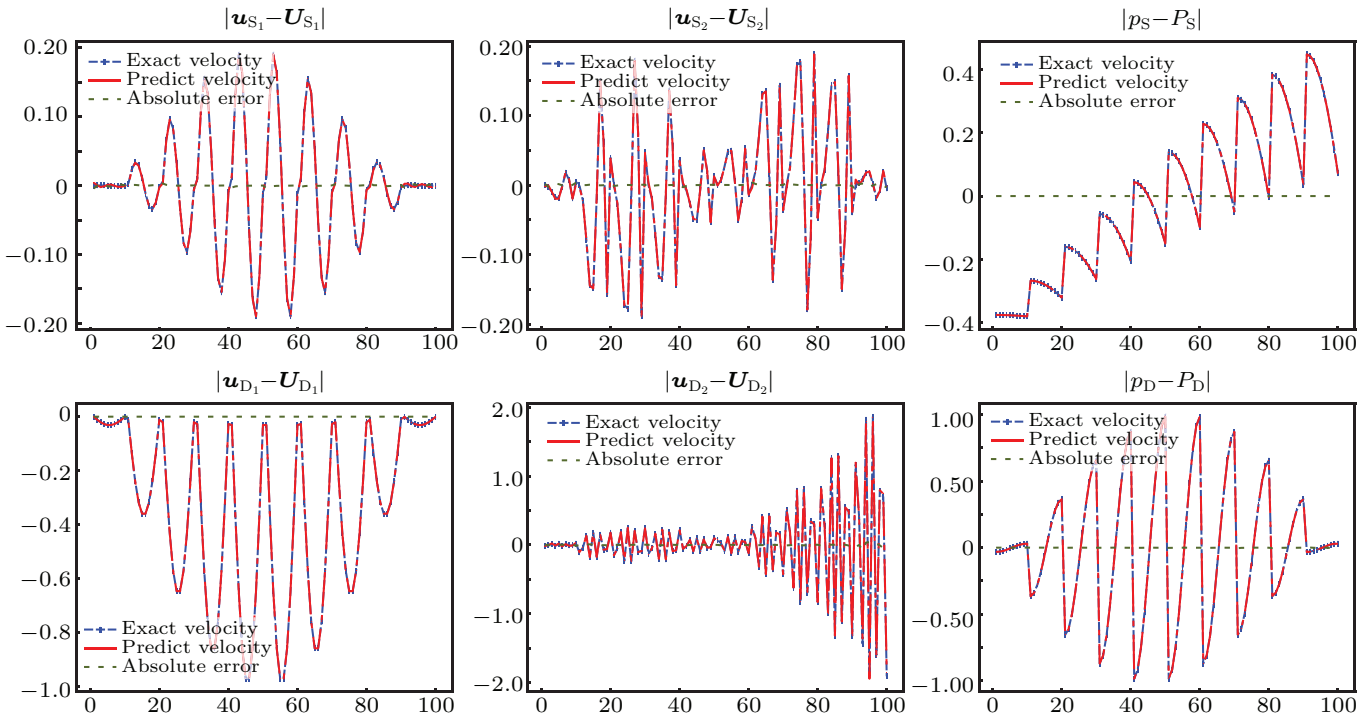


Fig. 5. The point-wise errors (Test 1).

Table 1. The relative errors of Test 1.

400 sampled points				
1 layer	U_S	P_S	U_D	P_D
err L^1	2.49×10^{0}	9.15×10^{0}	8.72×10^{0}	2.74×10^{-1}
err L^2	4.84×10^{0}	9.42×10^{0}	1.94×10^{0}	2.99×10^{-1}
2 layers	U_S	P_S	U_D	P_D
err L^1	4.85×10^{-1}	3.59×10^{0}	9.62×10^{-2}	4.03×10^{-2}
err L^2	9.01×10^{-1}	3.21×10^{0}	2.19×10^{-1}	4.18×10^{-2}
3 layers	U_S	P_S	U_D	P_D
err L^1	5.80×10^{-3}	5.26×10^{-2}	1.01×10^{-2}	3.25×10^{-3}
err L^2	1.09×10^{-2}	4.66×10^{-2}	2.29×10^{-2}	3.38×10^{-3}

