

## Recovering the Source Term in Elliptic Equation via Deep Learning: Method and Convergence Analysis

Chenguang Duan<sup>1,2</sup>, Yuling Jiao<sup>1,2</sup>, Jerry Zhijian Yang<sup>1,2,\*</sup>  
and Pingwen Zhang<sup>1,2,3</sup>

<sup>1</sup>*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China.*

<sup>2</sup>*Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072, China.*

<sup>3</sup>*School of Mathematical Sciences, Peking University, Beijing 100871, China.*

Received 26 October 2023; Accepted (in revised version) 29 March 2024.

---

**Abstract.** In this paper, we present a deep learning approach to tackle elliptic inverse source problems. Our method combines Tikhonov regularization with physics-informed neural networks, utilizing separate neural networks to approximate the source term and solution. Firstly, we construct a population loss and derive stability estimates. Furthermore, we conduct a convergence analysis of the empirical risk minimization estimator. This analysis yields a prior rule for selecting regularization parameters, determining the number of observations, and choosing the size of neural networks. Finally, we validate our proposed method through numerical experiments. These experiments also demonstrate the remarkable robustness of our approach against data noise, even at high levels of up to 50%.

**AMS subject classifications:** 65N15, 65N20, 65N21

**Key words:** Inverse source problem, deep neural network, stability estimate, convergence rate.

---

### 1. Introduction

Inverse source problems have attracted considerable interest in various scientific and engineering domains, which arise in practical applications over natural phenomena such as pollution source identification [4, 5, 10, 45], dislocation problems [6] and inverse problems of gravimetry [29]. Additionally, they have found extensive use in a range of biomedical imaging techniques, including photo-acoustic and thermo-acoustic tomography, optical tomography [2], electroencephalography (EEG) [38], magnetoencephalography (MEG) [26], and bioluminescence tomography (BLT) [60]. Of particular relevance to this paper is the modeling of the seawater intrusion phenomenon [7, 17], where the source term represents the pumping wells of freshwater within the context of seawater intrusion.

---

\*Corresponding author. Email address: zjyang.math@whu.edu.cn (J.-Z. Yang)

In this work, we aim at identifying the unknown source density in elliptic problems from interior measurements. Let  $\Omega \subseteq \mathbb{R}^d$  ( $d \geq 1$ ) be a simply connected bounded domain with sufficiently smooth boundary  $\partial\Omega$ . Consider the following second order elliptic equation with Neumann boundary condition:

$$\begin{aligned} -\Delta u + V(x)u &= f(x) \quad \text{in } \Omega, \\ \partial_n u &= g(x) \quad \text{on } \partial\Omega, \end{aligned} \tag{1.1}$$

where the potential function  $V$  and the boundary flux  $g$  are given. Further, the potential function  $V$  has a positive lower bound, that is,  $V(x) \geq V_0 > 0$  for each  $x \in \Omega$ . Let  $f^\dagger$  be the ground truth space-dependent source density and  $u^\dagger$  be the solution of (1.1) corresponding to the source density  $f^\dagger$ . The elliptic inverse source problems aim to recover the unknown source density from finite number of random samples generated from the following noisy model:

$$y^\delta(x) = u^\dagger(x) + \xi(x), \quad z^\delta(x) = \nabla u^\dagger(x) + \zeta(x), \quad x \sim U(\Omega), \tag{1.2}$$

where  $\xi(x)$  and  $\zeta_j(x)$ ,  $j \in [d]$  are noise terms, and  $U(\Omega)$  represents the uniform distribution on  $\Omega$ . Further, we assume that

$$\|\xi\|_{L^\infty(\Omega)} \leq \delta, \quad \|\zeta_j\|_{L^\infty(\Omega)} \leq \delta, \quad j \in [d], \tag{1.3}$$

where  $\delta$  is known as the noise level in the context of inverse problems. Notice that  $H^1(\Omega)$ -norm measurement in (1.2) is stronger than the usual  $L^2(\Omega)$ -norm measurement, the technical motivation for which is the necessity to establish stability estimates (see Theorem 2.1) for reconstructions. However,  $H^1(\Omega)$ -norm measurement also makes sense, which has been used in [36, 44]. For example, in the context of inverse problems of gravimetry [29], the gravitational force  $\nabla u^\dagger$  can be measured directly, and the measurement of gravitational field  $u^\dagger$  can be perceived by the noisy measurement of gravitational force. Besides, if only the  $L^2(\Omega)$ -norm measurement is available, the measurement of gradient can be obtained by some numerical differentiation methods after pre-smoothing the raw noisy data of  $u^\dagger$ . In addition, the provable  $H^1(\Omega)$ -norm estimation can also be obtained from noisy  $L^2(\Omega)$ -measurements via the finite element method [28] or deep Sobolev regression [18].

