



Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam



Current density impedance imaging with PINNs

Chenguang Duan, Junjun Huang, Yuling Jiao, Xiliang Lu*, Jerry Zhijian Yang

School of mathematics and statistics, Wuhan University, China



ARTICLE INFO

Keywords:

Inverse problems
EIT
CDII
PINNs
Neural networks

ARTICLE INFO

In this paper, we introduce CDII-PINNs, a computationally efficient method for solving CDII using PINNs in the framework of Tikhonov regularization. This method constructs a physics-informed loss function by merging the regularized least-squares output functional with an underlying differential equation, which describes the relationship between the conductivity and voltage. A pair of neural networks representing the conductivity and voltage, respectively, are coupled by this loss function. Then, minimizing the loss function provides a reconstruction. A rigorous theoretical guarantee is provided. We give an error analysis for CDII-PINNs and establish a convergence rate, based on prior selected neural network parameters in terms of the number of samples. The numerical simulations demonstrate that CDII-PINNs are efficient, accurate and robust to noise levels ranging from 1% to 20%.

1. Introduction

Electrical Impedance Tomography (EIT) is a medical imaging technique to recover the conductivity value of soft tissue from pairs of flux/voltage on its boundary. However, the problem of reconstructing the electrical conductivity distribution is severely ill-posed, which is very sensitive to noise and measurement errors. To avoid the drawback of EIT, people consider other conductivity imaging techniques that combine electrical impedance tomography (EIT) and magnetic resonance imaging (MRI). In this work, we consider the current density impedance imaging (CDII), one of the medical imaging techniques, which aims to image the current density distribution within a conductive medium [1]. CDII has potential applications in a variety of fields, including non-destructive testing, material science, and biomedical imaging. For example, it can be used in non-destructive testing of concrete to detect cracks and voids, in material science to study the microstructure of metals, and in biomedical imaging to detect tumors and other abnormalities in tissues.

Theoretical results on the uniqueness of the recovery and conditional stability of CDII have been studied in [1–4], respectively. Numerical reconstruction algorithms have been extensively studied, including the level set method [5,6], the weighted least-gradient method [1], and methods based on linearized reconstruction [7]. These algorithms have been shown to be effective in reconstructing the current density distribution from measured data. The finite element methods (FEM), regarded as traditional discretization techniques, are utilized to discretize the provided regularized functional. However, FEM faces significant challenges when confronted with highly noisy problems. In this work we propose a neural network-based method by combining Physics-Informed Neural Networks (PINNs) with the Tikhonov regularization. Compared to Galerkin approximation, our approach requires only a stochastic gradient descent solver and have been numerically shown to be very robust against measurement noise.

In recent years, Physics-Informed Neural Networks (PINNs) [8] have gained a lot of attention as a residual-based deep learning method for solving both forward and inverse problems in science and engineering. In [8], the authors explored numerically the recovery of constant functions in PDEs using PINNs. Building on this work, we propose to use PINNs to identify the parameters in

* Corresponding author.

E-mail address: xllv.math@whu.edu.cn (X. Lu).

CDII by constructing a residual-based loss function that involves two neural networks. These networks are used to simultaneously approximate the unknown conductivity coefficient function and the underlying solution. For solving inverse problems in PDEs, other deep learning methods have also been proposed, such as those in [9–12]. Here, we highlight the main differences between these methods and ours. In [9], a min–max loss function is induced by the weak form of the elliptic equations. This method requires an additional network to approximate the dual variable, which can make the training process more challenging than ours. On the other hand, in [10], the authors propose to handle the CDII problem by solving a forward problem using deep learning under the framework of weighted least-gradient [1]. In the paper [11], the authors investigate the inverse source problem within the framework of PINNs. However, their analysis differs significantly from ours, which focuses on consistency from the perspective of statistical learning theory. A related work, [12], considers the error analysis of PINNs for certain inverse problems. Nevertheless, this study does not provide guidance on how to choose the neural network parameters to ensure consistency.

Our main contributions are as follows.

- We introduce CDII-PINNs, a computationally efficient method for solving CDII using PINNs in the framework of nonparametric statistical learning [13]. Our numerical simulations demonstrate that CDII-PINNs achieve high accuracy and robustness to noise levels ranging from 1% to 20%.
- We present a comprehensive error analysis for CDII-PINNs, along with a prior rule for selecting appropriate neural network parameters, such as width and depth, based on the number of samples. This approach ensures that the estimation error is consistent and can be controlled to achieve the desired level of precision. To the best of our knowledge, this is the first theoretical result of its kind for PINNs in the context of inverse problems.

The remainder of the paper is organized as follows. Our CDII-PINNs and their convergence analysis are presented in Section 2 and Section 3, respectively. In Section 4, we provide numerical simulations. The conclusions and main proofs are presented in Section 5 and Appendix, respectively.

2. Methodology and preliminary

2.1. Problem formulation

Let $U \subseteq \mathbb{R}^d$ ($d > 1$) denote a bounded non-empty C^∞ -domain, referring to the space occupied by the object. Without loss of generality, we assume $U \subseteq (0, 1)^d$. The electric potential or voltage u in the interior of the domain is governed by the following second-order elliptic equation

$$\nabla \cdot (\gamma \nabla u) = 0 \quad \text{in } U, \tag{2.1}$$

corresponding to the unknown conductivity coefficient $\gamma \in L^\infty(U)$. In the idealistic situation, one assumes that the potential is measured everywhere on the boundary, that is, the Dirichlet trace

$$u = f \quad \text{on } \partial U. \tag{2.2}$$

The govern equation Eq. (2.1) together with the observations on the boundary Eq. (2.2) is a well-used mathematical model for EIT, called current density impedance imaging (CDII) [14]. Denote $J = -\gamma \nabla u$ the internal current density vector. In CDII, the magnitude of the current density field $a(x) = |J(x)|$ is observed on the whole of U . The goal is to determinate the conductivity distribution from the noisy measurement data. Moreover, we assume the conductivity γ belongs to the admissible set \mathcal{K} defined as

$$\mathcal{K} = \{\gamma \in L^\infty(U) : \gamma_0 \leq \gamma(x) \leq \gamma_1 \text{ a.e. in } U\}, \tag{2.3}$$

with $0 < \gamma_0 < \gamma_1 < \infty$.

Given a conductivity function $\gamma \in \mathcal{K}$, denote by u_γ the solution of Eqs. (2.1) and (2.2). Let us define a nonlinear operator $\mathcal{G} : \gamma \mapsto \gamma |\nabla u_\gamma|$, and consider an additive δ^2 -sub-Gaussian noise model

$$Y = \mathcal{G}[\gamma^\dagger](X) + \xi, \quad \xi \sim \text{i.i.d. subG}(0, \delta^2), \tag{2.4}$$

where the measurement noise ξ is independent of X and γ^\dagger . It is straightforward that $\mathbb{E}_{(X,Y)}[(Y - \mathcal{G}[\gamma^\dagger](X))^2] = \delta^2$. In this paper, we make no distinction between notations Y and a^δ .

2.2. Neural networks

A neural network $\phi : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_{D+1}}$ is a function defined by

$$\phi(x) = T_D(\varrho(T_{D-1}(\cdots \varrho(T_0(x)) \cdots))),$$

where the activation function ϱ is applied component-wisely and $T_\ell(x) := A_\ell x + b_\ell$ is an affine transformation with $A_\ell \in \mathbb{R}^{N_{\ell+1} \times N_\ell}$ and $b_\ell \in \mathbb{R}^{N_\ell}$ for $\ell = 0, \dots, D$. In this paper, we consider the case $N_0 = d$ and $N_{D+1} = 1$. The number D the depth of the neural network. Denote $S_i = \sum_{\ell=1}^i (\|A_\ell\|_0 + \|b_\ell\|_0)$ for $i = 1, \dots, D$, where $\|\cdot\|_0$ is the number of nonzero entries for a given vector. Hence S_i represents the nonzero weights on the first i -layers and then $S = S_D$ is the total number of nonzero weights.

We define the neural network class $\mathcal{N}_\varrho(D, S, \mathcal{B})$ as the collection of ϱ -neural networks with at most D layers and at most S nonzero weights and each weight are bounded by \mathcal{B} .

2.3. CDII-PINNs

By Tikhonov regularization, we can obtain a reconstruction by solving the optimization problem

$$\min_{\gamma \in \mathcal{K}} \left\{ J(\gamma) := |U| \mathbb{E}_{(X,Y)} [(Y - G[\gamma](X))^2] + \alpha \psi(\gamma) \right\}, \quad (2.5)$$

where ψ is the regularization functional. Observe that Eq. (2.5) is equivalent to a optimal control problem

$$\min_{\gamma, u} J(\gamma, u) = |U| \mathbb{E}_{(X,Y)} [(Y - G[\gamma](X))^2] + \alpha \psi(\gamma), \quad \text{subject to Eqs. (2.1)–(2.3)},$$

which can be converted to an unconstrained minimization problem by the penalty method

$$\min_{\gamma, u} L(\gamma, u) := J(\gamma, u) + G(\gamma, u). \quad (2.6)$$

For the sake of simplicity, we choose the penalty term as

$$G(\gamma, u) = \|\nabla \cdot (\gamma \nabla u)\|_{L^2(U)}^2 + \|Tu - f\|_{L^2(\partial U)}^2,$$

which is referred as the physic-informed loss function [8] for solving the constraint PDEs Eqs. (2.1) and (2.2) given γ . Furthermore, since

$$\mathbb{E}_{(X,Y)} [(Y - G[\gamma](X))^2] = \|G[\gamma^\dagger] - G[\gamma]\|_{L^2(U)}^2 + \delta^2,$$

it follows that $J(\gamma^\dagger, u^\dagger) \leq J(\gamma, u)$ for each (γ, u) if $\alpha = 0$. In addition, $G(\gamma, u) \geq 0$ for any (γ, u) and $G(\gamma^\dagger, u^\dagger) = 0$, which means that for $\alpha = 0$

$$(\gamma^\dagger, u^\dagger) \in \operatorname{argmin}_{\gamma, u} L(\gamma, u).$$

However, in inverse problems, the distribution of (X, Y) is typically unknown and only a random sample $S = \{(X_i, Y_i)\}_{i=1}^n \cup \{\bar{X}_i\}_{i=1}^n$ is available, where $\{(X_i, Y_i)\}_{i=1}^n$ are n independent copies of (X, Y) and $\{\bar{X}_i\}_{i=1}^n$ are n independent random variables drawn from the uniform distribution on ∂U . Based on the measurement data, we employ the Monte Carlo method to discretize the population risk Eq. (2.6) and yield the empirical risk

$$L_n(\gamma, u) := J_n(\gamma, u) + G_n(\gamma, u), \quad (2.7)$$

where the empirical objective functional and penalty term is given by

$$\begin{aligned} J_n(\gamma, u) &= \frac{1}{n} \sum_{i=1}^n (Y_i - \gamma |\nabla u|(X_i))^2 + \alpha \psi_n(\gamma), \quad \text{and} \\ G_n(\gamma, u) &= \frac{1}{n} \sum_{i=1}^n (\nabla \cdot (\gamma \nabla u)(X_i))^2 + (Tu(\bar{X}_i) - f(\bar{X}_i))^2, \end{aligned}$$

respectively. Here $\psi_n(\gamma)$ is a discrete version of $\psi(\gamma)$, satisfying $\mathbb{E}_S \psi_n(\gamma) = \psi(\gamma)$ for any fixed function $\gamma : U \rightarrow \mathbb{R}$.

Finally, we represent the conductivity γ and voltage u by two dependent neural networks, respectively and then substitute them into the empirical risk Eq. (2.7). An estimator of $(\gamma^\dagger, u^\dagger)$ can be obtained by empirical risk minimization, which reads

$$(\hat{\gamma}_n, \hat{u}_n) \in \operatorname{argmin}_{(\gamma, u) \in \mathcal{F}_\gamma \times \mathcal{F}_u} L_n(\gamma, u) := J_n(\gamma, u) + G_n(\gamma, u), \quad (2.8)$$

where \mathcal{F}_γ and \mathcal{F}_u are two neural network classes chosen by the users.

To summarize, our approach constructs a loss function by merging the measurement data with an underlying physical model. A pair of neural networks representing conductivity and voltage, respectively, are coupled by this physics-informed loss function. Therefore, we call our method as Physics-Informed Neural Networks for Current Density Impedance Imaging (CDII-PINNs). A detailed reconstruction procedure is shown in Algorithm 1.

In the classical methods for solving CDII, such as [1], to ensure that the reconstructions of conductivity and voltage are admissible, one must solve the forward problem after each update of the reconstructions, which is costly. In contrast, CDII-PINNs need not fit the measurement data and solve the forward problems alternatively, as the loss of which considers the measurement data and the PDE constraints simultaneously.

3. Error analysis

In this section, we present an error estimate to the cost functional (2.6). For simplicity we consider the case with regularization parameter $\alpha = 0$. Then the estimator $(\tilde{\gamma}_n, \tilde{u}_n) \in \mathcal{F}_\gamma \times \mathcal{F}_u$ can be evaluated via its expected excess risk

$$R(\tilde{\gamma}_n, \tilde{u}_n) = \mathbb{E}_S \left[\|\tilde{\gamma}_n |\nabla \tilde{u}_n| - \gamma^\dagger |\nabla u^\dagger|\|_{L^2(U)}^2 \right] + \mathbb{E}_S \left[G(\tilde{\gamma}_n, \tilde{u}_n) \right],$$

where the first term measures the difference between the magnitude of the recovered current density field and that of the ground truth current density field, and the second term reflects how the estimator satisfies the PDE constraint.

Algorithm 1 CDII-PINNs.

Input: A measurement data set $S = \{(X_i, Y_i)\}_{i=1}^n \cup \{\bar{X}_i\}_{i=1}^n$.

- 1: Construct neural networks (γ_ϕ, u_θ) , parameterized by (ϕ, θ) .
- 2: Initialize parameters (ϕ, θ) randomly.
- 3: **for** $k = 1 : K$ **do**
- 4: # Compute the loss (or empirical risk) based on the data set.
- 5: $L_n(\gamma_\phi, u_\theta) = J_n(\gamma_\phi, u_\theta) + G_n(\gamma_\phi, u_\theta)$ as defined in (2.7).
- 6: # Back propagation: compute the gradient of the loss w.r.t. (ϕ, θ) .
- 7: $(g_\phi, g_\theta) = \nabla_{(\phi, \theta)} L_n(\gamma_\phi, u_\theta)$.
- 8: # Update (ϕ, θ) by some SGD-type algorithm.
- 9: $(\phi, \theta) \leftarrow \text{SGD}\{(\phi, \theta), (g_\phi, g_\theta), \tau\}$. *
- 10: **end for**

Output: $\tilde{\gamma}_n = \gamma_\phi, \tilde{u}_n = u_\theta$.

* Denote by `SGD(parameters,gradient,learning rate)` a step of SGD update.

Assumption 3.1. Assume f is bounded, i.e., $\|f\|_{H^{3/2}(\partial U)} \leq B_f$ with $B_f \geq 1$.

Assumption 3.2 (Boundedness). Assume $\gamma^\dagger, u^\dagger$ and functions in $\mathcal{F}_\gamma, \mathcal{F}_u$ are bounded, which means

- (i) $\{\gamma^\dagger\} \cup \mathcal{F}_\gamma \subseteq \{\gamma : \|\gamma\|_{W^{1,\infty}(U)} \leq B_\gamma\}$ with $B_\gamma \geq 1$, and
- (ii) $\{u^\dagger\} \cup \mathcal{F}_u \subseteq \{u : \|u\|_{W^{2,\infty}(U)} \leq B_u\}$ with $B_u \geq 1$.

To measuring the complexity or compactness of a subset of a metric space in a quantitative way, we introduce the covering number and the metric entropy.