4.2. Test 2

In this example, we consider $\Omega_S=(0,1)^2$, $\Omega_D=(0,1)\times(1,2)$ and the interface $\Gamma=\{0<x<1,\ y=1\}$ with an analytical solution presented in Ref. [56]. We set $\boldsymbol{K}$ to be the identity tensor in $\mathbb{R}^{2\times2}$, $\mu=\rho=\beta=\nu=1$, and the exact solution is given by

$$\boldsymbol{u}_S=\begin{pmatrix}-\cos^2(\frac{\pi y}{2})\sin(\frac{\pi x}{2})\\ \frac{1}{4}\cos(\frac{\pi x}{2})(\sin(\pi y)+\pi y)\end{pmatrix},$$

$$p_S=\frac{\pi}{4}\cos\left(\frac{\pi x}{2}\right)\left(y-2\cos\left(\frac{\pi y}{2}\right)^2\right),$$

and

$$\boldsymbol{u}_D=\begin{pmatrix}-\frac{1}{8}\sin(\frac{\pi x}{2})\\ \frac{1}{4}\pi\cos(\frac{\pi x}{2})\end{pmatrix},\quad p_D=-\frac{\pi}{4}\cos\left(\frac{\pi x}{2}\right)y.$$

Naturally, the corresponding $\boldsymbol{f}_S$, f_D, and $\boldsymbol{g}_D$ can be calculated by the exact solution. Note that this example satisfies the interface conditions (7)–(9). According to Test 1, we choose appropriate data and hidden layer to solve the second example. Figure 6 and Table 2 show the accuracy of the CDNNs for solving the coupled problems in detail.

Table 2. The relative errors of Test 2.

400 sampled points				
1 layer	U_S	P_S	U_D	P_D
err L^1	1.82×10^{-2}	1.19×10^{-1}	1.57×10^{-2}	1.08×10^{-2}
err L^2	3.66×10^{-2}	1.23×10^{-1}	4.62×10^{-2}	1.11×10^{-2}
2 layers	U_S	P_S	U_D	P_D
err L^1	2.27×10^{-4}	1.74×10^{-3}	1.04×10^{-4}	4.67×10^{-5}
err L^2	4.21×10^{-4}	2.00×10^{-3}	3.18×10^{-4}	5.55×10^{-5}
3 layers	U_S	P_S	U_D	P_D
err L^1	1.65×10^{-4}	1.01×10^{-3}	1.13×10^{-4}	7.69×10^{-5}
err L^2	3.37×10^{-4}	1.50×10^{-3}	3.44×10^{-4}	8.32×10^{-5}