There have been extensive study devoted to the uniqueness and stability of inverse source problems [7, 8]. The uniqueness can be obtained by means of Holmgren's theorem and the regularity of the forward problem, as it was done in [8]. Further, the Lipschitz stability estimates for inverse source problems are proposed in [7].

Due to the ill-posed nature of inverse source problems [7, 21, 27, 56], constructing accurate and stable numerical approximations can be challenging. Several reconstruction methods have been developed to address this issue [1, 28, 39, 40, 45, 58, 61]. One popular approach involves reformulating the inverse source problem as an output least-squares PDE-constrained optimization problem, complemented with Tikhonov regularization [14, 30]. By formulating it as an optimization problem, classical optimization algorithms can then be employed for solution. In practical computation, one still needs to discretize Tikhonov functional and the PDE constraint, which is often achieved by the Galerkin finite element

method (FEM). However, the solution of PDE-constrained optimization problems necessitates the development of complex optimization algorithms tailored to each case. Furthermore, Galerkin approximation is mesh-dependent, resulting in exponential growth of computational cost with dimensionality, commonly referred to as the curse of dimensionality (CoD). Therefore, a simple yet effective approach is to convert the PDE-constrained optimization problem into an unconstrained problem by introducing an augmented objective functional that incorporates a penalty inspired by physics-informed neural networks (PINNs) [55]. By employing neural networks to approximate the unknown source density and the solution, and discretizing the objective functional using Monte Carlo methods, a mesh-free approach can be achieved, which is easy to implement and has the potential to alleviate the curse of dimensionality.

In this study, we present a novel approach based on neural networks for identifying the source density in elliptic equations using interior measurements. We also conduct a comprehensive analysis to determine the convergence rates of the reconstructions with respect to the noise level. To begin with, we establish stability estimates that serve as the basis for determining an appropriate regularization functional. Subsequently, we employ the Monte Carlo method to discretize the population risk. By minimizing the empirical risk within pre-specified neural network classes, we obtain reconstructions of the source density and solution. To analyze the convergence rates of the reconstructions, we demonstrate that the population risk is influenced by several factors. These include the approximation error, the generalization error, the error arising from data noise, and a regularization term controlled by the regularization parameter. Notably, we observe that while the approximation error decreases as the size of the neural network increases, the generalization error exhibits an opposite trend. Consequently, by striking a balance between approximation power and generalization ability, we can select neural network classes that offer optimal performance. By combining the analysis of population risk error with the aforementioned stability estimates, we derive the convergence rates of the reconstructions in relation to the noise level. Furthermore, our study provides valuable a priori guidance for selecting appropriate regularization parameters, determining the number of samples, and choosing the size of the neural networks. Finally, through a series of numerical experiments, we demonstrate the remarkable stability of our method against data noise.

In recent years, there has been an increasing interest in neural network-based methods for solving ill-posed problems [9, 19, 35, 36, 47, 55, 62, 63]. These approaches typically formulate a loss function tailored to the specific problem and use neural networks to approximate unknown functions. The population risk is then discretized using the Monte Carlo method. By adopting a mesh-free strategy, methods in this line offer a promising approach for addressing high-dimensional problems. One particularly notable and widely recognized framework within these works is physics-informed neural networks (PINNs) [55]. The original PINNs scheme has been applied to recover constant coefficients in PDEs [55] and to estimate derivative orders in fractional PDEs [25, 53]. Additionally, [50] utilized PINNs to solve Cauchy problems and data assimilation. For reconstructing non-constant coefficients in PDEs, a natural approach is to approximate both the solution and the unknown coefficient function using two neural networks. These networks are then coupled through a loss func-

tion that combines a least-squares data-fitting loss with a physics-informed loss. The idea can be traced back to its initial mention in [9], where coupled PINNs were compared with inverse weak adversarial networks (IWANs) [9] in numerical experiments. Subsequently, this methodology has been extended and applied to a variety of inverse problems, including holography inverse-design problems [47], parabolic inverse source problems [64], and current density impedance imaging [19]. The most relevant works are [62, 63], where the source term is recovered from measurements of the solution in partial domains, using the modified deep Galerkin method. In addition, [37] employed PINNs to solve elliptic distributed optimal control problems, which closely relates to the inverse source problem addressed in this study.