Definition 3.3 (Covering Number and Metric Entropy). Let (S, d) be a metric space, and $T \subseteq S$. A set $T_\varepsilon \subseteq S$ is called an ε -cover of T if and only if for each $t \in T$, there exists $t_\varepsilon \in T_\varepsilon$, such that $d(t, t_\varepsilon) \leq \varepsilon$. Moreover,

$$N(\varepsilon, T, d) := \inf \left\{ |T_\varepsilon| : T_\varepsilon \text{ is an } \varepsilon\text{-cover of } T \right\}$$

is called the ε -cover number of T , and $H(\varepsilon, T, d) = \log N(\varepsilon, T, d)$ is called the ε metric entropy of T .

For each estimator $(\tilde{\gamma}_n, \tilde{u}_n)$ taking values in $\mathcal{F}_\gamma \times \mathcal{F}_u$, we define the corresponding quantity

$$\Delta_n(\tilde{\gamma}_n, \tilde{u}_n) := \mathbb{E}_S \left[L_n(\tilde{\gamma}_n, \tilde{u}_n) - L_n(\hat{\gamma}_n, \hat{u}_n) \right],$$

which measures the difference between the expected empirical risk of $(\tilde{\gamma}_n, \tilde{u}_n)$ and the minimum over all functions in the hypothesis class $\mathcal{F}_\gamma \times \mathcal{F}_u$. It is obvious that $\Delta_n(\tilde{\gamma}_n, \tilde{u}_n) \geq 0$ and $\Delta_n(\tilde{\gamma}_n, \tilde{u}_n) = 0$ if and only if $\Delta_n(\tilde{\gamma}_n, \tilde{u}_n)$ is an empirical risk minimizer.

With the help of the preceding notations, we can decompose the expected excess risk as follows.

Lemma 3.4 (Error Decomposition). Under the noise model Eq. (2.4). Suppose Assumptions 3.1 and 3.2 are fulfilled. Then it holds for each estimator $(\tilde{\gamma}_n, \tilde{u}_n)$ taking values in $\mathcal{F}_\gamma \times \mathcal{F}_u$ that

$$\begin{aligned} R(\tilde{\gamma}_n, \tilde{u}_n) &\lesssim \inf_{\varepsilon > 0} \left\{ (B_\gamma^2 B_u^2 + B_f^2 + \delta^2)(H_\gamma + H_u)n^{-1} + (\delta + B_\gamma B_u)(B_\gamma + B_u)\varepsilon \right\} \\ &\quad + \inf_{(\gamma, u) \in \mathcal{F}_\gamma \times \mathcal{F}_u} R(\gamma, u) + \Delta_n(\tilde{\gamma}_n, \tilde{u}_n), \end{aligned}$$

provided $\varepsilon > 0$ and n large enough, where

$$H_\gamma^\varepsilon = \log N(\varepsilon, \mathcal{F}_\gamma, \|\cdot\|_{W^{1,\infty}(U)}) \quad \text{and} \quad H_u^\varepsilon = \log N(\varepsilon, \mathcal{F}_u, \|\cdot\|_{W^{2,\infty}(U)}).$$

The proof refers to Appendix A.1. By Lemma 3.4, the expected excess risk can be decomposed into three parts: approximation error, optimization error, and statistical error. The approximation error, defined as $\inf_{(\gamma, u) \in \mathcal{F}_\gamma \times \mathcal{F}_u} R(\gamma, u)$, is the minimum of the expected excess risk over all functions in the hypothesis class, measuring the expressive power of the functions in $\mathcal{F}_\gamma \times \mathcal{F}_u$. The optimization error $\Delta_n(\tilde{\gamma}_n, \tilde{u}_n)$ measures the difference between the estimator and the empirical risk minimizer. The remaining part is defined as the statistical error, which is caused by the discretization of the population risk. By choosing an appropriate ε , we can find that the statistical error converges to zero as the sample size $n \rightarrow \infty$. Furthermore, the convergence rate $\mathcal{O}(\frac{1}{\sqrt{n}})$ obtained in Lemma 3.4 is an improvement of $\mathcal{O}(\frac{1}{\sqrt{n}})$ in [10,15–17].

In this paper, we ignore the optimization error, and focus on the approximation error, the statistical error and the trade-off between them.

Table 1
Commonly-used exponential PU-admissible activation functions.

Name	Definition	PU-Decay (j, τ)	$\mathcal{B}_{\rho,0}$	$\mathcal{B}_{\rho,1}$	$\mathcal{B}_{\rho,2}$	$\mathcal{B}_{\rho,3}^{\text{a}}$
sigmoid	$\frac{1}{1+e^{-x}}$	$(j, 0)$ for any $j \in \mathbb{N}$	1	1/4	8/9	≤ 1
tanh	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	$(j, 0)$ for any $j \in \mathbb{N}$	1	1	1	2/3
softplus	$\ln(1 + e^x)$	$(j, 1)$ for any $j \in \mathbb{N}$	–	–	–	–
swish	$\frac{x}{1+e^{-x}}$	$(j, 1)$ for any $j \in \mathbb{N}$	–	–	–	–

^a $|\varphi(x)| \leq \mathcal{B}_{\rho,0}$, $|\varphi'(x)| \leq \mathcal{B}_{\rho,1}$, $|\varphi''(x)| \leq \mathcal{B}_{\rho,2}$ and $|\varphi'''(x)| \leq \mathcal{B}_{\rho,3}$ for each $x \in \bar{U}$.

3.1. Approximation error

In this part, we bound the approximation error. By the following lemma, it is sufficient to estimate the approximation error of functions by neural networks.

Lemma 3.5. Suppose [Assumption 3.2](#) is fulfilled. Then

$$\inf_{(\gamma,u) \in \mathcal{F}_\gamma \times \mathcal{F}_u} R(\gamma, u) \leq C_{\text{app}} \cdot B_\gamma^2 B_u^2 \left(\inf_{\gamma \in \mathcal{F}_\gamma} \|\gamma - \gamma^\dagger\|_{C^1(\bar{U})} + \inf_{u \in \mathcal{F}_u} \|u - u^\dagger\|_{H^2(U)} \right),$$

where C_{app} is a constant depending on γ_0, γ_1, U and B_f .

The proof refers to [Appendix A.2](#).

We now turn to investigate the function approximation problems. To this end, we first introduce a type of activation functions.

Definition 3.6 (*Exponential PU-Admissible Function* [18]). Let $j \in \mathbb{N}$ and $\tau \in \{0, 1\}$. We say that a function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is exponential (j, τ) -PU-admissible, if

- (a) φ is bounded if $\tau = 0$, and φ is Lipschitz continuous if $\tau = 1$;
- (b) There exists $R > 0$ such that $\varphi \in C^j(\mathbb{R} \setminus [-R, R])$;
- (c) $\varphi' \in W^{j-1, \infty}(\mathbb{R})$, if $j \geq 1$;
- (d) There exists $A = A(\varphi), B = B(\varphi) \in \mathbb{R}$ with $A < B$, some $C = C(\varphi, j) > 0$ and some $D = D(\varphi, j) > 0$ such that
 - (d.1) $|B - \varphi^{(\tau)}(x)| \leq C \exp(-Dx)$ for all $x > R$;
 - (d.2) $|A - \varphi^{(\tau)}(x)| \leq C \exp(Dx)$ for all $x < -R$;
 - (d.3) $|\varphi^{(k)}(x)| \leq C \exp(-D|x|)$ for all $x \in \mathbb{R} \setminus [-R, R]$ and all $k = \tau + 1, \dots, j$.

Several commonly-used exponential PU-admissible activation functions are shown in [Table 1](#).

Applying [Proposition 4.8](#) in [18] and [Lemma 3.5](#), we have the follow approximation error estimate.

Lemma 3.7 (*Approximation Error*). Let $\mu > 0, d, s \in \mathbb{N}_+$ and $U \subseteq (0, 1)^d$ be a domain. Suppose $\gamma^\dagger \in C^{s+1}(\bar{U})$ and $u^\dagger \in H^{s+2}(U)$. Suppose $D_\gamma, S_\gamma, D_u, S_u \in \mathbb{N}_+$, and φ is an exponential PU-admissible activation function. Suppose [Assumptions 3.1](#) and [3.2](#) are fulfilled. Set

- (i) $\mathcal{F}_\gamma = \mathcal{N}_\varphi(D_\gamma, S_\gamma, \mathcal{B}_\gamma)$ with $D_\gamma = C \log(d+s)$ and $\mathcal{B}_\gamma = CS_\gamma^{\frac{2s}{d}+7}$, and
- (ii) $\mathcal{F}_u = \mathcal{N}_\varphi(D_u, S_u, \mathcal{B}_u)$ with $D_u = C \log(d+s+1)$ and $\mathcal{B}_u = CS_u^{\frac{2s+2}{d}+7}$,

where the constant C depends on d, s, μ and U . Then

$$\inf_{(\gamma,u) \in \mathcal{F}_\gamma \times \mathcal{F}_u} R(\gamma, u) \leq C_{\text{app}} \cdot B_\gamma^2 B_u^2 \left\{ \|\gamma^\dagger\|_{C^{s+1}(\bar{U})} \cdot S_\gamma^{-\frac{s-\mu}{d}} + \|u^\dagger\|_{H^{s+2}(U)} \cdot S_u^{-\frac{s+1-\mu}{d}} \right\},$$

where C_{app} is a constant depending on $d, s, \mu, \gamma_0, \gamma_1, U$ and B_f .

The proof refers to [Appendix A.2](#). [Lemma 3.7](#) shows that as long as the number of parameters is large enough, the neural networks can approximate the ground truth conductivity and voltage with arbitrarily small errors. In addition, when the size of the neural networks is the same, the smoother the ground truth conductivity and voltage are, the smaller the approximation error will be.

3.2. Statistical error

An estimate of the statistical error is presented in this section. With reference to the definition of statistical error in [Lemma 3.4](#), an upper bound on the statistical error will be obtained if we can estimate the metric entropy of the hypothesis classes \mathcal{F}_γ and \mathcal{F}_u . In [19], the author gives the L^∞ covering number bounds for some commonly used classes of neural networks. However, in PDE-related problems, we need to estimate not only the L^∞ covering number of the set of networks, but also that of the set of network derivatives. In other words, we need to bound the $W^{m, \infty}$ ($m \geq 1$) covering numbers of the set of networks. In [15, 16],

the derivatives of ReLU ^{k} -networks are implemented by ReLU-type networks, whose covering numbers can be bounded via [19]. Unfortunately, this approach cannot be extended to networks with other activation functions. In [10,17], the covering numbers of the derivatives of the networks are estimated by converting them to the covering numbers of the parameter class, which is a compact subset of \mathbb{R}^n , by Lipschitz continuity. This method requires some tedious manipulations. In this paper, we consider the classes of neural networks as relatively compact subsets of the smooth function spaces. We then estimate their metric entropy using the well-known entropy bounds for classes of smooth functions.

Lemma 3.8 (Statistical Error). Let $\mu > 0$, $s \in \mathbb{N}_+$ and $S \in \mathbb{N}_+$. Suppose $U \subset\subset (0, 1)^d$ be a domain with C^∞ -boundary. Let φ be an activation function which, together with its up to third-order derivatives, are continuous and uniformly bounded. Suppose [Assumption 3.2](#) is fulfilled. Set

- (i) $\mathcal{F}_\gamma = \mathcal{N}_\varphi(D_\gamma, S, B_\gamma)$ with $D_\gamma = C \log(d+s)$ and $B_\gamma = CS^{\frac{2s}{d}+7}$, and
- (ii) $\mathcal{F}_u = \mathcal{N}_\varphi(D_u, S, B_u)$ with $D_u = C \log(d+s+1)$ and $B_u = CS^{\frac{2s+2}{d}+7}$,

where the constant C depends on d , s , μ and U . Then

$$\begin{aligned} & \inf_{\varepsilon > 0} \left\{ (B_\gamma^2 B_u^2 + B_f^2 + \delta^2)(H_\gamma + H_u)n^{-1} + (\delta + B_\gamma B_u)(B_\gamma + B_u)\varepsilon \right\} \\ & \leq C_{\text{sta}} \cdot S^{\frac{6(4d+s+1)}{d+1} \log^3(d+s+1)} n^{-\frac{1}{d+1}}, \end{aligned}$$

where c is an absolute constant and C_{sta} is a constant depending on d , s , φ , δ , B_γ , B_u and U .

The proof refers to [Appendix A.3](#). Note that both sigmoid and tanh satisfy the conditions in [Lemma 3.8](#), as shown in [Table 1](#).

3.3. Convergence rate

Up to now, we have bounded the approximation error and the statistical error, respectively. [Lemmas 3.7](#) and [3.8](#) show that the approximation error will decrease with the size of the neural networks, while at the same time the statistical error will increase. By making a trade-off between the two, we obtain the following theorem.

Theorem 3.9 (Convergence Rate). Under the noise model Eq. (2.4). Let $\mu > 0$, $d, s \in \mathbb{N}_+$ and $U \subset\subset (0, 1)^d$ be a domain with C^∞ -boundary. Suppose $\gamma^\dagger \in C^{s+1}(\bar{U})$ and $u^\dagger \in H^{s+2}(U)$. Let φ be an exponential PU-admissible activation function which, together with its up to third-order derivatives, are continuous and uniformly bounded. Suppose [Assumptions 3.1](#) and [3.2](#) are fulfilled. Set $\mathcal{F}_\gamma = \mathcal{F}_u = \mathcal{N}_\varphi(D, S, B)$ with

$$D = C(d, s, \mu, U), \quad S = \mathcal{O}\left(n^{\frac{1}{6(4d+s+1)\log^3(d+s+1)}}\right) \quad \text{and} \quad B = \mathcal{O}\left(n^{\frac{s+7d}{3d(4d+s+1)\log^3(d+s+1)}}\right).$$

Then it follows for each estimator $(\tilde{y}_n, \tilde{u}_n)$ taking values in $\mathcal{F}_\gamma \times \mathcal{F}_u$ that

$$R(\tilde{y}_n, \tilde{u}_n) = \mathcal{O}\left(n^{-\frac{s-\mu}{7d(4d+s+1)\log^3(d+s+1)}}\right).$$

Theoretical guarantees for deep learning-based methods for solving PDEs have attracted considerable attention. In [15,16,20], the convergence rate of deep Ritz methods and PINNs with ReLU ^{k} networks is presented. In particular, the minimax optimal rate is proved in [20]. An error estimate of deep Ritz method with different activation functions is given in [17]. In addition, [21] gives an error analysis for deep Ritz methods for eigenvalue problems. However, limited work has been done on error analysis of deep learning based methods for inverse problems. In [10], a convergence rate is proposed for a neural network method for solving CDII. To the best of our knowledge, [Theorem 3.9](#) is the first rigorous theoretical analysis of PINNs for reconstructing non-constant coefficients.

[Theorem 3.9](#) demonstrates the consistency of the expected excess risk, that is, both the approximation error and the statistical error in [Lemma 3.4](#) vanish as the sample size $n \rightarrow \infty$, provided the number of parameters in networks is large enough. In addition, the convergence rate $\simeq \mathcal{O}(n^{-\frac{s}{c-d^2}})$ depends on the dimension d and the smoothness s of the ground truth conductivity and voltage. We first consider the dependence of the convergence rate on d . Note that the rate $\mathcal{O}(n^{-\frac{1}{c-d^2}})$ in [Theorem 3.9](#) is slower than $\mathcal{O}(n^{-\frac{1}{c-d}})$ in [10], which is because the statistical error bound in [Lemma 3.8](#) is not sharp enough. However, the number of parameters required in this paper $\mathcal{O}(n^{\frac{1}{c-d}})$ is significantly smaller than that $\mathcal{O}(n^{\frac{1}{c-d}})$ in [10] in high-dimensional problems. Furthermore, taking the smoothness into consideration, the rate $\simeq \mathcal{O}(n^{-\frac{s}{c-d^2}})$ shows that our approach may overcome the curse of dimensionality, provided the ground truth functions are smooth enough.