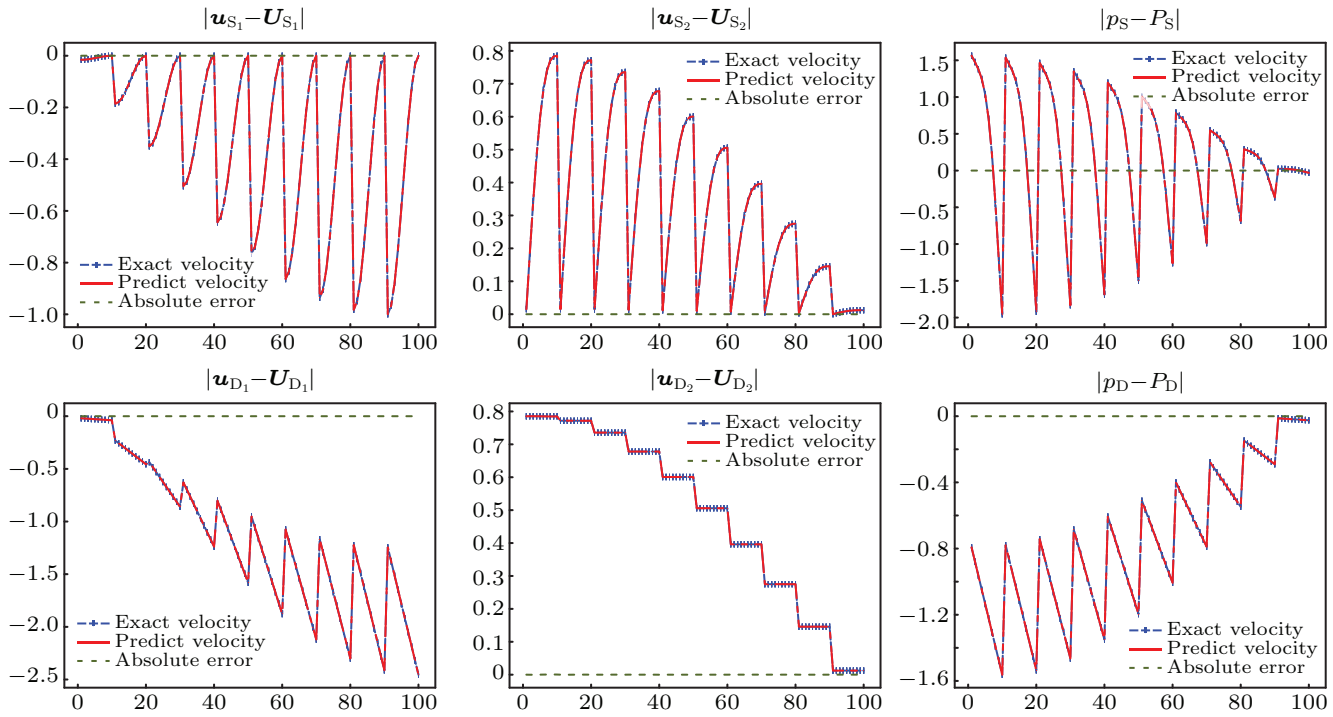


Fig. 6. The point-wise errors (Test 2).

4.3. Test 3

In this subsection, we solve coupling of the Stokes and Darcy–Forchheimer problems with highly oscillatory permeability over domains $\Omega_S = (0,1)\times(0,1/2)$, $\Omega_D = (0,1)\times(1/2,1)$ and the interface $\Gamma = \{0<x<1,\ y=1/2\}$ presented in Ref. [56]. Here we set $\mu=\rho=\beta=\nu=1$, $\boldsymbol{K}^{-1}=\rho\boldsymbol{I}$, and ρ is defined by

$$\rho = \frac{2+1.8\sin(2\pi x/\varepsilon)}{2+1.8\sin(2\pi y/\varepsilon)} + \frac{2+1.8\sin(2\pi y/\varepsilon)}{2+1.8\sin(2\pi x/\varepsilon)},$$

where $\varepsilon = 1/16$. The profile of ρ is shown in the figure. The exact solution is given by

$$\boldsymbol{u}_S = \begin{pmatrix} 16y\cos(\pi x)^2(y^2-0.25) \\ 8\pi\cos(\pi x)\sin(\pi x)(y^2-0.25)^2 \end{pmatrix}, \quad p_S = x^2,$$

$$\boldsymbol{u}_D = \begin{pmatrix} \sin(2\pi x)\cos(2\pi y) \\ -\cos(2\pi x)\sin(2\pi y) \end{pmatrix}, \quad p_D = \cos(2\pi x)\cos(2\pi y).$$

We calculate the relative errors in Table 3 to reflect ability of the CDNNs for solving the coupled problems with highly oscillatory permeability (Fig. 7). Figure 8 reveals that the CDNNs handle the highly oscillatory permeability coupled problems without losing accuracy.

4.4. Test 4

The problems that we have studied so far have the exact solution. In this example, we consider coupling of the Stokes and Darcy–Forchheimer problems with no exact solution over $\Omega_S = (-1/2,3/2)\times(0,2)$, $\Omega_D = (-1/2,3/2)\times(-2,0)$ and the interface $\Gamma = \{-1/2<x<3/2,\ y=0\}$. Specifically, in the Stokes region, the Dirichlet boundary condition is given by Kovasznay flow,[57]

$$\boldsymbol{u}_S = \begin{pmatrix} 1-\mathrm{e}^{\lambda x}\cos(2\pi y) \\ \frac{\lambda}{2\pi}\mathrm{e}^{\lambda x}\sin(2\pi y) \end{pmatrix},$$

where $\lambda = \frac{-8\pi^2}{1+\sqrt{1+64\pi^2}}$. Moreover, we set $\mu=\rho=\beta=\nu=1$ and $\boldsymbol{g}_D = \boldsymbol{0}$, $f_D = 0$, $\boldsymbol{f}_S = \boldsymbol{0}$. In addition, p_D satisfies the homogeneous Dirichlet boundary condition along $y=-2$, otherwise it has an homogeneous Neumann boundary condition. The permeability is taken to be $\boldsymbol{K}=\varepsilon\boldsymbol{I}$ and the $\varepsilon = 10^4$. Since the exact solution for this example is unavailable, we provide L^2 error of interface to demonstrate the accuracy of the CDNNs in Table 4. Obviously, the error decreases gradually with increase of the hidden layers. Furthermore, Figs. 9 and 10 display the exact solution and the results of CDNNs in detail.