While PINNs-based approaches have shown promising empirical performance in previous studies, most of them lack rigorous theoretical guarantees. A comprehensive theoretical understanding of a reconstruction method should address two important questions:

1. What are the convergence rates of the population risk of reconstructions?
2. How does the error of reconstructions converge as the population risk decreases?

Answering these questions would provide convergence rates for the reconstructions. The first question pertains to the statistical learning theory, while the second question relates to stability estimates for inverse problems. Unfortunately, to the best of our knowledge, none of the existing theoretical analyses of PINNs-based reconstruction approaches provide a complete answer to these two questions. In a recent work by [19], a convergence rate of the population risk is derived by considering the trade-off between the approximation error and the generalization error. However, due to the lack of stability estimates, this analysis does not specifically provide a convergence rate for the reconstruction itself. Other works such as [50, 63] primarily focus on the generalization error in their analysis, overlooking the approximation power of neural networks. As a result, these studies cannot fully address the first question and do not provide guidance on how to select the size of neural networks. To answer the second question, [50, 63] utilize conditional stability estimates for elliptic inverse source problems, assuming the unknown source density to be analytic. Conversely, the stability estimates established in our work require weaker regularity conditions for the ground truth source density.

### 1.1. Contributions

The contributions of this paper are summarized as follows:

1. Within the framework of Tikhonov regularization, we introduce a loss function for elliptic inverse source problems using physic-informed neural networks. Subsequently, we establish stability estimates that provide a solid theoretical foundation for the validity of our method.
2. We establish a rigorous convergence analysis for reconstructions  $\|u - u^\dagger\|_{H^1(\Omega)} = \mathcal{O}(\delta)$  and  $\|f - f^\dagger\|_{L^2(\Omega)} = \mathcal{O}(\delta^{1/2})$ , which also provide a priori guides for the selection of

hyper-parameters such as the regularization parameter, the number of samples and the size of neural networks.

3. Our reconstruction method is easy to implement and performs well on recovering both smooth and non-smooth source densities. It also shows remarkable robustness against data noise in numerical experiments, remaining highly accurate in presence of up to 50% noise.

## 1.2. Preliminaries and notations

**Definition 1.1** (Fully-connected neural networks). A *fully-connected neural network*  $\psi : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_{L+1}}$  is a function defined by

$$\psi(x) = T_L(\varrho(T_{L-1}(\cdots \varrho(T_0(x)) \cdots))),$$

where the activation function  $\varrho$  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  $\ell = 0, \dots, L$ .

Throughout this paper, we consider the case  $N_0 = d$  and  $N_{L+1} = 1$ . The parameter of the neural network  $\psi$  is defined as the collection of all its weights  $\theta = \{(A_\ell, b_\ell)\}_{\ell=0}^L$ . The positive integer  $L$  is the depth of the neural network and  $S := \sum_{\ell=1}^L (\|A_\ell\|_0 + \|b_\ell\|_0)$  is the total number of nonzero weights. Moreover, we denote by  $R$  the bound of nonzero weights in absolute value, that is,  $R = \max_{0 \leq \ell \leq L} \max\{\max_{i,j} |A_{\ell,ij}|, \max_i |b_{\ell,i}|\}$ . Finally, we define the function class  $\mathcal{N}_\varrho(L, S, R)$  as the collection of  $\varrho$ -neural networks with at most  $L$  layers, at most  $S$  nonzero weights and each weight are bounded by  $R$ . The approximation property of tanh-neural networks have been established as follows.