Remark 3.10. For simplicity of presentation, this paper only considers the case where the regularization parameter α is zero. In general, if the regularization parameter $\alpha > 0$, the error bound of the population loss is given by

$$\mathbb{E}_S \left[L_\lambda(\hat{q}, \hat{u}) \right] \leq \mathcal{O}\left(n^{-\frac{s-\mu}{7d(4d+s+1)\log^3(d+s+1)}}\right) + \mathcal{O}(\alpha) + \mathcal{O}(\delta^2).$$

By choosing the sample size n and the regularization parameter α are certain order of δ , we can obtain a convergence rate of the population loss in terms of the noise level δ .

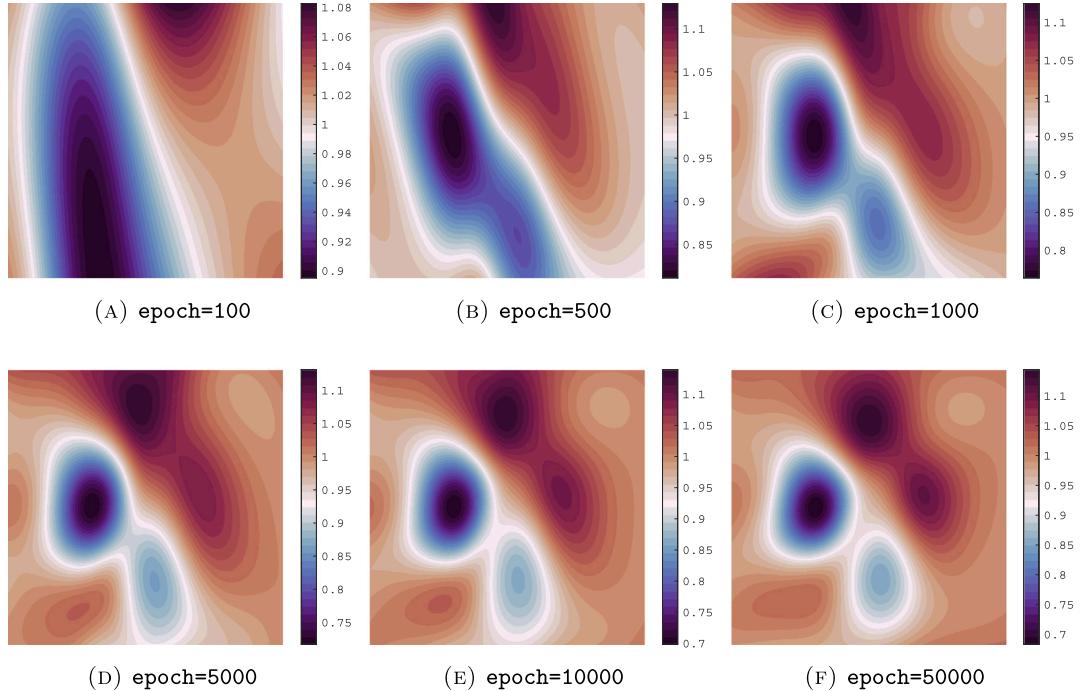


Fig. 1. The reconstructions of conductivity after different epochs. $\delta = 1\%$.

Remark 3.11. The analysis in this work focuses on the error estimation to the cost functional, which is applicable to any regularization parameter α . If the error analysis to the conductivity γ is needed, we need to choose the regularization parameter α properly, to ensure the stability of unknown conductivity from the cost functional with respect to the noise level. In that case, the number of samples should also depend on the noise level. We will leave the analysis to the conductivity γ in the future work.

4. Numerical experiments and discussions

Let $U = (0, 1)^2$. We use the two-to-one voltage potential of f , which equals the trace of the harmonic function $u_h(x, y) = y$. Given an admissible pair (γ, f) , we solve numerically the problem

$$\nabla \cdot (\gamma \nabla u) = 0, \quad u|_{\partial\Omega} = f. \quad (4.1)$$

Once the solution u is found, the interior data $a = \gamma |\nabla u|$ are computed. Then the noisy data a^δ is generated by adding Gaussian random noise

$$a^\delta(x) = a^\dagger(x) + \delta \cdot a^\dagger(x)\xi(x), \quad \xi(x) \sim N(0, 1).$$

In all of our experiments, we parameterize networks γ_ϕ and u_θ as four-layer \tanh -MLP with width 32 (or 64), of which the parameters are initialized by Xavier's method [22]. For each example, we use $n = 100000$ interior measurement data points and $n = 100000$ boundary points and run Algorithm 1 for 50000 epochs with batch size 2048. We minimize the loss by ADAM [23], and the learning rate is set as 1.0×10^{-3} , which does not change during the training. Thanks to the robustness of our methods, there is no need for denoising and other preprocessing to the noisy measurements. We use the relative L^2 -error $\text{err}(\hat{\gamma}) = \|\gamma^\dagger - \hat{\gamma}\|_2 / \|\gamma^\dagger\|_2$ to measure the accuracy of the reconstruction $\hat{\gamma}$. In Examples 4.1 and 4.2, we use L^2 -regularization $\psi(\gamma) = \|\gamma\|_{L^2(U)}$ with regularization parameter $\alpha = 1.0 \times 10^{-5}$, while both L^2 and TV regularization $\psi(\gamma) = |\nabla \gamma|_{L^1(U)}$ are applied in Example 4.3 with $\alpha = 1.0 \times 10^{-3}$ and $\alpha = 1.0 \times 10^{-3}$, respectively.

We solve Eq. (4.1) by MATLAB 2022b. Our models are implemented by PyTorch 12.1 [24], and trained with one NVIDIA Tesla V100 GPU.

Example 4.1. We employ the four-mode model conductivity distribution [1]

$$\gamma^\dagger(x, y) = 1 + \gamma_s(x, y),$$

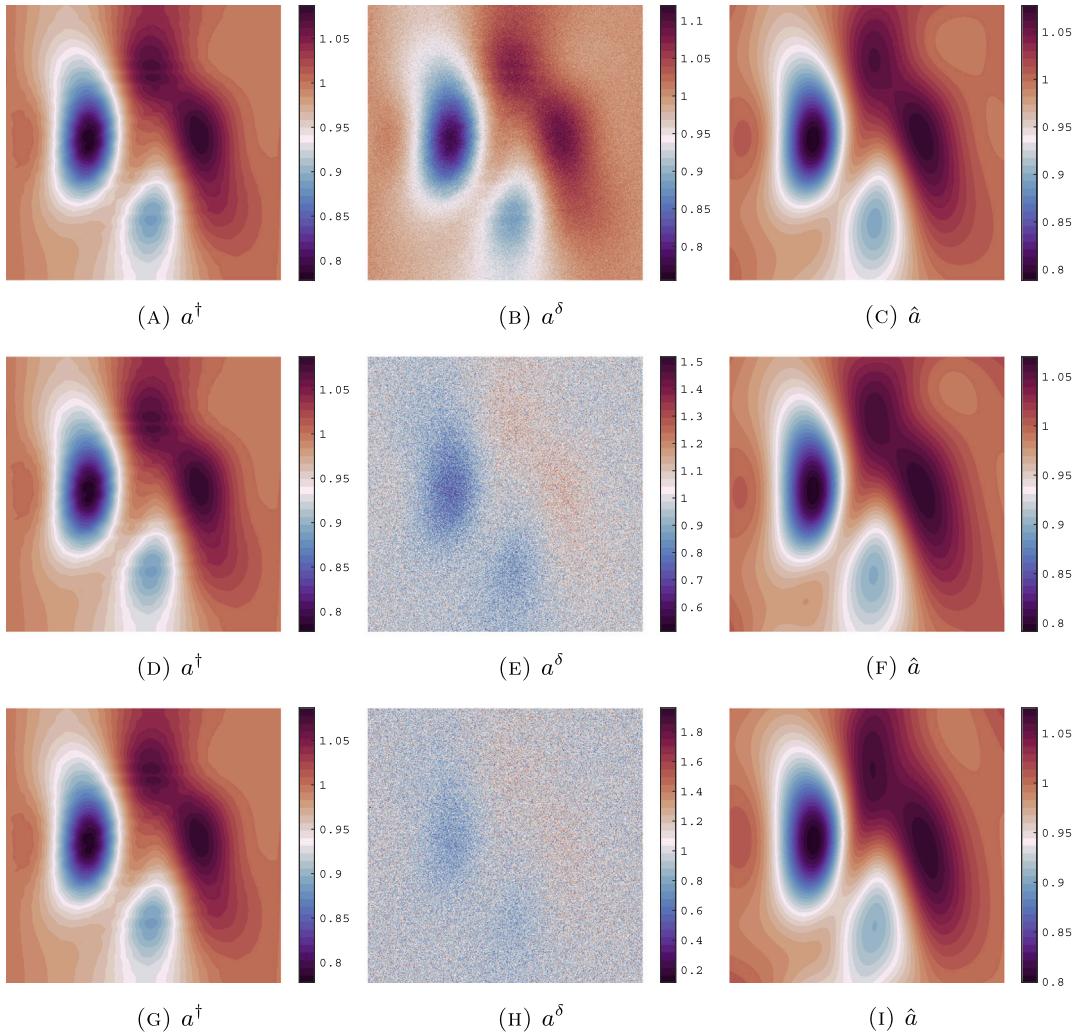


Fig. 2. The ground truth a^\dagger , noisy measurements a^δ , and reconstruction \hat{a} by our method in Example 4.1. $\delta = 1\%$ (top) $\delta = 10\%$ (middle) $\delta = 20\%$ (bottom).

where γ_s is a function with support in U , which is given by

$$\begin{aligned} \gamma_s(x, y) &= 0.3 \cdot (a(x, y) - b(x, y) - c(x, y)), \\ a(x, y) &= 0.3 \cdot (1 - 3(2x - 1))^2 \cdot \exp[-9 \cdot (2x - 1)^2 - (6y - 2)^2], \\ b(x, y) &= \left(\frac{3(2x - 1)}{5} - 27 \cdot (2x - 1)^3 - (3 \cdot (2y - 1))^5 \right) \exp[-(9 \cdot (2x - 1)^2 + 9 \cdot (2y - 1)^2)], \\ c(x, y) &= \exp[-(3 \cdot (2x - 1) + 1)^2 - 9 \cdot (2y - 1)^2]. \end{aligned}$$

We first apply our proposed method to the reconstruction of conductivity and voltage from data with 1% noise. Fig. 1 displays the reconstructed conductivity obtained by training the neural network for different epochs. Our results demonstrate that a rough approximation of the conductivity can be obtained after a few epochs, while further training is required to capture the fine-scale details. More exactly, we find that the neural network first learns to capture the low-frequency features, before fitting the high-frequency information. These findings align with the frequency principle [25].

Table 2 presents the relative errors of recovered conductivity and voltage from data with different noise levels, showing that our method is robust to the noise. Fig. 2 compares the measurement and recovered data at different noise levels. The reconstructed conductivity and voltage from the data at different noise levels and their point-wise absolute error are shown in Figs. 3 and 4.

Example 4.2 (CDII with Discontinuous Conductivity). Let $\rho(x, y) = \|(x, y) - (\frac{1}{2}, \frac{1}{2})\|_2$.

$$\gamma^\dagger(x, y) = 1 + \chi_{\{x > \frac{1}{2}\}} \exp(-2\rho^2).$$

Table 2
The relative L^2 -error of the recovered conductivity $\hat{\gamma}$ and voltage \hat{u} in Example 4.1.

δ	$\hat{\gamma}$		
	1%	10%	20%
err(γ)	2.86×10^{-2}	3.22×10^{-2}	3.75×10^{-2}
err(u)	3.03×10^{-3}	3.85×10^{-3}	4.22×10^{-3}

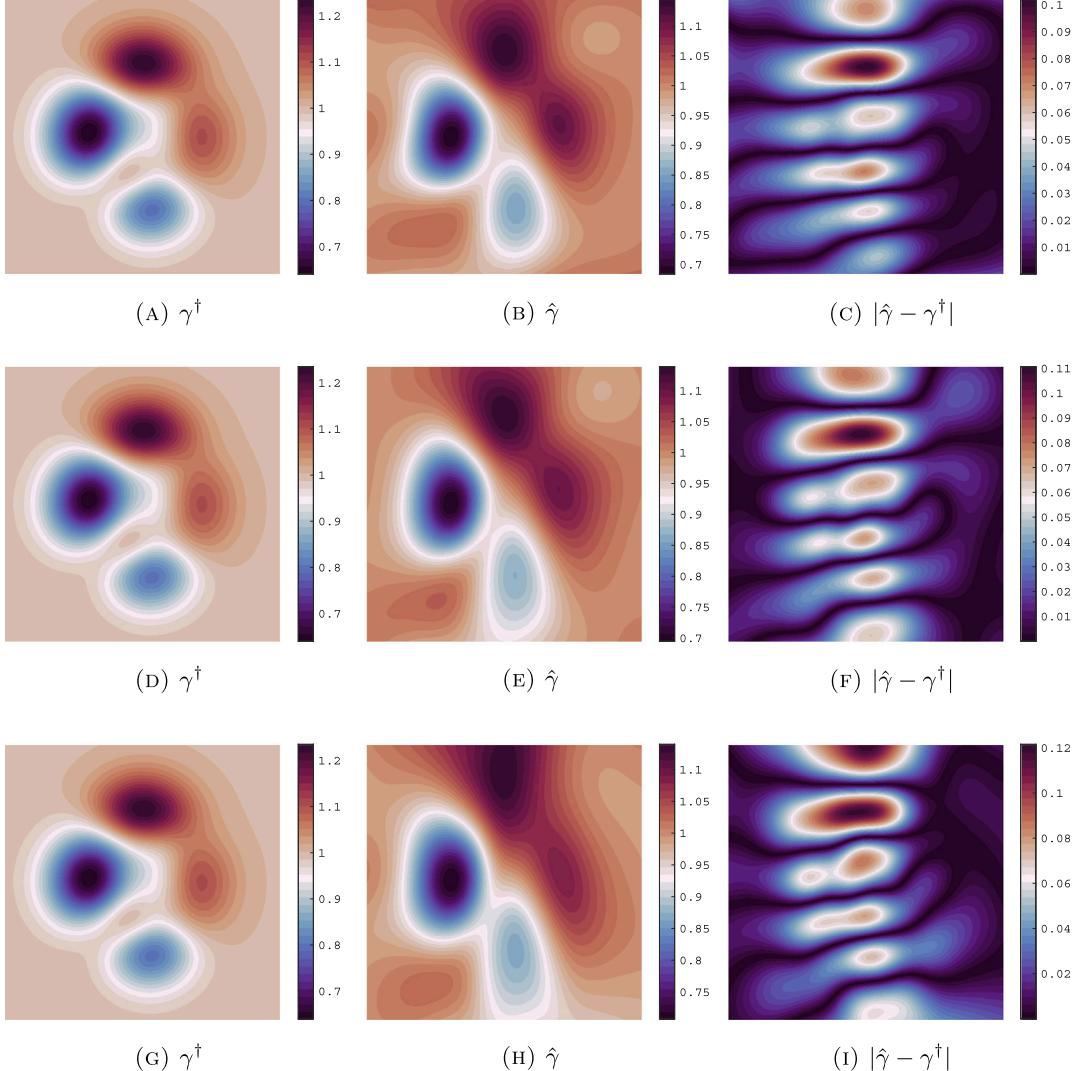


Fig. 3. The ground truth conductivity γ^\dagger , reconstruction $\hat{\gamma}$ by our method, and the point-wise absolute error $|\hat{\gamma} - \gamma^\dagger|$ in Example 4.1. $\delta = 1\%$ (top) $\delta = 10\%$ (middle) $\delta = 20\%$ (bottom).