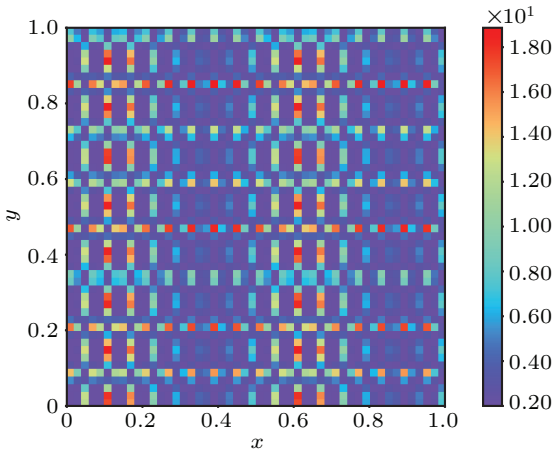


Fig. 7. The value of ρ (Test 3).

Table 3. The relative errors of Test 3.

400 sampled points				
1 layer	$\boldsymbol{U}_{\mathrm{S}}$	P_{S}	$\boldsymbol{U}_{\mathrm{D}}$	P_{D}
err L^1	3.59×10^{-1}	5.03×10^{0}	5.52×10^{-2}	7.59×10^{-2}
err L^2	6.79×10^{-1}	5.20×10^{0}	1.73×10^{-1}	7.07×10^{-2}
2 layers	$\boldsymbol{U}_{\mathrm{S}}$	P_{S}	$\boldsymbol{U}_{\mathrm{D}}$	P_{D}
err L^1	8.42×10^{-4}	9.41×10^{-3}	1.15×10^{-3}	1.32×10^{-3}
err L^2	1.65×10^{-3}	1.05×10^{-2}	3.43×10^{-3}	1.56×10^{-3}
3 layers	$\boldsymbol{U}_{\mathrm{S}}$	P_{S}	$\boldsymbol{U}_{\mathrm{D}}$	P_{D}
err L^1	1.89×10^{-4}	3.04×10^{-3}	2.97×10^{-4}	6.70×10^{-5}
err L^2	3.65×10^{-4}	3.37×10^{-3}	8.98×10^{-4}	8.40×10^{-5}

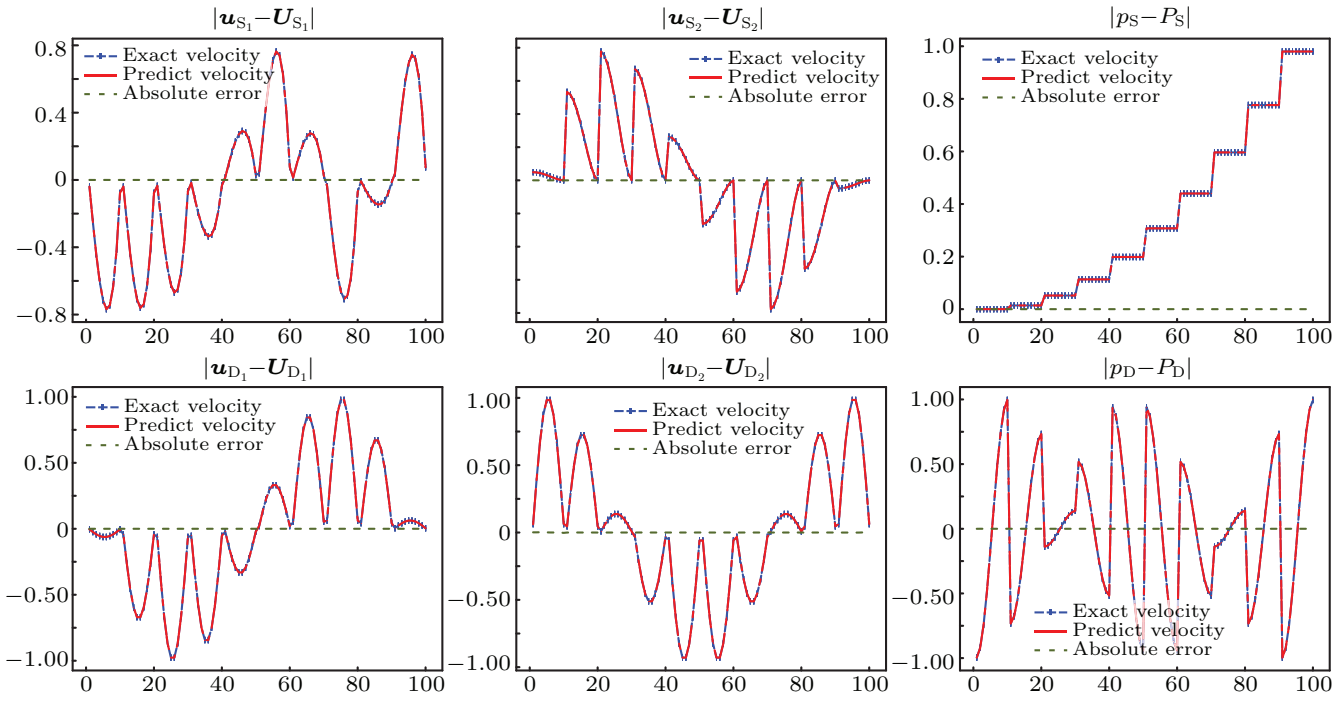


Fig. 8. The point-wise errors (Test 3).