**Lemma 1.1** (cf. Gühring & Raslan [24, Theorem 3.8]). Let  $d, s, k \in \mathbb{N}_+$  such that  $0 \leq k < s$ . Let  $\varrho = \tanh$  and  $\mathcal{H}_{s,d} = \{g : \|g\|_{H^s((0,1)^d)} \leq 1\}$ . Then it holds that

$$\sup_{g \in \mathcal{H}_{s,d}} \inf_{\psi \in \mathcal{N}_\varrho(L,S,R)} \|\psi - g\|_{H^k((0,1)^d)} \leq \varepsilon,$$

provided that  $L = c \log(d+k)$ ,  $S = C \varepsilon^{-d/(s-k-\mu)}$  and  $R = C \varepsilon^{-(9d+2k+4\mu)/(2(s-k-\mu))-2}$ , where  $c, \mu > 0$  are absolute constants and the constant  $C$  depends on  $s, k, d$  and  $\mu$ .

To measure the complexity of the function class, we next introduce Rademacher complexity [12], which plays an important role in generalization analysis [12, 51].

**Definition 1.2** (Rademacher Complexity). Let  $\Omega \subseteq \mathbb{R}^d$  ( $d \geq 1$ ) be a bounded domain and let  $\mu$  be a measure on  $\mathcal{X}$ . Suppose that  $\mathcal{G}$  is a family of functions mapping from  $\mathcal{X}$  to  $\mathbb{R}$  and  $\{x_i\}_{i=1}^m$  is a set of samples i.i.d. drawn from  $\mu$ . Let  $\sigma = \{\sigma_i\}_{i=1}^m$  be a set of i.i.d. Rademacher variables and independent of  $\{x_i\}_{i=1}^m$ . Then the empirical Rademacher complexity of  $\mathcal{G}$  with respect to the sample set  $\{x_i\}_{i=1}^m$  is defined as

$$\widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G}) = \mathbb{E}_\sigma \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^m \sigma_i g(x_i) \right],$$

where  $\mathbb{E}_\sigma[\cdot] = \mathbb{E}[\cdot | \{x_i\}_{i=1}^m]$ . Further, the Rademacher complexity of  $\mathcal{G}$  is the expectation of the empirical Rademacher complexity over all samples of size  $m$  drawn according to  $\mu^m$ , that is,

$$\mathfrak{R}_m(\mathcal{G}) = \mathbb{E}_{\{x_i\}_{i=1}^m} [\widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G})].$$

The rest of the paper is organized as follows. Section 2 introduces the loss function for source recovery in elliptic equations and discusses stability estimates. In Section 3, we present a data-dependent oracle inequality (Section 3.1) that characterizes the population risk of reconstructions. Moreover, convergence rates for reconstructions are derived in Section 3.2. To complement our theoretical analysis, a series of numerical experiments is presented in Section 4, demonstrating the exceptional stability of our method in the presence of data noise. The proofs of the theoretical results are provided in Section 5. The conclusions and discussions are summarized in Section 6.

## 2. Reconstruction Methods

We will study in this section the stability estimates, based on which we propose a neural network-based method for the recovery of unknown source density in the elliptic system (1.1). Throughout this section, we assume that  $V \in L^\infty(\Omega)$ ,  $f \in L^2(\Omega)$  and  $g \in H^{1/2}(\partial\Omega)$  in (1.1).

In view of the ill-posedness of elliptic inverse source problems, researchers usually employ Tikhonov regularization to transform (1.1) and (1.2) into an optimal control problem

$$\begin{aligned} \min_{(u,f)} J_\lambda(u,f) &= \|u - y^\delta\|_{L^2(\Omega)}^2 + \|\nabla u - z^\delta\|_{L^2(\Omega)}^2 + \lambda \mathcal{R}(f) \\ \text{subject to} \quad &\begin{cases} -\Delta u + V(x)u = f(x) & \text{in } \Omega, \\ \partial_n u = g(x) & \text{on } \partial\Omega, \end{cases} \end{aligned} \tag{2.1}$$

where  $\lambda > 0$  is the regularization parameter and  $\mathcal{R}$  is the pre-specified regularization functional. Inspired by physics-informed neural networks (PINNs), instead of putting a hard-constraint in the optimization, we treat the PDE constraints in (2.1) as a soft-constraint, which leads to the following unconstrained optimization problem:

$$(u_\lambda^\delta, f_\lambda^\delta) \in \arg \min_{(u,f)} L_\lambda(u,f) = J_\lambda(u,f) + R_{\text{int}}(u,f) + R_{\text{bdy}}(u), \tag{2.2}$$

where

$$R_{\text{int}}(u,f) = \|f + \Delta u - Vu\|_{L^2(\Omega)}^2, \quad R_{\text{bdy}}(u) = \|\partial_n u - g\|_{L^2(\partial\Omega)}^2.$$