We apply our approach to reconstruct the conductivity and voltage from the measurement data with 1%, 10%, and 20% noise, respectively. The relative L^2 -error of the recovered voltage \hat{u} and conductivity $\hat{\gamma}$ are shown in Table 3. Our results demonstrate that the proposed method can effectively reconstruct both the conductivity and voltage from data sets with noise levels of up to 20%, while maintaining high accuracy. At the same time, the accuracy does not decrease significantly as the noise level increases. We also compare the ground truth measurement data, noisy data, and recovered data in Fig. 5. As shown in Figs. 6 and 7, the error in the reconstructed conductivity and voltage is primarily concentrated at the discontinuous interface, which aligns with the findings of [10]. Additionally, our method has no significant advantage over that in [10] on this example, mainly because our method is based on a strong form of the PDE.

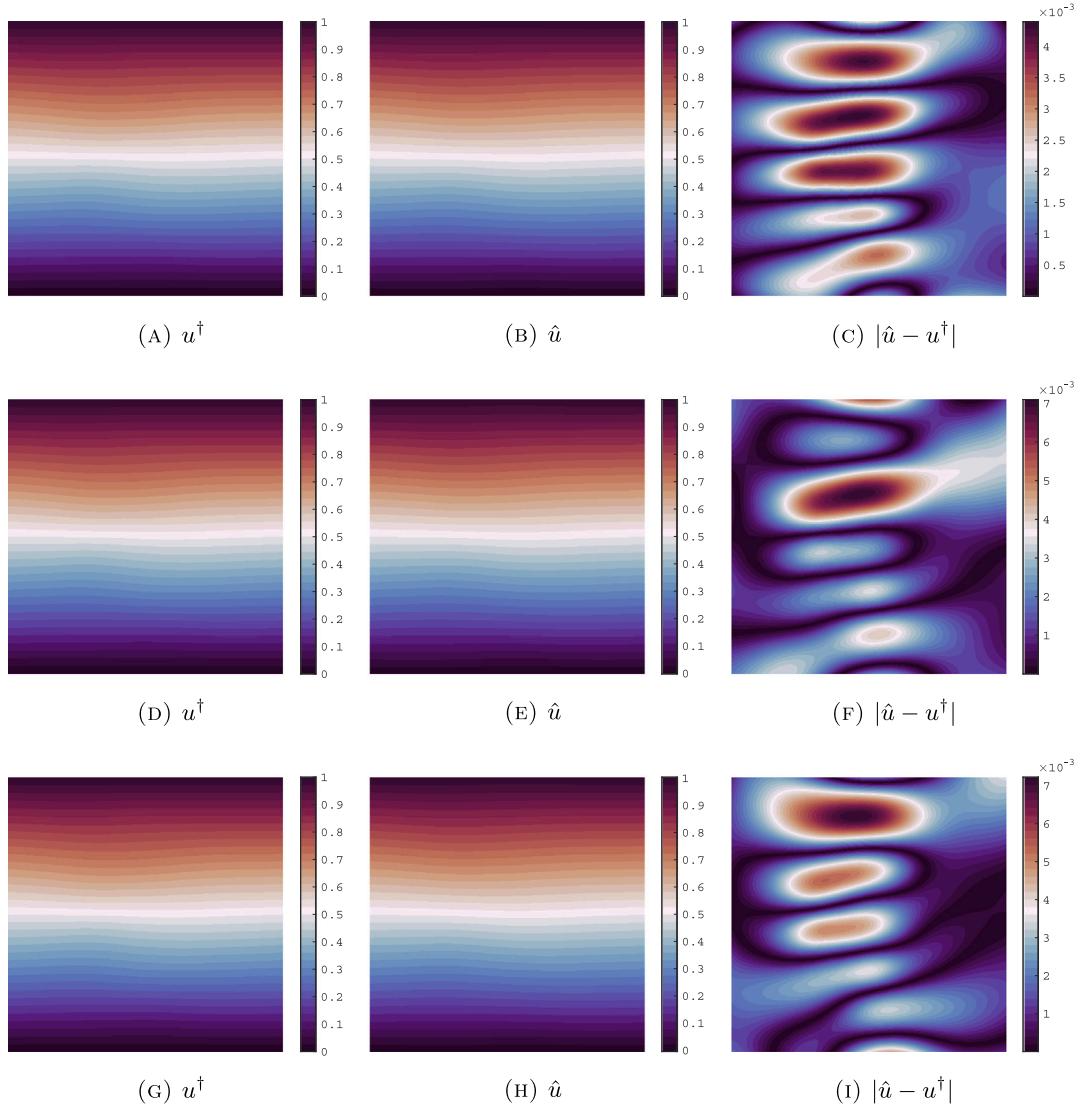


Fig. 4. The ground truth voltage u^\dagger , reconstruction \hat{u} by our method, and the point-wise absolute error $|\hat{u} - u^\dagger|$ in Example 4.1. $\delta = 1\%$ (top) $\delta = 10\%$ (middle) $\delta = 20\%$ (bottom).

Table 3
The relative L^2 -error of the recovered conductivity $\hat{\gamma}$ and voltage \hat{u} in Example 4.2.

δ	1%	10%	20%
err(γ)	3.71×10^{-2}	4.38×10^{-2}	4.81×10^{-2}
err(u)	8.48×10^{-3}	9.08×10^{-3}	9.67×10^{-3}

Example 4.3 (CDII with Disjoint Modes in Conductivity). Let $\Omega_1 = \{(x, y) : 100(x - 0.3)^2 + 36(y - 0.7)^2 - 72(x - 0.3)(y - 0.7) < 1\}$ and $\Omega_2 = \{(x, y) : 36(x - 0.6)^2 + 36(y - 0.4)^2 < 1\}$. We set the ground truth conductivity as

$$\gamma^\dagger(x, y) = 1 + \chi_{\Omega_1} - \chi_{\Omega_2}.$$

In this example, we set the noise level $\delta = 10\%$ and use the TV-regularization. Referring to (2.7) in [10], we replace the term $|\nabla \gamma(x)|$ in TV-regularization with Huber function

$$h(\gamma) = \begin{cases} |\nabla \gamma(x)|, & |\nabla \gamma(x)| \geq \zeta, \\ \frac{|\nabla \gamma(x)|^2}{2\zeta} + \frac{\zeta}{2}, & \text{otherwise,} \end{cases}$$

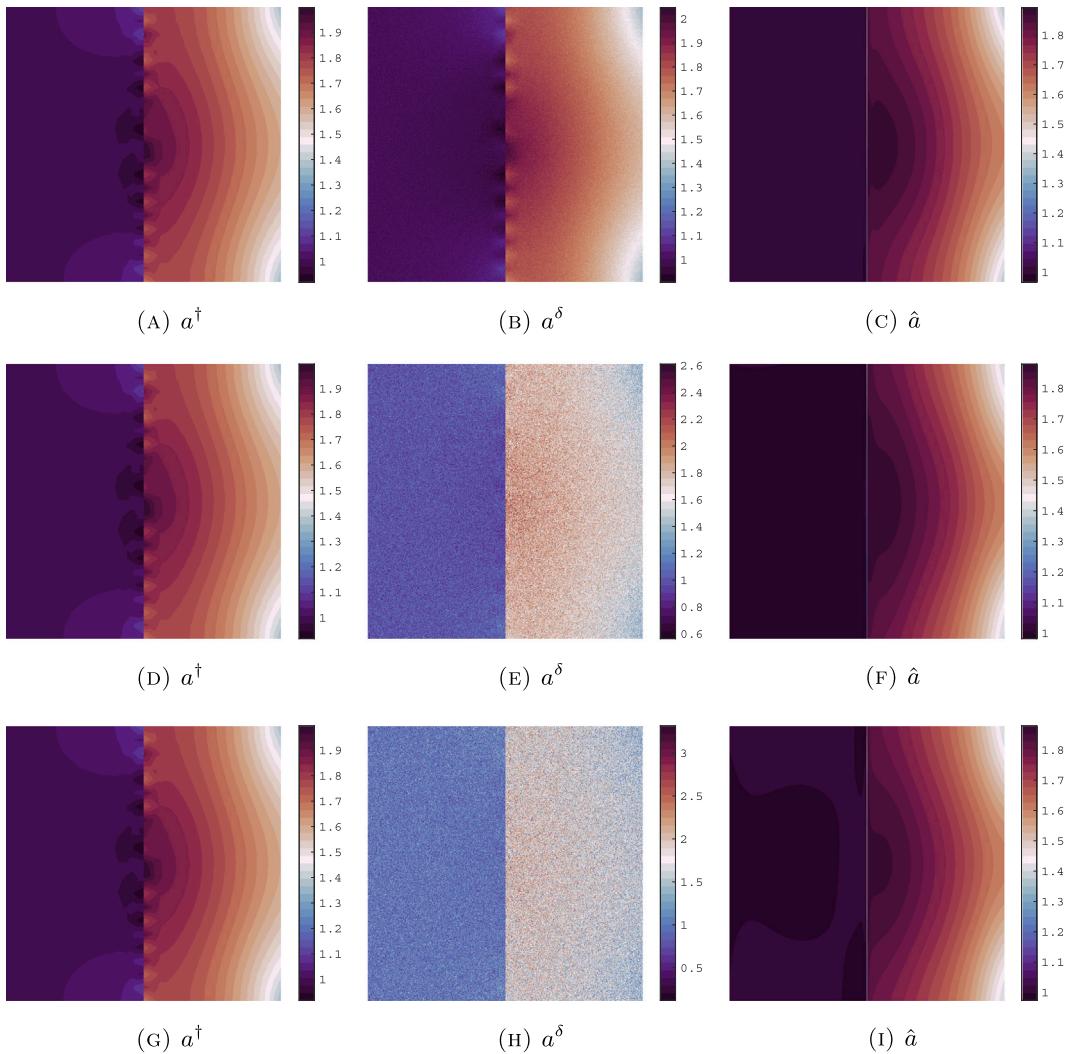


Fig. 5. The ground truth a^\dagger , noisy measurements a^δ , and reconstruction \hat{a} by our method in Example 4.2. $\delta = 1\%$ (top) $\delta = 10\%$ (middle) $\delta = 20\%$ (bottom).

Table 4
The relative L^2 -error of the recovered conductivity $\hat{\gamma}$ and voltage \hat{u} obtained by L^2 -regularization and TV-regularization in Example 4.3.

	L^2 -regularization	TV-regularization
$\text{err}(\gamma)$	7.19×10^{-2}	5.98×10^{-2}
$\text{err}(u)$	7.62×10^{-3}	7.13×10^{-3}

where $\zeta > 0$ is a small constant. We set $\zeta = 0.001$ in this example (see Table 4).

The noisy data a^δ and its reconstructed counterpart \hat{a} are illustrated in Fig. 8. Fig. 9 shows a comparison between ground truth conductivity γ^\dagger and its reconstruction $\hat{\gamma}$. As can be seen, our method determines the shape of the different modes precisely, and thanks to the TV-regularization, it can capture the nature of the piece-wise constants of the ground truth conductivity, while the L^2 -regularization cannot. A comparison between the ground truth voltage u^\dagger and its reconstruction \hat{u} is shown in Fig. 10.

5. Conclusions

In this paper, we propose a PINNs-based method for solving CDII. We first use two neural networks to represent the conductivity and voltage and then construct the loss function over the measurement data. Our method directly reconstructs the conductivity and voltage by minimizing the loss function, avoiding iterative updates. We present an error estimate and give a convergence rate. At

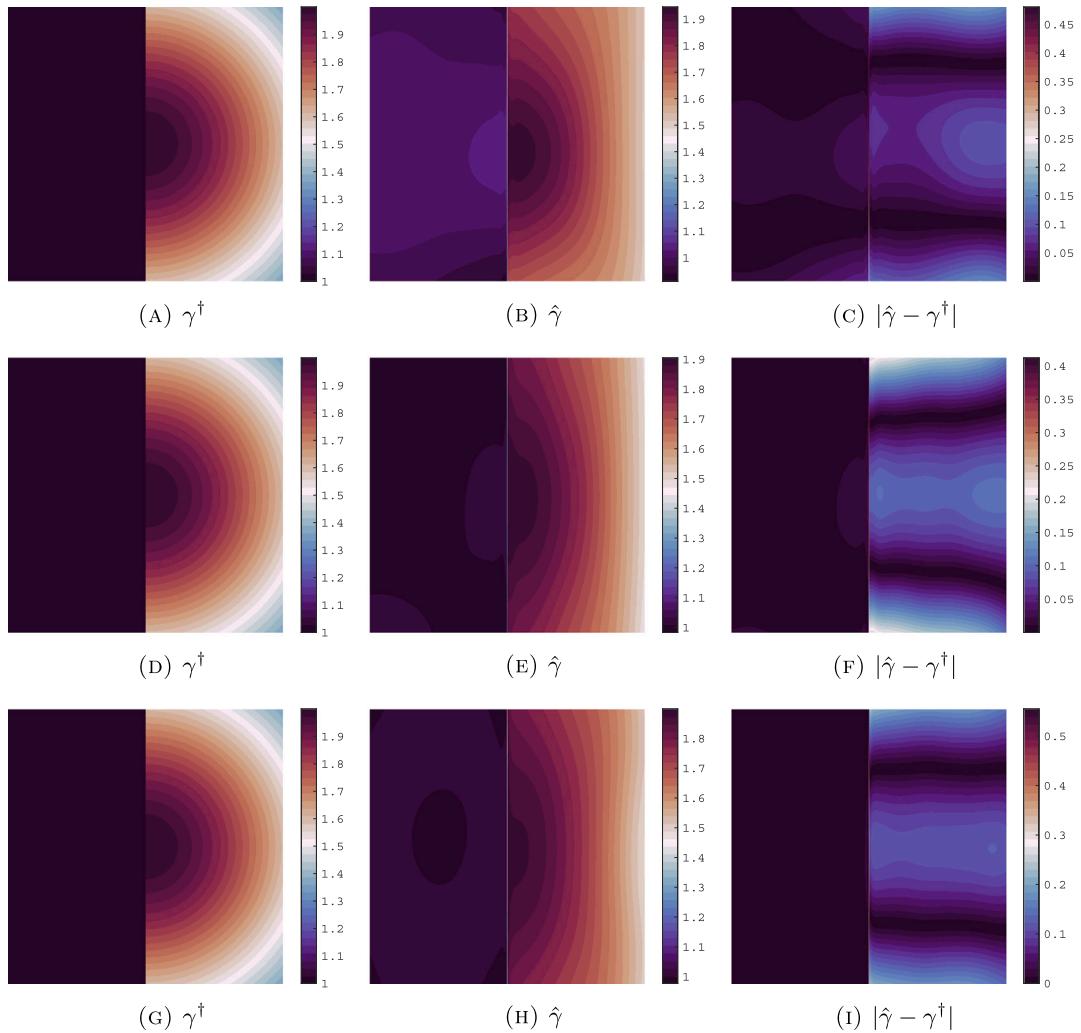


Fig. 6. The ground truth conductivity γ^\dagger , reconstruction $\hat{\gamma}$ by our method, and the point-wise absolute error $|\hat{\gamma} - \gamma^\dagger|$ in [Example 4.2](#). $\delta = 1\%$ (top) $\delta = 10\%$ (middle) $\delta = 20\%$ (bottom).

the same time, our error analysis can provide a way to choose the structure of the neural networks. The stability of CDII will be taken into account in future work. The method has shown robustness to noise in numerical experiments. Finally, the coupled-PINNs method can be directly extended to general inverse problems, although we have only considered CDII in this paper. Simultaneously, the error analysis framework and techniques are readily applicable.

Data availability

No data was used for the research described in the article.

Acknowledgments

We would like to thank the anonymous referees and associated editor for their useful comments and suggestions, which have led to considerable improvements in the paper. This work is supported by the National Key Research and Development Program of China (No. 2023YFA1000103), by the National Science Foundation of China (No. 12125103, No. 12071362, No. 12371424, No. 12371441). The numerical calculations have been done at the Supercomputing Center of Wuhan University.