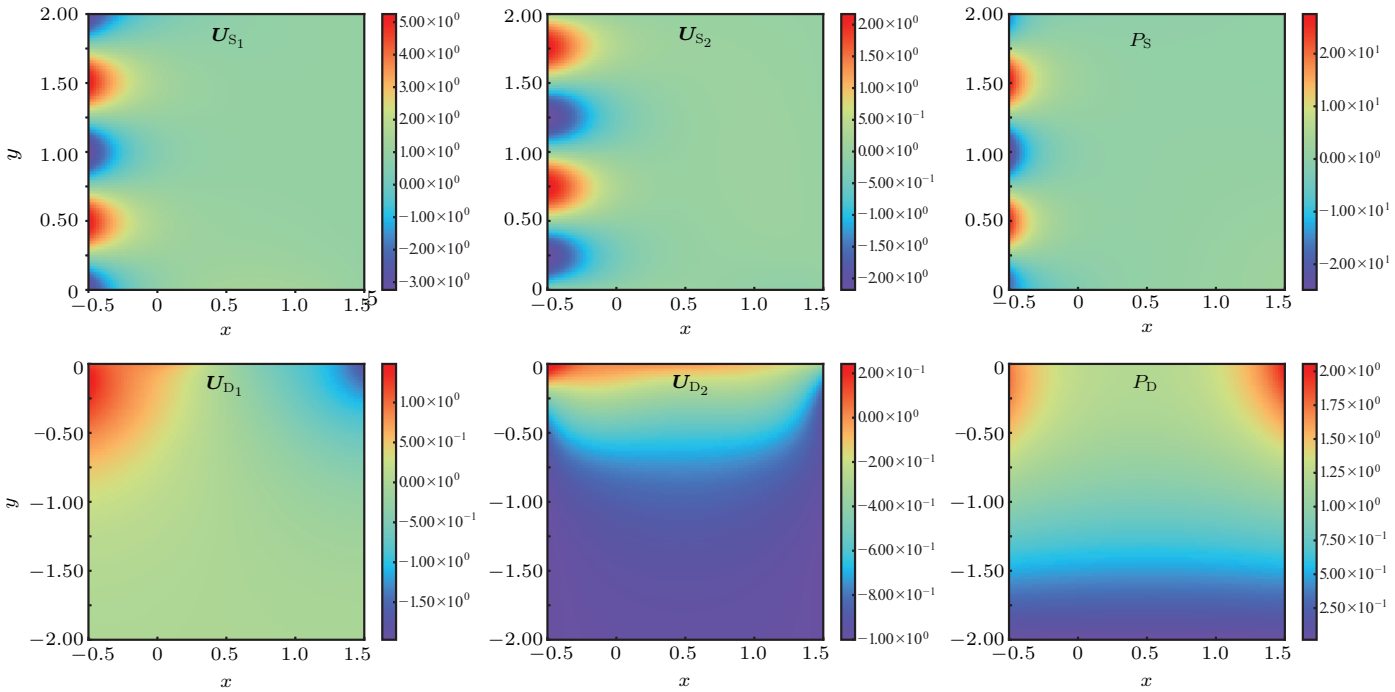


Fig. 9. The results of CDNNs (Test 4).

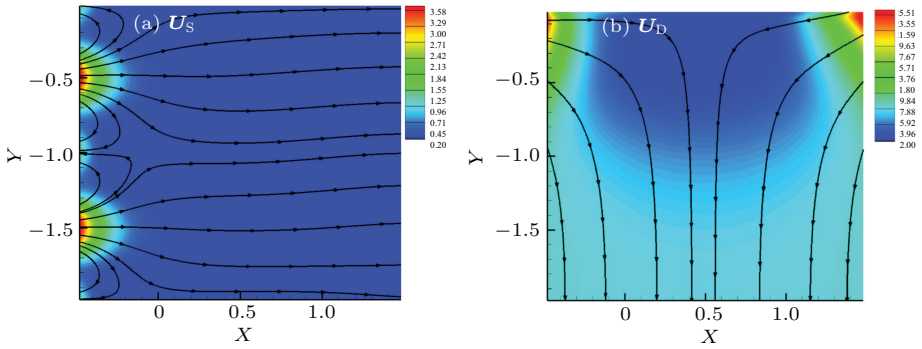


Fig. 10. The velocity flows of Stokes and Darcy (Test 4).