In addition, the functional  $L_\lambda(\cdot, \cdot)$  is called the population risk. The following theorem demonstrates that reconstructions converge to the ground truth as excess risks decrease, when we choose appropriate regularization functional.

**Theorem 2.1** (Stability Estimates for Elliptic Equations). *Under the noise model (1.2). Let  $\mathcal{R}$  be the squared  $H^1(\Omega)$ -norm — i.e.  $\mathcal{R}(f) = \|f\|_{H^1(\Omega)}^2$ . Suppose that  $(u^\dagger, f^\dagger) \in H^2(\Omega) \times H^1(\Omega)$*

satisfies (1.1), and let the population risk  $L_\lambda(\cdot, \cdot)$  defined as (2.2). Then for each  $(u, f) \in H^2(\Omega) \times H^1(\Omega)$  the following inequalities hold:

$$\begin{aligned}\|u - u^\dagger\|_{H^1(\Omega)} &\leq L_\lambda^{1/2}(u, f)C\delta, \\ \|f - f^\dagger\|_{L^2(\Omega)} &\leq C(1 + \lambda^{-1/4}L_\lambda^{1/4}(u, f))(L_\lambda^{1/4}(u, f) + \delta^{1/2}),\end{aligned}$$

where  $C$  is a positive constant only depending on  $\Omega$ ,  $\|V\|_{L^\infty(\Omega)}$  and  $\|f^\dagger\|_{H^1(\Omega)}$ .

**Remark 2.1** (Convergence Rates in Population Level). With the aid of Theorem 2.1, we can establish convergence rates of  $(u_\lambda^\delta, f_\lambda^\delta)$  defined as (2.2). In fact, it is apparent that  $L_\lambda(u_\lambda^\delta, f_\lambda^\delta) \leq \mathcal{O}(\delta^2) + \mathcal{O}(\lambda)$ . Under the parameter choice  $\lambda = \mathcal{O}(\delta^2)$ , we have the convergence rates  $\|u_\lambda^\delta - u^\dagger\|_{H^1(\Omega)} = \mathcal{O}(\delta)$  and  $\|f_\lambda^\delta - f^\dagger\|_{L^2(\Omega)} = \mathcal{O}(\delta^{1/2})$ .

Since there is no way to calculate expectations and integrals in (2.2) directly, we employ Monte Carlo methods to approximate them by empirical averages. Let  $\{x_i^\Omega\}_{i=1}^m$  and  $\{x_i^\Gamma\}_{i=1}^m$  be sets of independently and identically distributed samples drawn from  $U(\Omega)$  and  $U(\partial\Omega)$ , respectively. Then we define the empirical objective functional by

$$\widehat{J}_\lambda(u, f) = \frac{|\Omega|}{m} \sum_{i=1}^m (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 + \frac{|\Omega|}{m} \sum_{i=1}^m \|\nabla u(x_i^\Omega) - z^\delta(x_i^\Omega)\|_2^2 + \lambda \widehat{\mathcal{R}}(f),$$

where

$$\widehat{\mathcal{R}}(f) = \frac{|\Omega|}{m} \sum_{i=1}^m (f^2(x_i^\Omega) + \|\nabla f(x_i^\Omega)\|_2^2)$$

is the empirical counterpart of the squared  $H^1(\Omega)$ -norm of  $f$ . Therefore the empirical risk is given by

$$\widehat{L}_\lambda(u, f) = \widehat{J}_\lambda(u, f) + \widehat{R}_{\text{int}}(u, f) + \widehat{R}_{\text{bdy}}(u), \quad (2.3)$$

where the empirical physics-informed penalty terms are defined by

$$\begin{aligned}\widehat{R}_{\text{int}}(u, f) &= \frac{|\Omega|}{m} \sum_{i=1}^m (f(x_i^\Omega) + \Delta u(x_i^\Omega) - V(x_i^\Omega)u(x_i^\Omega))^2, \\ \widehat{R}_{\text{bdy}}(u) &= \frac{|\partial\Omega|}{m} \sum_{i=1}^m (\partial_n u(x_i^\Gamma) - g(x_i^\Gamma))^2.\end{aligned}$$