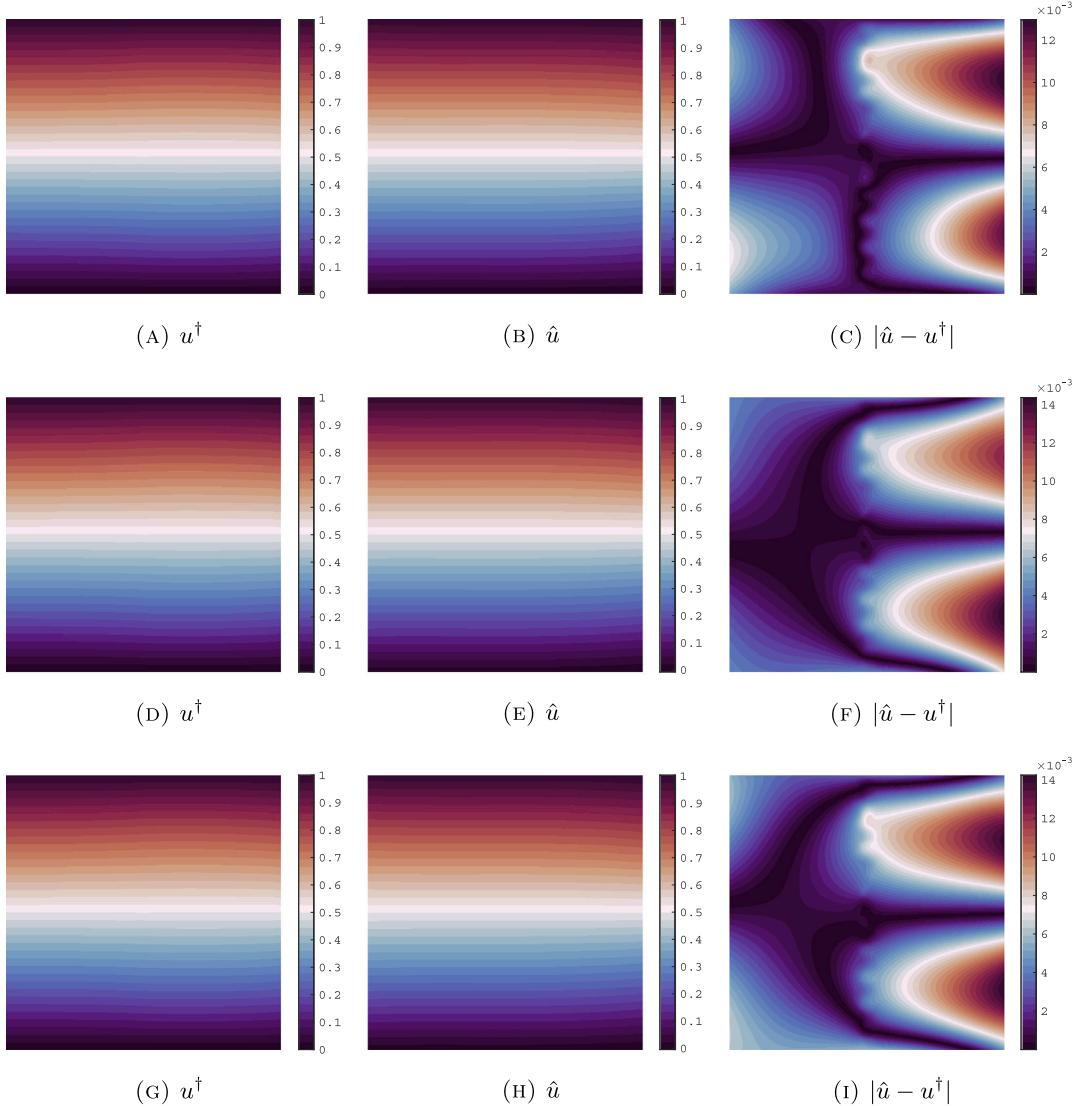


Fig. 7. The ground truth voltage u^\dagger , reconstruction \hat{u} by our method, and the point-wise absolute error $|\hat{u} - u^\dagger|$ in Example 4.2. $\delta = 1\%$ (top) $\delta = 10\%$ (middle) $\delta = 20\%$ (bottom).

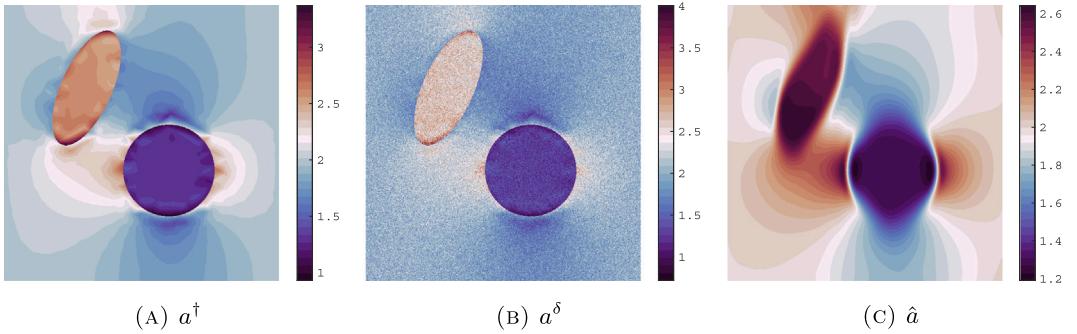


Fig. 8. The ground truth a^\dagger , noisy measurements a^δ with $\delta = 10\%$, and reconstruction \hat{a} by our method in Example 4.3. TV-regularization.

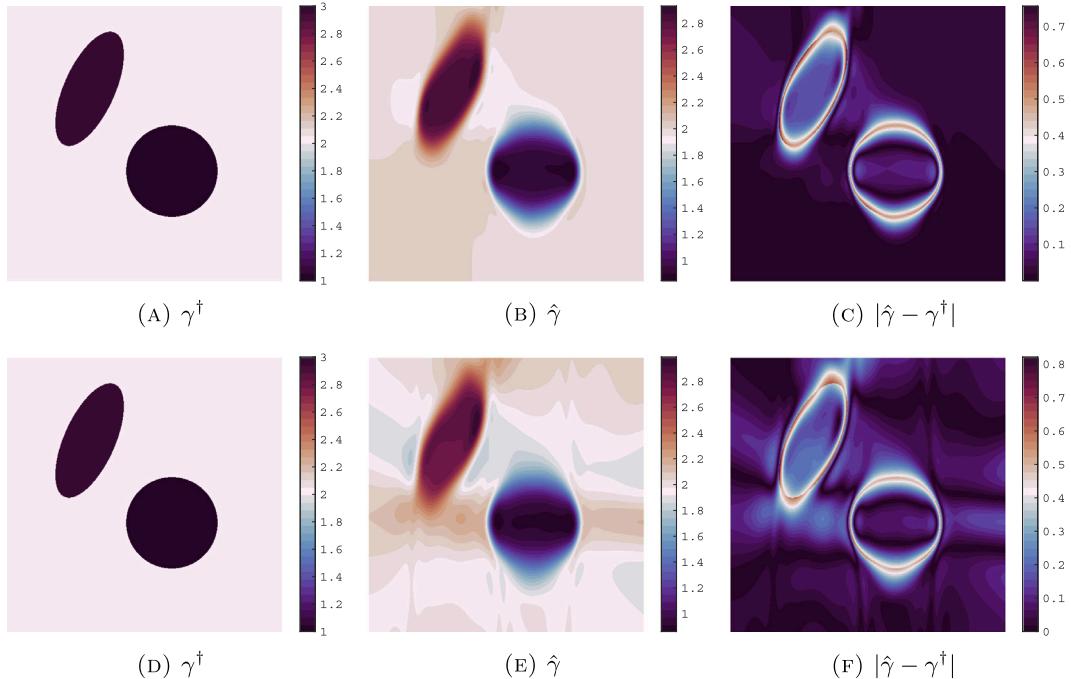


Fig. 9. The ground truth conductivity γ^\dagger , reconstruction $\hat{\gamma}$ by our method, and the point-wise absolute error $|\hat{\gamma} - \gamma^\dagger|$ in Example 4.3. $\delta = 10\%$. TV-regularization (top) L^2 -regularization (bottom).

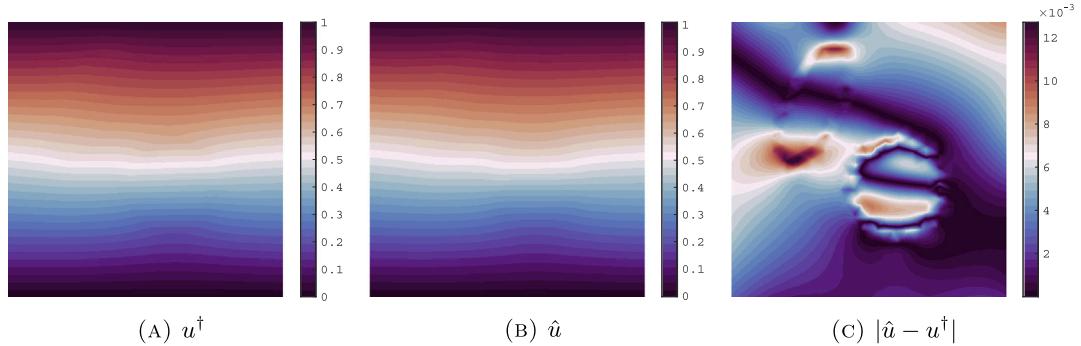


Fig. 10. The ground truth voltage u^* , reconstruction \hat{u} by our method, and the point-wise absolute error $|\hat{u} - u^*|$ in Example 4.3. $\delta = 10\%$ TV-regularization.

Appendix. Proofs

A.1. The proof of the error decomposition

Inspired by the proof of Lemma 4 in [26], we prove Lemma 3.4 in this section.

For an estimator $(\tilde{y}_n, \tilde{u}_n) \in \mathcal{F}_Y \times \mathcal{F}_U$ depending upon a data sample S , we introduce its expected empirical risk

$$R_n(\tilde{Y}_n, \tilde{u}_n) := \mathbb{E}_S \left[\frac{1}{n} \sum_{i=1}^n (\tilde{Y}_n | \nabla \tilde{u}_n |(X_i) - \gamma^\dagger | \nabla u^\dagger |(X_i))^2 \right] + \mathbb{E}_S \left[G_n(\tilde{Y}_n, \tilde{u}_n) \right],$$

and recall its optimization error

$$\Delta_n(\tilde{\gamma}_n, \tilde{u}_n) = \mathbb{E}_S \left[L_n(\tilde{\gamma}_n, \tilde{u}_n) - L_n(\hat{\gamma}_n, \hat{u}_n) \right].$$

Lemma A.1. Let \mathcal{F}_γ and \mathcal{F}_u be two function classes with $N_\gamma = N(\epsilon, \mathcal{F}_\gamma, \|\cdot; W^{1,\infty}(U)\|)$ and $N_u = N(\epsilon, \mathcal{F}_u, \|\cdot; W^{2,\infty}(U)\|)$. Suppose Assumptions 3.1 and 3.2 are fulfilled. Then it holds for each estimator $(\tilde{\gamma}_n, \tilde{u}_n)$ taking values in $\mathcal{F}_\gamma \times \mathcal{F}_u$ that

$$R(\tilde{\gamma}_n, \tilde{u}_n) \lesssim R_n(\tilde{\gamma}_n, \tilde{u}_n) + (B_\gamma^2 B_u^2 + B_f^2) \frac{\log(N_\gamma N_u)}{n} + B_\gamma B_u (B_\gamma + B_u) \epsilon,$$

with $\delta > 0$ and n large enough.

Proof. The proof is divided to four parts as follows.

Step 1. For any $\epsilon > 0$, let $C_\gamma = \{\gamma_k : k = 1, \dots, N_\gamma\}$ be the minimal ϵ -cover of \mathcal{F}_γ with respect to $\|\cdot\|_{W^{1,\infty}(U)}$, and let $C_u = \{u_\ell : \ell = 1, \dots, N_u\}$ be the minimal ϵ -cover of \mathcal{F}_u with respect to $\|\cdot\|_{W^{2,\infty}(U)}$. Then for an estimator $(\tilde{\gamma}_n, \tilde{u}_n) \in \mathcal{F}_\gamma \times \mathcal{F}_u$, there exists $(\gamma_{k^*}, u_{\ell^*}) \in C_\gamma \times C_u$, such that

$$\|\gamma_{k^*} - \tilde{\gamma}_n\|_{W^{1,\infty}(U)} \leq \epsilon \quad \text{and} \quad \|u_{\ell^*} - \tilde{u}_n\|_{W^{2,\infty}(U)} \leq \epsilon.$$

Without loss of generality, we can assume that $\|\gamma_k\|_{W^{1,\infty}(U)} \leq B_\gamma$ and $\|u_\ell\|_{W^{2,\infty}(U)} \leq B_u$. Generate a ghost sample $S' = \{(X'_i, Y'_i)\}_{i=1}^n \cup \{\bar{X}'_i\}_{i=1}^n$ independent of S , where (X'_i, Y'_i) and \bar{X}'_i are n independent copies of (X, Y) and \bar{X} , respectively. Then we have

$$\begin{aligned} |R(\tilde{\gamma}_n, \tilde{u}_n) - R_n(\tilde{\gamma}_n, \tilde{u}_n)| &= \left| \mathbb{E}_{S'} \left[\frac{1}{n} \sum_{i=1}^n (\tilde{\gamma}_n |\nabla \tilde{u}_n|(Z'_i) - \gamma^\dagger |\nabla u^\dagger|(Z'_i))^2 \right] \right. \\ &\quad \left. - \mathbb{E}_S \left[\frac{1}{n} \sum_{i=1}^n (\tilde{\gamma}_n |\nabla \tilde{u}_n|(Z_i) - \gamma^\dagger |\nabla u^\dagger|(Z_i))^2 \right] \right. \\ &\quad \left. + \mathbb{E}_{S'} \left[\frac{1}{n} \sum_{i=1}^n g(\tilde{\gamma}_n, \tilde{u}_n, Z'_i) \right] - \mathbb{E}_S \left[\frac{1}{n} \sum_{i=1}^n g(\tilde{\gamma}_n, \tilde{u}_n, Z_i) \right] \right| \\ &\leq \mathbb{E}_S \mathbb{E}_{S'} \left[\left| \frac{1}{n} \sum_{i=1}^n \varphi_{k^*, \ell^*}(Z_i, Z'_i) \right| \right] + 28 B_\gamma B_u (B_\gamma + B_u) \epsilon, \end{aligned} \tag{5.1}$$

where

$$\begin{aligned} g(\gamma, u, X, \bar{X}) &:= (\nabla \cdot (\gamma \nabla u)(X))^2 + (Tu(\bar{X}) - f(\bar{X}))^2 \\ h(\gamma_k, u_\ell, X, \bar{X}) &:= (\gamma_k |\nabla u_\ell|(X) - \gamma^\dagger |\nabla u^\dagger|(X))^2 + g(\gamma_k, u_\ell, X, \bar{X}), \\ \varphi_{k, \ell}(X, \bar{X}, X', \bar{X}') &:= h(\gamma_k, u_\ell, X, \bar{X}) - h(\gamma_k, u_\ell, X', \bar{X}'), \end{aligned}$$

for $k = 1, \dots, N_\gamma$ and $\ell = 1, \dots, N_u$.