Table 4. The errors in interface of Test 4 ($K = 10000$).

	400 sampled points		
	Condition 1	Condition 2	Condition 3
1 layer	6.49×10^{-2}	9.14×10^{-2}	3.03×10^{-2}
2 layers	3.67×10^{-5}	4.74×10^{-2}	6.27×10^{-4}
3 layers	3.37×10^{-5}	7.53×10^{-3}	1.44×10^{-5}

4.5. Test 5

We conclude this section with a physical flow, where $\Omega_{\mathrm{S}} = (0,1) \times (1,2)$, $\Omega_{\mathrm{D}} = (0,1)^2$ and the interface $\Gamma = \{0 < x < 1,\ y = 1\}$. In Ω_{S}, the boundaries of the cavity are walls with no-slip condition, except for the upper boundary where a uniform tangential velocity $\boldsymbol{u}_{\mathrm{S}}(x,2) = (1,0)^{\mathrm{T}}$ is imposed, which is driven cavity flow. More precisely, we enforce homogeneous Neumann and Dirichlet boundary conditions, respectively, on $\Gamma_{D,N} = \{x = 0 \text{ or } y = 0\}$ and $\Gamma_{D,D} = \{x = 1\}$. In addition, we set $\boldsymbol{K}$ to be the identity tensor in $\mathbb{R}^{2\times 2}$, $\mu = \rho = \beta = \nu = 1$, and $f_{\mathrm{D}} = 0$, $\boldsymbol{f}_{\mathrm{S}} = \boldsymbol{0}$, $\boldsymbol{g}_{\mathrm{D}} = \boldsymbol{0}$. The results of the CDNNs are depicted in Fig. 11. More vividly, we display the velocity flows of free-flow and porous media zones in Fig. 12.

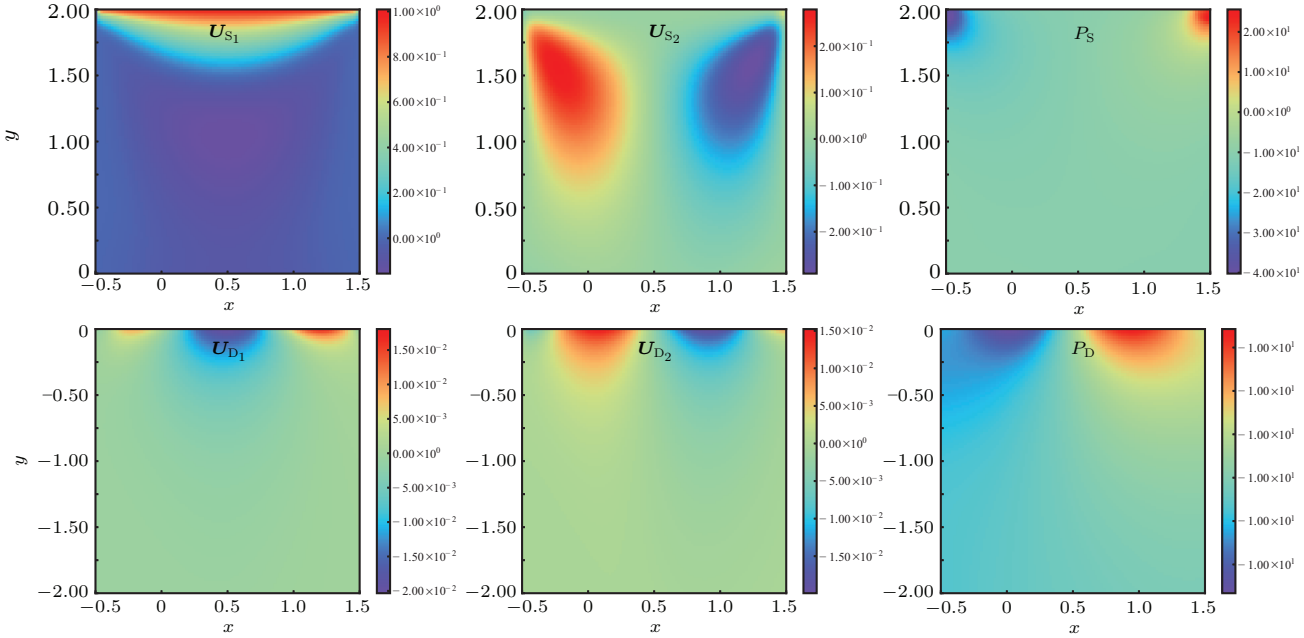


Fig. 11. The results of the CDNNs (Test 5).