It is straightforward to verify that  $\mathbb{E}\widehat{L}_\lambda(u, f) = L_\lambda(u, f)$  for each fixed  $(u, f)$ . We next select two classes of neural networks  $\mathcal{U} \subseteq H^2(\Omega)$  and  $\mathcal{F} \subseteq H^1(\Omega)$ , namely hypothesis classes. Minimizing the empirical risk (2.3) with respect to  $(u, f) \in \mathcal{U} \times \mathcal{F}$  yields an estimator of  $(u^\dagger, f^\dagger)$  as follows:

$$(\widehat{u}_\lambda^\delta, \widehat{f}_\lambda^\delta) \in \arg \min_{(u, f) \in \mathcal{U} \times \mathcal{F}} \widehat{L}_\lambda(u, f), \quad (2.4)$$

which is called the empirical risk minimizer (ERM). See Algorithm 2.1 for more details.

**Algorithm 2.1** Recover the Unknown Source in Elliptic Equations.

- 
- 1: Sample point sets  $\{x_i^\Omega\}_{i=1}^m \subseteq \Omega$  and  $\{x_i^\Gamma\}_{i=1}^m \subseteq \partial\Omega$ .
  - 2: Construct neural networks  $(u(\theta), f(\phi)) \in H^2(\Omega) \times H^1(\Omega)$ .
  - 3: Initialize parameters  $(\theta, \phi)$  randomly.
  - 4: **for**  $k = 1 : \text{num\_epochs}$  **do**
  - 5:     Compute the empirical risk defined by (2.3)
- $$\widehat{L}_\lambda(u(\theta), f(\phi)) = \widehat{J}_\lambda(u(\theta), f(\phi)) + \widehat{R}_{\text{int}}(u(\theta), f(\phi)) + \widehat{R}_{\text{bdy}}(u(\theta)).$$
- 6:     Back propagation:  $(g_\theta, g_\phi) = \nabla_{(\theta, \phi)} \widehat{L}_\lambda(u(\theta), f(\phi))$ .
  - 7:     Update  $(\theta, \phi)$  by a SGD-type algorithm  $(\theta, \phi) \leftarrow \text{SGD}\{(\theta, \phi), (g_\theta, g_\phi), \alpha\}$ .
  - 8: **end for**
- Output:** Estimator  $(u(\theta), f(\phi))$ .
- 

### 3. Convergence Analysis

Up to now, we have proposed convergence rates of our method in population level (see details in Remark 2.1). However, the rate of convergence of the reconstruction  $(\widehat{u}_\lambda^\delta, \widehat{f}_\lambda^\delta)$  is what we are really interested in. We will study in this section the oracle inequality and convergence rates of methods proposed in Section 2 for the recovery of the source density in the elliptic system (1.1). Throughout this section, we assume that  $V \in L^\infty(\Omega)$ ,  $f \in L^2(\Omega)$  and  $g \in H^{1/2}(\partial\Omega)$  in (1.1).

#### 3.1. Oracle inequality

In this section, we present an oracle inequality that decomposes the population risk of the reconstruction into four components: approximation error, generalization error, regularization term, and noise term. Oracle inequalities have been extensively studied in the field of non-parametric regression, as evidenced by previous works such as [13, 16, 22, 34, 42, 43, 52, 57]. In addition, several works have specifically proposed oracle inequalities (in the expectation form) for PINNs-type methods, including [19, 31, 48].