Step 2. Set $r_{k, \ell} = \max\{A, \mathbb{E}_{(X, \bar{X})}^{1/2}[h(\gamma_k, u_\ell, X, \bar{X})]\}$ with $A > 0$ for $k = 1, \dots, N_\gamma$ and $\ell = 1, \dots, N_u$, then

$$\begin{aligned} r_{k^*, \ell^*}^2 &= \max\{A^2, \mathbb{E}_{(X, \bar{X})}[h(\gamma_{k^*}, u_{\ell^*}, X, \bar{X})|S]\} \\ &\leq A^2 + \mathbb{E}_{(X, \bar{X})}[h(\tilde{\gamma}_n, \tilde{u}_n, X, \bar{X})|S] + 14 B_\gamma B_u (B_\gamma + B_u) \epsilon, \end{aligned} \tag{5.2}$$

where the last inequality is follows from $\|\tilde{\gamma}_n - \gamma_{k^*}\|_{W^{1,\infty}(U)} \leq \epsilon$ and $\|\tilde{u}_n - u_{\ell^*}\|_{W^{2,\infty}(U)} \leq \epsilon$. Define a random variable

$$T = \max_{k, \ell} \left| \frac{1}{nr_{k, \ell}} \sum_{i=1}^n \varphi_{k, \ell}(X_i, \bar{X}_i, X'_i, \bar{X}'_i) \right|, \tag{5.3}$$

then by (5.1) we obtain by using Cauchy-Schwarz and AM-GM inequality

$$\begin{aligned} |R(\tilde{\gamma}_n, \tilde{u}_n) - R_n(\tilde{\gamma}_n, \tilde{u}_n)| &\leq \mathbb{E}_{S, S'}[r_{k^*, \ell^*} T] + 28 B_\gamma B_u (B_\gamma + B_u) \epsilon \\ &\leq \frac{1}{2} \mathbb{E}_S[r_{k^*, \ell^*}^2] + \frac{1}{2} \mathbb{E}_{S, S'}[T^2] + 28 B_\gamma B_u (B_\gamma + B_u) \epsilon. \end{aligned} \tag{5.4}$$

Step 3. Now we turn to estimate $\mathbb{E}_S[r_{k^*, \ell^*}^2]$ and $\mathbb{E}_{S, S'}[T^2]$. By (5.2), $\mathbb{E}_S[r_{k^*, \ell^*}^2]$ can be evaluated by

$$\begin{aligned} \mathbb{E}_S[r_{k^*, \ell^*}^2] &\leq A^2 + \mathbb{E}_S \mathbb{E}_{(X, \bar{X})}[h(\tilde{\gamma}_n, \tilde{u}_n, X, \bar{X})|S] + 14 B_\gamma B_u (B_\gamma + B_u) \epsilon \\ &\leq A^2 + R(\tilde{\gamma}_n, \tilde{u}_n) + 14 B_\gamma B_u (B_\gamma + B_u) \epsilon. \end{aligned} \tag{5.5}$$

In order to estimate $\mathbb{E}_{S, S'}[T^2]$, we bound the tail probability $\mathbb{P}\{T > t\}$ by Bernstein inequality. Observe that $\mathbb{E}[\varphi_{k, \ell}(X, \bar{X}, X', \bar{X}')] = 0$, $|\varphi_{k, \ell}(X, \bar{X}, X', \bar{X}')| \leq 12 B_\gamma^2 B_u^2 + 2(B_f^2)$, and

$$\begin{aligned} \sigma_\varphi^2 &= \text{Var}(\varphi_{k, \ell}(X, \bar{X}, X', \bar{X}')) = 2 \text{Var}(h(\gamma_k, u_\ell, X, \bar{X})) \leq 2 \mathbb{E}_{(X, \bar{X})}[h^2(\gamma_k, u_\ell, X, \bar{X})] \\ &\leq (12 B_\gamma^2 B_u^2 + 2B_f^2) \mathbb{E}_{(X, \bar{X})}[h(\gamma_k, u_\ell, X, \bar{X})] \leq (12 B_\gamma^2 B_u^2 + 2B_f^2) r_{k, \ell}^2, \end{aligned}$$

where the second inequality is owing to $h(\gamma_k, u_\ell, X, \bar{X}) \geq 0$. Then using Bernstein inequality gives

$$\begin{aligned} \mathbb{P}\left\{ \left| \frac{\sum_{i=1}^n \varphi_{k, \ell}(X_i, \bar{X}_i, X'_i, \bar{X}'_i)}{n} \right| \geq t \right\} &\leq 2 \exp\left(-\frac{1}{4(6B_\gamma^2 B_u^2 + B_f^2)} \cdot \frac{nt^2}{t/3 + r_{k, \ell}^2}\right). \end{aligned} \tag{5.6}$$

For ease of notations, we denote $B = 4(6B_\gamma^2 B_u^2 + B_f^2)$. Since $A \leq \min_{k,\ell} r_{k,\ell}$, we obtain by (5.6)

$$\begin{aligned}\mathbb{P}\{T \geq t\} &\leq \sum_{k=1}^{N_\gamma} \sum_{\ell=1}^{N_u} \mathbb{P}\left\{\left|\frac{\sum_{i=1}^n \varphi_{k,\ell}(Z_i, Z'_i)}{nr_{k,\ell}}\right| \geq t\right\} \\ &\leq N_\gamma N_u \max_{k,\ell} \mathbb{P}\left\{\left|\frac{\sum_{i=1}^n \varphi_{k,\ell}(Z_i, Z'_i)}{n}\right| \geq tr_{k,\ell}\right\} \\ &\leq 2N_\gamma N_u \exp\left(-\frac{1}{B} \cdot \frac{nt^2}{t/3A+1}\right) \leq 2N_\gamma N_u \exp\left(-\frac{3Ant}{2B}\right),\end{aligned}$$

where the last inequality holds for $t \geq 3A$. Hence for $b \geq 3A$, it follows that

$$\begin{aligned}\mathbb{E}_{S,S'}[T^2] &= \int_0^\infty \mathbb{P}\{T^2 \geq u\} du = \int_0^\infty \mathbb{P}\{T^2 \geq \sqrt{u}\} du \\ &\leq b^2 + 2N_\gamma N_u \int_{b^2}^\infty \exp\left(-\frac{3An\sqrt{u}}{2B}\right) du \\ &\leq b^2 + 4N_\gamma N_u \left(\frac{2B}{3An}\right)^2 \left(1 + \frac{3Anb}{2B}\right) \exp\left(-\frac{3Anb}{2B}\right).\end{aligned}$$

Assume $5 \leq \log N_\gamma N_u \leq n$, setting $b = \frac{2B}{3An} \log(N_\gamma N_u)$ gives

$$\mathbb{E}_{S,S'}[T^2] \leq \left(\frac{2B}{3An}\right)^2 (\log^2(N_\gamma N_u) + \log(N_\gamma N_u) + 4) \leq 2\left(\frac{2B}{3A}\right)^2 \left(\frac{\log(N_\gamma N_u)}{n}\right)^2. \quad (5.7)$$

Step 4. Combining (5.4)(5.5)(5.7) yields

$$\begin{aligned}|R(\tilde{\gamma}_n, \tilde{u}_n) - R_n(\tilde{\gamma}_n, \tilde{u}_n)| \\ \leq \frac{1}{2} R(\tilde{\gamma}_n, \tilde{u}_n) + \frac{1}{2} A^2 + \left(\frac{2B}{3A}\right)^2 \left(\frac{\log(N_\gamma N_u)}{n}\right)^2 + 35B_\gamma B_u(B_\gamma + B_u)\varepsilon.\end{aligned}$$

Setting $A = \sqrt{\frac{2B \log(N_\gamma N_u)}{9n}}$ gives the result. \square

In order to estimate the expected empirical risk, we first introduce the following lemma.

Lemma A.2. Let $\eta_j \sim \text{subG}(0, \delta^2)$ for $j = 1, \dots, N$, then

$$\mathbb{E}\left[\max_{1 \leq j \leq N} \eta_j^2\right] \leq 4\varepsilon^2(1 + \log N).$$

Proof. For any $0 < t < (2\delta^2)^{-1}$, we can derive

$$\begin{aligned}\exp\left(t\mathbb{E}\left[\max_{1 \leq j \leq N} \eta_j^2\right]\right) &\leq \mathbb{E}\left[\max_{1 \leq j \leq N} \exp(t\eta_j^2)\right] \leq N\mathbb{E}[\exp(t\eta_1^2)] \\ &\leq \frac{N}{\sqrt{2\pi\varepsilon^2}} \int_{\mathbb{R}} \exp(tx^2) \exp\left(-\frac{x^2}{2\delta^2}\right) dx \\ &= \frac{N}{\sqrt{1 - 2\delta^2t}}.\end{aligned}$$

Therefore

$$\mathbb{E}\left[\max_{1 \leq j \leq N} \eta_j^2\right] \leq \frac{1}{t} \log \frac{N}{\sqrt{1 - 2\delta^2t}}.$$

Setting $t = (4\delta^2)^{-1}$ completes the proof. \square

Lemma A.3. Suppose Assumptions 3.1 and 3.2 hold.

$$R_n(\tilde{\gamma}_n, \tilde{u}_n) \lesssim \inf_{(\gamma,u) \in \mathcal{F}_\gamma \times \mathcal{F}_u} R(\gamma, u) + \Delta_n(\tilde{\gamma}_n, \tilde{u}_n) + \frac{\delta^2 \log(N_\gamma N_u)}{n} + (\delta + B_\gamma B_u)(B_\gamma + B_u)\varepsilon.$$

Proof. We divide the proof into three steps.

Step 1. For any fixed $(\gamma, u) \in \mathcal{F}_\gamma \times \mathcal{F}_u$, it is easy to show that

$$\mathbb{E}_S[L_n(\tilde{\gamma}_n, \tilde{u}_n)] \leq \mathbb{E}_S[L_n(\gamma, u)] + \Delta_n(\tilde{\gamma}_n, \tilde{u}_n). \quad (5.8)$$

By definition, we have

$$\begin{aligned}
\mathbb{E}_S[L_n(\tilde{\gamma}_n, \tilde{u}_n)] &= \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i))^2\right] + \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n g(\tilde{\gamma}_n, \tilde{u}_n, X_i, \bar{X}_i)\right] \\
&= \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n (\gamma^\dagger |\nabla u^\dagger|(X_i) - \tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i) + \xi_i)^2\right] + \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n g(\tilde{\gamma}_n, \tilde{u}_n, X_i, \bar{X}_i)\right] \\
&= R_n(\tilde{\gamma}_n, \tilde{u}_n) + \delta^2 - \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i \tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i)\right],
\end{aligned} \tag{5.9}$$

and

$$\begin{aligned}
\mathbb{E}_S[L_n(\gamma, u)] &= \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n (Y_i - \gamma |\nabla u|(X_i))^2\right] + \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n g(\gamma, u, X_i, \bar{X}_i)\right] \\
&= \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n (\gamma^\dagger |\nabla u^\dagger|(X_i) - \gamma |\nabla u|(X_i) + \xi_i)^2\right] + \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n g(\gamma, u, X_i, \bar{X}_i)\right] \\
&= \|\gamma^\dagger |\nabla u^\dagger| - \gamma |\nabla u|\|_{L^2(U)} + \delta^2 + G(\gamma, u).
\end{aligned} \tag{5.10}$$

Combining (5.8)(5.9)(5.10) gives

$$\begin{aligned}
R_n(\tilde{\gamma}_n, \tilde{u}_n) &\leq \|\gamma^\dagger |\nabla u^\dagger| - \gamma |\nabla u|\|_{L^2(U)}^2 + G(\gamma, u) + \left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i \tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i)\right] \right| + \Delta_n(\tilde{\gamma}_n, \tilde{u}_n).
\end{aligned} \tag{5.11}$$

Step 2. Now we show

$$\begin{aligned}
&\left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i \tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i)\right] \right| \\
&\leq \frac{1}{2} R_n(\tilde{\gamma}_n, \tilde{u}_n) + \frac{8\delta^2(1 + \log(N_\gamma N_u))}{n} + 2(\delta + B_\gamma B_u)(B_\gamma + B_u)\varepsilon.
\end{aligned} \tag{5.12}$$

In fact, for estimators $(\tilde{\gamma}_n, \tilde{u}_n) \in \mathcal{F}_\gamma \times \mathcal{F}_u$, there exists k' and ℓ' , such that $\|\gamma_{k'} - \tilde{\gamma}_n\|_{W^{1,\infty}(U)} \leq \varepsilon$ and $\|u_{\ell'} - \tilde{u}_n\|_{W^{2,\infty}(U)} \leq \varepsilon$. Then we have

$$\begin{aligned}
&\left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i \tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i)\right] \right| \\
&\leq \left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i (\tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i) - \gamma_{k'} |\nabla u_{\ell'}|(X_i))\right] \right| \\
&\quad + \left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))\right] \right| \\
&\leq \frac{2(B_\gamma + B_u)\delta}{n} \mathbb{E}_S\left[\sum_{i=1}^n |\xi_i|\right] + \left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))\right] \right| \\
&\leq 2\sqrt{\frac{2}{\pi}}(B_\gamma + B_u)\delta\varepsilon + \left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))\right] \right|.
\end{aligned} \tag{5.13}$$

Let $\eta_{k,\ell}$ be random variables defined as

$$\eta_{k,\ell} = \frac{\sum_{i=1}^n \xi_i (\gamma_k |\nabla u_\ell|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))}{(\sum_{i=1}^n (\gamma_k |\nabla u_\ell|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))^2)^{1/2}}, \tag{5.14}$$

then each $\eta_{k,\ell}$ follows sub-Gaussian distribution $\text{subG}(\delta^2)$ conditionally on $S = \{(X_i, Y_i)\}_{i=1}^n \cup \{\bar{X}_i\}_{i=1}^n$. By using Cauchy–Schwarz and AM–GM inequality, we find

$$\begin{aligned}
&\left| \mathbb{E}_S\left[\frac{2}{n} \sum_{i=1}^n \xi_i (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))\right] \right| \\
&\leq \frac{2}{n} \left| \mathbb{E}_S\left[\left(\sum_{i=1}^n (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))^2\right)^{1/2} \eta_{k',\ell'}\right] \right| \\
&\leq \frac{2}{n} \mathbb{E}_S\left[\sum_{i=1}^n (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))^2\right]^{1/2} \mathbb{E}_S\left[\eta_{k',\ell'}^2\right]^{1/2} \\
&\leq \frac{1}{2} \mathbb{E}_S\left[\frac{1}{n} \sum_{i=1}^n (\gamma_{k'} |\nabla u_{\ell'}|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))^2\right] + \frac{2}{n} \mathbb{E}_S\left[\eta_{k',\ell'}^2\right]
\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{2} \mathbb{E}_S \left[\frac{1}{n} \sum_{i=1}^n (\tilde{\gamma}_n |\nabla \tilde{u}_n|(X_i) - \gamma^\dagger |\nabla u^\dagger|(X_i))^2 \right] + 2B_\gamma B_u (B_\gamma + B_u) \delta + \frac{2}{n} \mathbb{E}_S \left[\max_{k,\ell} \eta_{k,\ell}^2 \right] \\ &\leq \frac{1}{2} R_n(\tilde{\gamma}_n, \tilde{u}_n) + \frac{8\epsilon^2(1 + \log N_\gamma N_u)}{n} + 2B_\gamma B_u (B_\gamma + B_u) \delta, \end{aligned} \quad (5.15)$$

where the last inequality is owing to $G_n(\tilde{\gamma}_n, \tilde{u}_n) \geq 0$ and [Lemma A.2](#). Combining [\(5.13\)](#) and [\(5.15\)](#), we obtain [\(5.12\)](#).