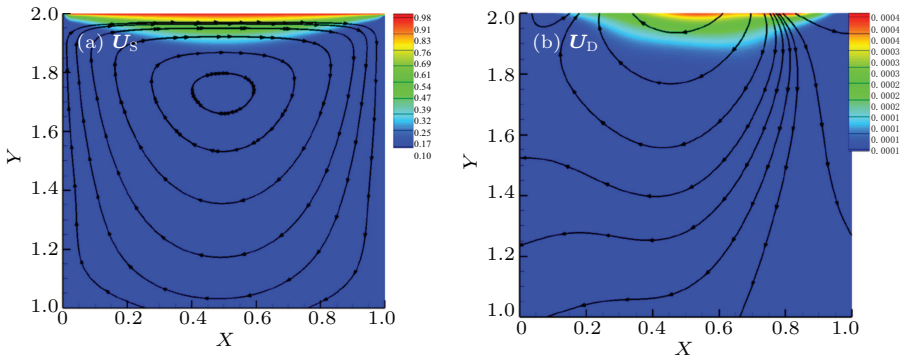


Fig. 12. The velocity flows of Stokes and Darcy (Test 5).

5. Conclusions

In summary, we have proposed the CDNNs to study the coupled Stokes and Darcy–Forchheimer problems. Our method compiles the interface conditions of the coupled problems into the networks properly and can be served as an efficient alternative to the complex coupled problems. CDNNs avoid limitations of the traditional methods, such as decoupling, grid construction and the complicated interface conditions. Furthermore, it is meshfree and parallel, it can solve multiple variables independently at the same time. Specially, we provide the convergence of the loss function and the convergence of the CDNNs to the exact solution. The numerical results are consistent with our theory sufficiently. Moreover, we leave the following issues subject to our future works: (1) combining data-driven with model-driven to solve the high dimensional coupled problems, (2) considering the specific size of the networks through theoretical analysis, (3) combining traditional numerical methods with deep learning to solve more complicated high dimensional coupled problems.

Acknowledgements

Project supported in part by the National Natural Science Foundation of China (Grant No. 11771259), the Special Support Program to Develop Innovative Talents in the Region of Shaanxi Province, the Innovation Team on Computationally Efficient Numerical Methods Based on New Energy Problems in Shaanxi Province, and the Innovative Team Project of Shaanxi Provincial Department of Education (Grant No. 21JP013).