Before proceeding, we introduce some notations as follows. Suppose  $B_{\mathcal{U}} \geq 1$  and  $B_{\mathcal{F}} \geq 1$  are two constant defined as follows:

$$\begin{aligned} B_{\mathcal{U}} &= \sup_{u \in \mathcal{U} \cup \{u^\dagger\}} \left\{ \|u\|_{L^\infty(\Omega)} + d \max_{j \in [d]} \|\partial_j u\|_{L^\infty(\Omega)} + d \max_{j \in [d]} \|\partial_{jj}^2 u\|_{L^\infty(\Omega)} \right\}, \\ B_{\mathcal{F}} &= \sup_{f \in \mathcal{F} \cup \{f^\dagger\}} \left\{ \|f\|_{L^\infty(\Omega)} + d \max_{j \in [d]} \|\partial_j f\|_{L^\infty(\Omega)} \right\}. \end{aligned}$$

Further, for simplicity of notation we define the function classes  $\partial_j \mathcal{U} = \{\partial_j u : u \in \mathcal{U}\}$ ,  $\partial_{jj}^2 \mathcal{U} = \{\partial_{jj}^2 u : u \in \mathcal{U}\}$ ,  $\partial_n \mathcal{U} = \{\partial_n u : u \in \mathcal{U}\}$  and  $\partial_j \mathcal{F} = \{\partial_j f : f \in \mathcal{F}\}$ , and also define

$$\widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}^{(2)}(\mathcal{U}) = \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) + d \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{U}) + d \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_{jj}^2 \mathcal{U}),$$

$$\widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}^{(1)}(\mathcal{F}) = \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{F}) + d \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{F}).$$

For the population risk proposed in (2.2), we now derive a distribution-free oracle inequality in the high-probability form.

**Theorem 3.1** (A Data-Dependent Oracle Inequality). *Let  $(\hat{u}_\lambda^\delta, \hat{f}_\lambda^\delta)$  be the reconstructions defined by (2.4). Then for each  $\lambda \in (0, 1)$  and  $\tau \in (0, 1)$ , the following inequality holds with probability at least  $1 - (13 + 7d)\tau$ :*

$$L_\lambda(\hat{u}_\lambda^\delta, \hat{f}_\lambda^\delta) \leq C\mathcal{E}_{\text{app}}(\mathcal{U}, \mathcal{F}) + C\mathcal{E}_{\text{gen}}(\mathcal{U}, \mathcal{F}, n) + 2\lambda\|f^\dagger\|_{H^1(\Omega)}^2 + C\delta^2,$$

where the approximation error  $\mathcal{E}_{\text{app}}$  and the generalization error  $\mathcal{E}_{\text{gen}}$  are defined as

$$\begin{aligned} \mathcal{E}_{\text{app}}(\mathcal{U}, \mathcal{F}) &= \inf_{u \in \mathcal{U}} \|u - u^\dagger\|_{H^2(\Omega)}^2 + \inf_{f \in \mathcal{F}} \|f - f^\dagger\|_{H^1(\Omega)}^2, \\ \mathcal{E}_{\text{gen}}(\mathcal{U}, \mathcal{F}, m) &= (B_{\mathcal{U}} + B_{\mathcal{F}} + \delta) \left\{ \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}^{(2)}(\mathcal{U}) + \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}^{(1)}(\mathcal{F}) \right\} \\ &\quad + B_{\mathcal{U}} \widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_n \mathcal{U}) + \left\{ B_{\mathcal{U}}^2 + B_{\mathcal{F}}^2 + \delta^2 \right\} \sqrt{\frac{\log(1/\tau)}{2m}}, \end{aligned}$$

and  $C$  is a positive constant only depending on  $d, \Omega, \|V\|_{L^\infty(\Omega)}$  and  $\|g\|_{L^\infty(\partial\Omega)}$ .

**Remark 3.1** (Approximation Error). Lemma 1.1 shows that if we choose neural network classes  $(\mathcal{U}, \mathcal{F})$  with sufficiently large depth, number of non-zero samples and weight radius, the approximation error  $\mathcal{E}_{\text{app}}$  in Theorem 3.1 can be arbitrary small.

**Remark 3.2** (Generalization Error). Observe from Theorem 3.1 that the generalization error is dominated by empirical Rademacher complexities, which only depends on data sets  $\{x_i^\Omega\}_{i=1}^m$  and  $\{x_i^\Gamma\}_{i=1}^m$  (data-dependent).