Step 3. Using [\(5.11\)](#) and [\(5.12\)](#), we have

$$\begin{aligned} R_n(\tilde{\gamma}_n, \tilde{u}_n) &\leq 2(\|\gamma |\nabla u| - \gamma^\dagger |\nabla u^\dagger|\|_{L^2(U)}^2 + G(\gamma, u)) + \frac{16\epsilon^2(1 + \log(N_\gamma N_u))}{n} \\ &\quad + 4\sqrt{\frac{2}{\pi}} (B_\gamma + B_u) \epsilon \delta + 4B_\gamma B_u (B_\gamma + B_u) \epsilon + \Delta_n(\tilde{\gamma}_n, \tilde{u}_n), \end{aligned} \quad (5.16)$$

which completes the proof. \square

Proof of Lemma 3.4. Combining [Lemmas A.1](#) and [A.3](#) yields the conclusion. \square

A.2. The proof of the approximation error bound

Lemma A.4 (Lipschitz Continuity of Forward Operator). Suppose $f \in H^{3/2}(\partial U)$. Given a function γ , we denote $u(\gamma)$ the weak solution of

$$\nabla \cdot (\gamma \nabla u) = 0, \text{ in } U, \quad Tu = f, \text{ on } \partial U.$$

Then the mapping $\gamma \mapsto u(\gamma)$ is Lipschitz continuous from $\mathcal{K} \cap C^1(\bar{U})$ to $H^2(U)$, i.e., for any $\gamma, \gamma + \delta\gamma \in \mathcal{K} \cap C^1(\bar{U})$, there holds

$$\|u(\gamma + \delta\gamma) - u(\gamma)\|_{H^2(U)} \leq C_{\text{Lip}} B_f \|\delta\gamma\|_{C^1(\bar{U})},$$

where C_{Lip} is a constant depending on γ_0, γ_1 and U .

Proof. Let $\delta u = u(\gamma + \delta\gamma) - u(\gamma)$. Then δu is the unique solution of

$$-\nabla \cdot (\gamma \nabla \delta u) = \nabla \cdot (\delta\gamma \nabla u(\gamma + \delta\gamma)), \text{ in } U, \quad T\delta u = 0, \text{ on } \partial U.$$

It is easy to verify that

$$\begin{aligned} \|\delta u\|_{H^2(U)} &\leq c_1 \|\nabla \cdot (\delta\gamma \nabla u(\gamma + \delta\gamma))\|_{L^2(U)} \\ &\leq c_1 \|u(\gamma + \delta\gamma)\|_{H^2(U)} \|\delta\gamma\|_{C^1(\bar{U})} \\ &\leq c_1 c_2 \|f\|_{H^{3/2}(\partial U)} \|\delta\gamma\|_{C^1(\bar{U})}, \end{aligned}$$

where c_1, c_2 are constants depending on γ_0, γ_1 and U . \square

Proof of Lemma 3.5. For each fixed $(\gamma, u) \in \mathcal{F}_\gamma \times \mathcal{F}_u$, recall the definition of the excess risk

$$R(\gamma, u) = \|\gamma |\nabla u| - \gamma^\dagger |\nabla u^\dagger|\|_{L^2(U)}^2 + G(\gamma, u). \quad (5.17)$$

We first consider the first term in Eq. [\(5.17\)](#). By triangular inequality and AM–GM inequality, we have

$$\begin{aligned} \|\gamma |\nabla u| - \gamma^\dagger |\nabla u^\dagger|\|_{L^2(U)}^2 &\leq 2\|\gamma |\nabla u| - \gamma^\dagger |\nabla u|\|_{L^2(U)}^2 + 2\|\gamma^\dagger |\nabla u| - \gamma^\dagger |\nabla u^\dagger|\|_{L^2(U)}^2 \\ &\leq 2B_u^2 \|\gamma - \gamma^\dagger\|_{L^2(U)}^2 + 2B_\gamma^2 \|u - u^\dagger\|_{H^1(U)}^2. \end{aligned} \quad (5.18)$$

We next investigate the second term in Eq. [\(5.17\)](#). Denote by u_γ the solution of

$$\nabla \cdot (\gamma \nabla u) = 0, \text{ in } U, \quad Tu = f, \text{ on } \partial U.$$

Then it follows that

$$\begin{aligned} G(\gamma, u) &= \|\nabla \cdot (\gamma \nabla(u - u_\gamma))\|_{L^2(U)}^2 + \|Tu - Tu_\gamma\|_{L^2(\partial U)}^2 \\ &\leq B_\gamma^2 \|u - u_\gamma\|_{H^2(U)}^2 + C_{\text{tr}}^2 \|u - u_\gamma\|_{H^1(U)}^2 \leq (B_\gamma^2 + C_{\text{tr}}^2) \|u - u_\gamma\|_{H^2(U)}^2, \end{aligned}$$

where the first inequality is due to the trace theorem, and the constant C_{tr} only depends on U . By [Lemma A.4](#), it holds

$$\|u_\gamma - u^\dagger\|_{H^2(U)} \leq C_{\text{Lip}} B_f \|\gamma - \gamma^\dagger\|_{C^1(\bar{U})},$$

and consequently,

$$G(\gamma, u) \leq 2(B_\gamma^2 + C_{\text{tr}}^2) C_{\text{Lip}}^2 B_f^2 \|\gamma - \gamma^\dagger\|_{C^1(\bar{U})}^2 + 2(B_\gamma^2 + C_{\text{tr}}^2) \|u - u^\dagger\|_{H^2(U)}^2. \quad (5.19)$$

Combining Eqs. [\(5.18\)](#) and [\(5.19\)](#) completes the proof. \square

Lemma A.5. Let $s \in \mathbb{N}_+$ and $\mu > 0$. Suppose φ is an exponential PU-admissible activation function.

(i) For each $\gamma^\dagger \in C^{s+1}(\bar{U})$, there exists a φ -network $\gamma \in \mathcal{N}_\varphi(\mathcal{D}_\gamma, S_\gamma, \mathcal{B}_\gamma)$ such that

$$\|\gamma - \gamma^\dagger\|_{C^1(\bar{U})} \leq \|\gamma^\dagger\|_{C^{s+1}(\bar{U})} \cdot S_\gamma^{-\frac{s-\mu}{d}}, \quad \text{with } \mathcal{D}_\gamma = C \log(d+s) \text{ and } \mathcal{B}_\gamma = CS_\gamma^{\frac{2s}{d}+7}.$$

(ii) For any $u^\dagger \in H^{s+2}(U)$, there exists a φ -network $u \in \mathcal{N}_\varphi(\mathcal{D}_u, S_u, \mathcal{B}_u)$ such that

$$\|u - u^\dagger\|_{H^2(U)} \leq \|u^\dagger\|_{H^{s+2}(U)} \cdot S_u^{-\frac{s+1-\mu}{d}}, \quad \text{with } \mathcal{D}_u = C \log(d+s+1) \text{ and } \mathcal{B}_u = CS_u^{\frac{2s+2}{d}+7}.$$

Here C is a constant depending on d , s , μ and U .

Proof. A direct conclusion of Proposition 4.8 in [18]. \square

Proof of Lemma 3.7. Combining Lemmas 3.5 and A.5 yields the desired result. \square

A.3. The proof of the statistical error bound

Lemma A.6. Let $m \in \mathbb{N}$ and $U \subseteq [0, 1]^d$ be a domain. For each neural network $\phi \in \mathcal{N}_\varphi(\mathcal{W}, S, \mathcal{B})$, we have $\phi \in C^m(\bar{U})$ and

$$\begin{aligned} \|\phi\|_{C^0(\bar{U})} &= \mathcal{O}(N_{\max} \mathcal{B}), \quad \|\phi\|_{C^1(\bar{U})} = \mathcal{O}(N_{\max}^D \mathcal{B}^D), \\ \|\phi\|_{C^2(\bar{U})} &= \mathcal{O}(N_{\max}^{D(2D+1)} \mathcal{B}^{D(D+1)}), \\ \|\phi\|_{C^3(\bar{U})} &= \mathcal{O}(N_{\max}^{D(2D^2+2D+1)} \mathcal{B}^{D(3D^2+3D+1)}), \end{aligned}$$

where $N_{\max} = \max\{N_\ell : \ell = 0, \dots, D\}$.

Proof. Recall the definition of the neural network ϕ :

$$\begin{aligned} \phi_0(x) &= x, \\ \phi_\ell(x) &= \varphi(T_D(\phi_{\ell-1}(x))), \quad \ell = 1, \dots, D-1, \\ \phi(x) &= T_D(\phi_{D-1}(x)). \end{aligned}$$

Let $B_{\varphi,i}$ ($i = 0, 1, 2, 3$) be the bounds of the activation function φ and its derivatives, i.e., $|\varphi(x)| \leq B_{\varphi,0}$, $|\varphi'(x)| \leq B_{\varphi,1}$, $|\varphi''(x)| \leq B_{\varphi,2}$ and $|\varphi'''(x)| \leq B_{\varphi,3}$, for each $x \in \bar{U}$. Then it holds

$$|\phi(x)| \leq N_D B B_{\varphi,0}.$$

Let $\phi_\ell(x) = (\phi_\ell^1(x), \dots, \phi_\ell^{N_\ell}(x))$, then we deduce

$$\partial_{x_q} \phi_\ell^m(x) = \sum_{i=1}^{N_\ell} A_\ell^{m,i} \varphi'(\phi_{\ell-1}^i(x)) \partial_{x_q} \phi_{\ell-1}^i(x), \quad \text{for } 1 \leq q \leq d,$$

and consequently,

$$|\partial_{x_q} \phi_\ell^m(x)| \leq N_\ell B B_{\varphi,1} \max_{1 \leq i \leq N_\ell} |\partial_{x_q} \phi_{\ell-1}^i(x)|, \quad \text{for } 1 \leq q \leq d.$$

Denote $B_{\ell,1} = \max_m \sup_{x \in \bar{U}} |\partial_{x_q} \phi_\ell^m(x)|$, we have

$$B_{\ell,1} \leq \left(\prod_{i=1}^{\ell} N_i \right) B^\ell B_{\varphi,1}^\ell \leq N_{\max}^\ell B^\ell B_{\varphi,1}^\ell \quad \text{and} \quad \sup_{x \in \bar{U}} |\partial_{x_p} \phi(x)| \leq N_{\max}^D B^D B_{\varphi,1}^D. \quad (5.20)$$

We next consider the bound of $\partial_{x_q} \partial_{x_p} \phi_\ell^m(x)$. It is easy to show that

$$\begin{aligned} \partial_{x_q} \partial_{x_p} \phi_\ell^m(x) &= \partial_{x_q} \left(\sum_{i=1}^{N_\ell} A_\ell^{m,i} \varphi'(\phi_{\ell-1}^i(x)) \partial_{x_p} \phi_{\ell-1}^i(x) \right) \\ &= \sum_{i=1}^{N_\ell} A_\ell^{m,i} \partial_{x_q} \left\{ \varphi'(\phi_{\ell-1}^i(x)) \right\} \partial_{x_p} \phi_{\ell-1}^i(x) + \sum_{i=1}^{N_\ell} A_\ell^{m,i} \varphi'(\phi_{\ell-1}^i(x)) \partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x) \\ &= \sum_{i=1}^{N_\ell} A_\ell^{m,i} \varphi''(\phi_{\ell-1}^i(x)) \partial_{x_q} \phi_{\ell-1}^i(x) \partial_{x_p} \phi_{\ell-1}^i(x) + \sum_{i=1}^{N_\ell} A_\ell^{m,i} \varphi'(\phi_{\ell-1}^i(x)) \partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x). \end{aligned}$$

Thus for $1 \leq p, q \leq d$, we have

$$|\partial_{x_q} \partial_{x_p} \phi_\ell^m(x)| \leq N_\ell B B_{\varphi,2} B_{\ell-1,1}^2 + N_\ell B B_{\varphi,1} \max_{1 \leq i \leq N_\ell} |\partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x)|.$$

Then using Eq. (5.20) gives

$$B_{\ell,2} \lesssim N_{\max}^{\ell} \left(\prod_{i=1}^{\ell} B_{i,1}^2 \right) B^{\ell} (B_{\ell,1} B_{\ell,2})^{\ell} \lesssim N_{\max}^{\ell(2\ell+1)} B^{\ell(\ell+1)} B_{\ell,1}^{\ell(\ell+1)} B_{\ell,2}^{\ell}, \quad (5.21)$$

and

$$|\partial_{x_q} \partial_{x_p} \phi(x)| \lesssim N_{\max}^{D(2D+1)} B^{D(D+1)} B_{\ell,1}^{D(D+1)} B_{\ell,2}^D.$$

Similarly, we now bound $\partial_{x_r} \partial_{x_q} \partial_{x_p} \phi_{\ell}^m(x)$. It follows that

$$\begin{aligned} & \partial_{x_r} \partial_{x_q} \partial_{x_p} \phi_{\ell}^m(x) \\ &= \sum_{i=1}^{N_{\ell}} A_{\ell}^{m,i} \partial_{x_r} \left\{ \rho''(\phi_{\ell-1}^i(x)) \partial_{x_q} \phi_{\ell-1}^i(x) \partial_{x_p} \phi_{\ell-1}^i(x) \right\} \\ &\quad + \sum_{i=1}^{N_{\ell}} A_{\ell}^{m,i} \partial_{x_r} \left\{ \rho'(\phi_{\ell-1}^i(x)) \partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x) \right\} \\ &= \sum_{i=1}^{N_{\ell}} A_{\ell}^{m,i} \rho'''(\phi_{\ell-1}^i(x)) \partial_{x_r} \phi_{\ell-1}^i(x) \partial_{x_q} \phi_{\ell-1}^i(x) \partial_{x_p} \phi_{\ell-1}^i(x) \\ &\quad + \sum_{i=1}^{N_{\ell}} A_{\ell}^{m,i} \rho''(\phi_{\ell-1}^i(x)) \left\{ \partial_{x_r} \partial_{x_q} \phi_{\ell-1}^i(x) \partial_{x_p} \phi_{\ell-1}^i(x) + \partial_{x_q} \phi_{\ell-1}^i(x) \partial_{x_r} \partial_{x_p} \phi_{\ell-1}^i(x) \right\} \\ &\quad + \sum_{i=1}^{N_{\ell}} A_{\ell}^{m,i} \rho''(\phi_{\ell-1}^i(x)) \partial_{x_r} \phi_{\ell-1}^i(x) \partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x) \\ &\quad + \sum_{i=1}^{N_{\ell}} A_{\ell}^{m,i} \rho'(\phi_{\ell-1}^i(x)) \partial_{x_r} \partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x). \end{aligned}$$

Hence

$$|\partial_{x_r} \partial_{x_q} \partial_{x_p} \phi_{\ell}^m(x)| \leq N_{\ell} B B_{\ell,3} B_{\ell-1,1}^3 + 3 N_{\ell} B B_{\ell,2} B_{\ell-1,1} B_{\ell-1,2} + N_{\ell} B B_{\ell,1} \max_{1 \leq i \leq N_{\ell}} |\partial_{x_r} \partial_{x_q} \partial_{x_p} \phi_{\ell-1}^i(x)|.$$

Then we have for $1 \leq p, q, r \leq d$

$$\begin{aligned} \sup_{\bar{U}} |\partial_{x_r} \partial_{x_q} \partial_{x_p} \phi(x)| &\lesssim N_{\max}^D \left(\prod_{\ell=1}^D B_{\ell,1}^3 + \prod_{\ell=1}^D B_{\ell,1} B_{\ell,2} \right) B^D (B_{\ell,1} B_{\ell,2} B_{\ell,3})^D \\ &\lesssim N_{\max}^{D(2D^2+2D+1)} B^{D(3D^2+3D+1)} B_{\ell,1}^{D(3D^2+3D+1)} B_{\ell,2}^{D(D+1)} B_{\ell,2}^D. \end{aligned}$$

This completes the proof. \square

Lemma A.7 (Extension). Let $m \in \mathbb{N}$ and $U \subset \subset (0, 1)^d$ be a domain with C^∞ -boundary. Then for each $\psi \in C^m(\bar{U})$, there exists a bounded and compact support extension $E\psi \in C_0^m([0, 1]^d)$, such that

- (i) $E\psi(x) = \psi(x)$ in \bar{U} ,
- (ii) $\|E\psi\|_{C^m([0, 1]^d)} \leq C \|\psi\|_{C^m(\bar{U})}$,

where C is a constant depending only on U .