References

[1] Li J, Bai Y and Zhao X 2023 *Modern Numerical Methods for Mathematical Physics Equations* (Beijing: Science Press) p. 10 (in Chinese)
[2] Li J, Lin X and Chen Z 2022 *Finite Volume Methods for the Incompressible Navier–Stokes Equations* (Berlin: Springer) p. 15
[3] Li J 2019 *Numerical Methods for the Incompressible Navier–Stokes Equations* (Beijing: Science Press) p. 8
[4] Saffman P G 1971 *Stud. Appl. Math.* **50** 93
[5] Forchheimer P 1901 *Zeitz. Ver. Duetch Ing.* **45** 1782 (in Japanese)
[6] Park E J 1995 *SIAM J. Numer. Anal.* **32** 865
[7] Kim M Y and Park E J 1999 *Comput. Math. Appl.* **38** 113
[8] Park E J 2005 *Numer. Methods Part. Differ. Equ.* **21** 213
[9] Discacciati M, Miglio E and Quarteroni A 2002 *Appl. Numer. Math.* **43** 57
[10] Layton W J, Schieweck F and Yotov I 2003 *SIAM J. Numer. Anal.* **40** 2195
[11] Rivière B 2005 *J. Sci. Comput.* **22** 479
[12] Rivière B and Yotov I 2005 *SIAM J. Numer. Anal.* **42** 1959
[13] Burman E and Hansbo P 2007 *J. Comput. Appl. Math.* **198** 35
[14] Gatica G N, Oyarzúa R and Sayas F J 2011 *Math. Comput.* **80** 1911
[15] Girault V, Vassilev D and Yotov I 2014 *Numer. Math.* **127** 93
[16] Lipnikov K, Vassilev D and Yotov I 2014 *Numer. Math.* **126** 321
[17] Qiu C X, He X M, Li J and Lin Y P 2020 *J. Comput. Phys.* **411** 109400
[18] Li R, Gao Y L, Li J and Chen Z X 2018 *J. Comput. Appl. Math.* **334** 111
[19] He Y N and Li J 2010 *Int. J. Numer. Anal. Mod.* **62** 647
[20] Liu X, Li J and Chen Z X 2018 *J. Comput. Appl. Math.* **333** 442
[21] Li J, Mei L Q and He Y N 2006 *Appl. Math. Comput.* **182** 24
[22] Zhu L P, Li J and Chen Z X 2011 *J. Comput. Appl. Math.* **235** 2821
[23] Krizhevsky A, Sutskever I and Hinton G E 2012 *Commun. ACM* **64** 84
[24] Hinton G, Deng L, Yu D, *et al.* 2012 *IEEE Signal Proc. Mag.* **29** 82
[25] He K M, Zhang X Y, Ren S Q, *et al.* 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* June 27–30, 2016, Las Vegas, NV, USA p. 770
[26] Cotter N E 1990 *IEEE Trans. Neural Networks* **4** 290
[27] Hornik K, Stinchcombe M and White H 1989 *Neural Networks* **2** 359
[28] Hornik K, Stinchcombe M and White H 1990 *Neural Networks* **3** 551
[29] Hornik K 1991 *Neural Networks* **4** 251
[30] Cybenko G 1989 *Math. Control Signal.* **2** 303
[31] Telgrasky M 2016 *Proc. Mach. Learn. Res.* **49** 1517
[32] Mhaskar H, Liao Q L and Poggio T 2016 arXiv:1603.00988v4[cs.LG]
[33] Khoo Y, Lu J F and Ying L X 2017 arXiv:1707.03351[math.NA]
[34] Li J, Yue J, Zhang W, *et al.* 2022 *J. Sci. Comput.* (accepted)
[35] Li J, Zhang W and Yue J 2021 *Int. J. Numer. Anal. Model.* **18** 427
[36] Yue J and Li J 2022 *Int. J. Numer. Methods Fluids.* **94** 1416
[37] Yue J and Li J 2023 *Appl. Math. Comput.* **437** 127514
[38] Fan Y W, Lin L, Ying L X, *et al.* 2018 arXiv:1807.01883[math.NA]
[39] Wang M, Cheung S W, Chung E T, *et al.* 2018 arXiv:1810.12245[math.NA]
[40] Li X 1996 *Neurocomputing* **12** 327
[41] Lagaris I E, Likas A C and Fotiadis D I 1998 *IEEE Trans. Neural Network* **9** 987
[42] Lagaris I E, Likas A C and Papageorgiou D G 2000 *IEEE Trans. Neural Network* **11** 1041
[43] McFall K S and Mahan J R 2009 *IEEE Trans. Neural Network* **20** 1221
[44] Raissi M, Perdikaris P and Karniadakis G E 2017 arXiv:1711.10561[cs.AI]
[45] Raissi M, Perdikaris P and Karniadakis G E 2017 arXiv:1711.10566[cs.AI]
[46] Raissi M, Perdikaris P and Karniadakis G E 2019 *J. Comput. Phys.* **378** 686
[47] Yang L, Meng X H and Karniadakis G E 2021 *J. Comput. Phys.* **425** 109913
[48] Rao C P, Sun H and Liu Y 2020 arXiv: 2006.08472v1[math.NA]
[49] Olivier P and Fablet R 2020 arXiv:2002.01029 [physics.comp-ph]
[50] Lu L, Meng X H, Mao Z P, *et al.* 2021 *SIAM Rev.* **63** 208
[51] Fang Z W and Zhan J 2020 *IEEE Access* **8** 26328
[52] Pang G F, Lu L and Karniadakis G E 2019 *SIAM J. Sci. Comput.* **41** A2603
[53] Zhu Y H, Zabaras N, Koutsourelakis P S, *et al.* 2019 *J. Comput. Phys.* **394** 56
[54] Sirignano J and Spiliopoulos K 2018 *J. Comput. Phys.* **375** 1339
[55] Beaver G S and Joseph D D 1967 *J. Fluid Mech.* **30** 197
[56] Zhao L, Chung E T, Park E J and Zhou G 2021 *SIAM J. Numer. Anal.* **59** 1
[57] Kovasznay L I G 1948 *Math. Proc. Cambridge* **44** 58
[58] Léon Bottou 2012 *Lecture Notes in Computer Science* Grégoire M, Genevieve B O and Klaus R M (eds) (Berlin: Springer) pp. 430–445