Proof. For each function $\psi \in C^m(\bar{U})$, by Theorem 5.24 [27], there exists $E_1\psi \in C^m([0, 1]^d)$. Let $\zeta \in C_0^m([0, 1]^d)$ be a cut-off function such that

- (i) $0 \leq \zeta(x) \leq 1$ for each $x \in [0, 1]^d$,
- (ii) $\zeta(x) = 1$ for each $x \in \bar{U}$, and
- (iii) $|\partial^\alpha \zeta| \leq C / (\text{dist}(U, \partial(0, 1)^d))^{|a|}$.

Then $E\psi := \zeta E_1\psi \in C_0^m([0, 1]^d)$ is the desired function of ψ . The proof is completed. \square

Lemma A.8. Let $U \subset \subset (0, 1)^d$ be a domain with C^∞ -boundary. Let $k, s \in \mathbb{N}$ with $k < s$, and C_B^s be the norm-ball of radius B in $C^s(\bar{U})$. Then

$$H(\varepsilon, C_B^s, C^k(\bar{U})) \leq C \cdot \left(\frac{B}{\varepsilon} \right)^{\frac{d}{s-k}},$$

where C is a constant depending on s, k and U .

Proof. Notice that

$$\left\{ \partial^\alpha f : f \in C_B^s, |\alpha| = k \right\} \subseteq C_B^{s-k}.$$

Let $C_{B,\varepsilon}^{s-k}$ be a $\|\cdot\|_{C(\bar{U})}$ δ -cover of C_B^{s-k} with $|C_{B,\varepsilon}^{s-k}| = N(\delta, C_B^{s-k}, C(\bar{U}))$. Then for each $f \in C_B^s$ and $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = k$, there exist $\pi^\alpha(f) \in C_{B,\varepsilon}^{s-k}$ such that

$$\|\pi^\alpha(f) - \partial^\alpha f\|_{C(\bar{U})} \leq \delta. \quad (5.22)$$

Without loss of generality, we assume $\alpha_j \geq 1$, and define

$$F(x_1, \dots, x_d) = \int_{x_{j,0}}^{x_j} \pi^\alpha(f)(x_1, \dots, x_{j-1}, z, x_{j+1}, \dots, x_d) dz,$$

where $x_{j,0} = \min\{x_j : (x_1, \dots, x_j, \dots, x_d) \in \bar{U}\}$. It is clear that $\partial_j F = \pi^\alpha(f)$. Denote $\tilde{\alpha} = (\alpha_1, \dots, \alpha_j - 1, \dots, \alpha_d)$. By Lemma A.7, there exist an extension $E(F - \partial^{\tilde{\alpha}} f) \in C_0^1([0, 1]^d)$, such that $E(F - \partial^{\tilde{\alpha}} f)(x) = (F - \partial^\alpha f)(x)$ for each $x \in \bar{U}$, and consequently

$$\|F - \partial^{\tilde{\alpha}} f\|_{C(\bar{U})} \leq \|E(F - \partial^{\tilde{\alpha}} f)\|_{C([0,1]^d)}. \quad (5.23)$$

Using Poincaré's inequality, we have

$$\|E(F - \partial^{\tilde{\alpha}} f)\|_{C([0,1]^d)} \leq C \|\partial_j E(F - \partial^{\tilde{\alpha}} f)\|_{C([0,1]^d)} = C \|\tilde{E}(\pi^\alpha(f) - \partial^\alpha f)\|_{C([0,1]^d)}, \quad (5.24)$$

where \tilde{E} is an extension operator, satisfying $\tilde{E}(\pi^\alpha(f) - \partial^\alpha f) \in C_0([0, 1]^d)$ and

$$\|\tilde{E}(\pi^\alpha(f) - \partial^\alpha f)\|_{C([0,1]^d)} \leq C \|\pi^\alpha(f) - \partial^\alpha f\|_{C(\bar{U})}, \quad (5.25)$$

where C is a constant depending on U . Here the existence of \tilde{E} can be guaranteed by Lemma A.7. Combining Eqs. (5.22)–(5.25) yields

$$\max \left\{ \|F - \partial^{\tilde{\alpha}} f\|_{C(\bar{U})}, \|\partial_j(F - \partial^{\tilde{\alpha}} f)\|_{C(\bar{U})} \right\} \leq C \|\pi^\alpha(f) - \partial^\alpha f\|_{C(\bar{U})} \leq C\delta,$$

where C is a constant depending on U .

Repeating the same procedure as above, we can construct a function $\pi(f) \in C^k(\bar{U})$, such that

$$\max_{0 \leq |\alpha| \leq k} \|\partial^\alpha(\pi(f) - f)\|_{C(\bar{U})} \leq C\delta,$$

which means $C_{B,\varepsilon}^s := \{\pi(f) : f \in C_B^s\}$ is a $\|\cdot\|_{C(\bar{U})}$ $C\delta$ -cover of C_B^s satisfying $|C_{B,\varepsilon}^s| = |C_{B,\varepsilon}^{s-k}|$. Then

$$N(C\delta, C_B^s, C^k(\bar{U})) \leq N(\delta, C_{B,\varepsilon}^{s-k}, C(\bar{U})).$$

Setting $\varepsilon = C\delta$ and by Theorem 4.3.36 in [13], we complete the proof. \square

Proof of Lemma 3.8. It is sufficient to estimate the metric entropy of the following function classes:

- (i) $\mathcal{F}_\gamma = \mathcal{N}_\rho(D_\gamma, S, \mathcal{B}_\gamma)$ with $D_\gamma = C \log(d + s)$ and $\mathcal{B}_\gamma = CS^{\frac{2s}{d}+7}$, and
- (ii) $\mathcal{F}_u = \mathcal{N}_\rho(D_u, S, \mathcal{B}_u)$ with $D_u = C \log(d + s + 1)$ and $\mathcal{B}_u = CS^{\frac{2s+2}{d}+7}$.

By Lemma A.6, we have

$$\mathcal{F}_\gamma \subseteq C_{B_\gamma}^2, \quad \text{with } \mathcal{B}_\gamma = \mathcal{O}(N_{\gamma,\max}^{D_\gamma(2D_\gamma+1)} \mathcal{B}_\gamma^{D_\gamma(D_\gamma+1)}), \quad (5.26)$$

and

$$\mathcal{F}_u \subseteq C_{B_u}^3, \quad \text{with } \mathcal{B}_u = \mathcal{O}(N_{u,\max}^{D_u(2D_u^2+2D_u+1)} \mathcal{B}_u^{D_u(3D_u^2+3D_u+1)}). \quad (5.27)$$

Consequently, we have

$$H(\varepsilon, \mathcal{F}_\gamma, \|\cdot\|_{W^{1,\infty}(U)}) \leq H(\varepsilon, C_{B_\gamma}^2, \|\cdot\|_{W^{1,\infty}(U)}), \quad (5.28)$$

and

$$H(\varepsilon, \mathcal{F}_u, \|\cdot\|_{W^{2,\infty}(U)}) \leq H(\varepsilon, C_{B_u}^3, \|\cdot\|_{W^{2,\infty}(U)}). \quad (5.29)$$

Applying Lemma A.8 and (5.26) and (5.27) yields

$$\begin{aligned} H(\varepsilon, \mathcal{F}_\gamma, \|\cdot\|_{W^{1,\infty}(U)}) &\leq C \cdot \left(\frac{N_{\gamma,\max}^{D_\gamma(2D_\gamma+1)} \mathcal{B}_\gamma^{D_\gamma(D_\gamma+1)}}{\varepsilon} \right)^d, \quad \text{and} \\ H(\varepsilon, \mathcal{F}_u, \|\cdot\|_{W^{2,\infty}(U)}) &\leq C \cdot \left(\frac{N_{u,\max}^{D_u(2D_u^2+2D_u+1)} \mathcal{B}_u^{D_u(3D_u^2+3D_u+1)}}{\varepsilon} \right)^d. \end{aligned}$$

Substituting $N_{\gamma,\max} \leq S$ and $N_{u,\max} \leq S$ gives

$$\max\{H_\gamma^\varepsilon, H_u^\varepsilon\} \lesssim \mathcal{O}(S^{6(4d+s+1)\log^3(d+s+1)}\varepsilon^{-d}). \quad (5.30)$$

Setting

$$\varepsilon = \left(\frac{B_\gamma^2 B_u^2 + B_f^2 + \delta^2}{(\delta + B_\gamma B_u)(B_\gamma + B_u)} \frac{S^{6(4d+s+1)\log^3(d+s+1)}}{n} \right)^{\frac{1}{d+1}}$$

and letting $m = s + 1$ yield the desired result. \square

References

- [1] Adrian Nachman, Alexandru Tamasan, Alexandre Timonov, Recovering the conductivity from a single measurement of interior data, *Inverse Problems* 25 (3) (2009) 035014, <http://dx.doi.org/10.1088/0266-5611/25/3/035014>.
- [2] Nicholas Hoell, Amir Moradifam, Adrian Nachman, Current density impedance imaging of an anisotropic conductivity in a known conformal class, *SIAM J. Math. Anal.* 46 (3) (2014) 1820–1842.
- [3] Carlos Montalto, Plamen Stefanov, Stability of coupled-physics inverse problems with one internal measurement, *Inverse Problems* 29 (12) (2013) 125004.
- [4] Robert Lopez, Amir Moradifam, Stability of current density impedance imaging, *SIAM J. Math. Anal.* 52 (5) (2020) 4506–4523, <http://dx.doi.org/10.1137/19M126520X>.
- [5] Adrian Nachman, Alexandru Tamasan, Alexandre Timonov, Conductivity imaging with a single measurement of boundary and interior data, *Inverse Problems* 23 (6) (2007) 2551, <http://dx.doi.org/10.1088/0266-5611/23/6/017>.
- [6] Alexandru Tamasan, Alexandre Timonov, Johann Veras, Stable reconstruction of regular 1-harmonic maps with a given trace at the boundary, *Appl. Anal.* 94 (6) (2015) 1098–1115.
- [7] Hassan Yazdanian, Kim Knudsen, Numerical conductivity reconstruction from partial interior current density information in three dimensions, *Inverse Problems* 37 (10) (2021) 105010.
- [8] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* (ISSN: 0021-9991) 378 (2019) 686–707, <http://dx.doi.org/10.1016/j.jcp.2018.10.045>, URL <https://www.sciencedirect.com/science/article/pii/S0021999118307125>.
- [9] Bao Gang, Xiaojing Ye, Yaohua Zhang, Haomin Zhou, Numerical solution of inverse problems by weak adversarial networks, *Inverse Problems* 36 (11) (2020) 115003, <http://dx.doi.org/10.1088/1361-6420/abb447>.
- [10] Bangti Jin, Xiyao Li, Xiliang Lu, Imaging conductivity from current density magnitude using neural networks, *Inverse Problems* 38 (7) (2022) 075003, <http://dx.doi.org/10.1088/1361-6420/ac6d03>.
- [11] Mengmeng Zhang, Qianxiao Li, Jijun Liu, On stability and regularization for data-driven solution of parabolic inverse source problems, *J. Comput. Phys.* (ISSN: 0021-9991) 474 (2023) 111769, <http://dx.doi.org/10.1016/j.jcp.2022.111769>, URL <https://www.sciencedirect.com/science/article/pii/S0021999122008324>.
- [12] Siddhartha Mishra, Roberto Molinaro, Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for PDEs, *IMA J. Numer. Anal.* (ISSN: 0272-4979) 42 (2) (2021) 981–1022, <http://dx.doi.org/10.1093/imanum/drab032>, arXiv:<https://academic.oup.com/imanum/article-pdf/42/2/981/43373937/drab032.pdf>.
- [13] Evarist Giné, Richard Nickl, Mathematical foundations of infinite-dimensional statistical models, in: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2015, <http://dx.doi.org/10.1017/CBO9781107337862>.
- [14] Thomas Widlak, Otmar Scherzer, Hybrid tomography for conductivity imaging, *Inverse Problems* 28 (8) (2012) 084008, <http://dx.doi.org/10.1088/0266-5611/28/8/084008>.
- [15] Chenguang Duan, Yuling Jiao, Yanming Lai, Dingwei Li, Xiliang Lu, Jerry Zhijian Yang, Convergence rate analysis for deep ritz method, *Commun. Comput. Phys.* (ISSN: 1991-7120) 31 (4) (2022) 1020–1048, <http://dx.doi.org/10.4208/cicp.OA-2021-0195>, URL http://global-sci.org/intro/article_detail/cicp/20375.html.
- [16] Yuling Jiao, Yanming Lai, Dingwei Li, Xiliang Lu, Fengru Wang, Yang Wang, Jerry Zhijian Yang, A rate of convergence of physics informed neural networks for the linear second order elliptic PDEs, *Commun. Comput. Phys.* (ISSN: 1991-7120) 31 (4) (2022) 1272–1295, <http://dx.doi.org/10.4208/cicp.OA-2021-0186>, URL http://global-sci.org/intro/article_detail/cicp/20384.html.
- [17] Yuling Jiao, Yanming Lai, Yisu Lo, Yang Wang, Yunfei Yang, Error Analysis of Deep Ritz Methods for Elliptic Equations, 2021, <http://dx.doi.org/10.48550/ARXIV.2107.14478>, URL <https://arxiv.org/abs/2107.14478>.
- [18] Ingo Gühring, Mones Raslan, Approximation rates for neural networks with encodable weights in smoothness spaces, *Neural Netw.* (ISSN: 0893-6080) 134 (2021) 107–130, <http://dx.doi.org/10.1016/j.neunet.2020.11.010>, URL <https://www.sciencedirect.com/science/article/pii/S0893608020303956>.
- [19] Martin Anthony, Peter L. Bartlett, *Neural network learning: Theoretical foundations*, vol. 9, cambridge university press Cambridge, 1999.
- [20] Yiping Lu, Haoxuan Chen, Jianfeng Lu, Lexing Ying, Jose Blanchet, Machine learning for elliptic PDEs: Fast rate generalization bound, neural scaling law and minimax optimality, in: International Conference on Learning Representations, 2022, URL <https://openreview.net/forum?id=mhYUBYNoGz>.
- [21] Yulong Lu, Jianfeng Lu, Min Wang, A Priori Generalization Analysis of the Deep Ritz Method for Solving High Dimensional Elliptic Partial Differential Equations, in: Mikhail Belkin, Samory Kpotufe (Eds.), Proceedings of Thirty Fourth Conference on Learning Theory, in: Proceedings of Machine Learning Research, vol. 134, PMLR, 2021, pp. 3196–3241, URL <https://proceedings.mlr.press/v134/lu21a.html>.
- [22] Xavier Glorot, Yoshua Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Yee Whye Teh, Mike Titterington (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, in: Proceedings of Machine Learning Research, vol. 9, PMLR, Chia Laguna Resort, Sardinia, Italy, 2010, pp. 249–256, URL <https://proceedings.mlr.press/v9/glorot10a.html>.
- [23] D. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Yee Whye Teh, Mike Titterington (Eds.), *Proceedings of the 3rd International Conference for Learning Representations*, Vol. 500, 2015.
- [24] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, Soumith Chintala, PyTorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035, URL <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [25] John Zhi-Qin Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, Zheng Ma, Frequency principle: Fourier analysis sheds light on deep neural networks, *Commun. Comput. Phys.* (ISSN: 1991-7120) 28 (5) (2020) 1746–1767, <http://dx.doi.org/10.4208/cicp.OA-2020-0085>, URL http://global-sci.org/intro/article_detail/cicp/18395.html.
- [26] Johannes Schmidt-Hieber, Nonparametric regression using deep neural networks with ReLU activation function, *Ann. Statist.* 48 (4) (2020) 1875–1897, <http://dx.doi.org/10.1214/19-AOS1875>.
- [27] Robert A. Adams, John J.F. Fournier, Sobolev Spaces, Elsevier, 2003.