**Remark 3.3.** The oracle inequality shown in Theorem 3.1 demonstrates that the population risk achieved by the reconstructions is nearly as small as the summation of the optimal risk (noise term)  $\mathcal{O}(\delta^2)$  and the regularization parameter  $\mathcal{O}(\lambda)$ . This means that  $L_\lambda(\hat{u}_\lambda^\delta, \hat{f}_\lambda^\delta) = \mathcal{O}(\delta^2) + \mathcal{O}(\lambda)$ , when we choose appropriate neural network classes  $(\mathcal{U}, \mathcal{F})$  and sufficiently large number of samples  $m$ .

### 3.2. Convergence rates

The oracle inequality proposed in Section 3.1 shows that for fixed noise level and regularization parameter, the population risk is dominated by the approximation error and generalization error, which depend on the selection of neural network classes  $(\mathcal{U}, \mathcal{F})$ . Lemma 1.1 shows how the approximation error decreases as the size of neural networks increase. The generalization error shall be discussed in this section.

We first estimate the bounds of function classes associated with  $(\mathcal{U}, \mathcal{F})$ .

**Lemma 3.1.** Let the activation  $\varrho$  be set as  $\varrho = \tanh$ . Let the neural network classes  $(\mathcal{U}, \mathcal{F})$  be  $\mathcal{U} = \mathcal{N}_\varrho(L_{\mathcal{U}}, S_{\mathcal{U}}, R_{\mathcal{U}})$  and  $\mathcal{F} = \mathcal{N}_\varrho(L_{\mathcal{F}}, S_{\mathcal{F}}, R_{\mathcal{F}})$ . It follows that

$$\begin{aligned} B_{\mathcal{U}} &\leq R_{\mathcal{U}} S_{\mathcal{U}} + dR_{\mathcal{U}}^{L_{\mathcal{U}}} S_{\mathcal{U}}^{L_{\mathcal{U}}} + 2dR_{\mathcal{U}}^{2L_{\mathcal{U}}} S_{\mathcal{U}}^{2L_{\mathcal{U}}}, \\ B_{\mathcal{F}} &\leq R_{\mathcal{F}} S_{\mathcal{F}} + dR_{\mathcal{F}}^{L_{\mathcal{F}}} S_{\mathcal{F}}^{L_{\mathcal{F}}}. \end{aligned}$$

It remains to consider Rademacher complexities of function classes defined in Theorem 3.1. How to compute Rademacher complexity of a neural network class has been widely discussed in the literature on statistical learning theory [3, 51]. However, there is no uniform method for calculating Rademacher complexity of the derivative class of neural networks because the derivative operator is not Lipschitz continuous. We employ a similar argument to [19, 32, 35–37] to calculate Rademacher complexity of the tanh-neural network and its derivative class as follows.

**Lemma 3.2.** Let the activation  $\varrho$  be set as  $\varrho = \tanh$ . Let the neural network class  $\mathcal{U}$  be  $\mathcal{U} = \mathcal{N}_\varrho(L_{\mathcal{U}}, S_{\mathcal{U}}, R_{\mathcal{U}})$ . It follows that

$$\begin{aligned} \sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) &\leq c \sqrt{S_{\mathcal{U}}^3 R_{\mathcal{U}}^2 L_{\mathcal{U}} \log(S_{\mathcal{U}} R_{\mathcal{U}}) \frac{\log m}{m}}, \\ \max_{j \in [d]} \sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{U}) &\leq c \sqrt{S_{\mathcal{U}}^{2L+1} R_{\mathcal{U}}^{2L} L_{\mathcal{U}} \log(S_{\mathcal{U}} R_{\mathcal{U}}) \frac{\log m}{m}}, \\ \max_{j \in [d]} \sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_{jj}^2 \mathcal{U}) &\leq c \sqrt{S_{\mathcal{U}}^{4L+1} R_{\mathcal{U}}^{4L} L_{\mathcal{U}} \log(S_{\mathcal{U}} R_{\mathcal{U}}) \frac{\log m}{m}}, \\ \sup_{\{x_i^\Gamma\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_n \mathcal{U}) &\leq c \sqrt{d S_{\mathcal{U}}^{2L+1} R_{\mathcal{U}}^{2L} L_{\mathcal{U}} \log(S_{\mathcal{U}} R_{\mathcal{U}}) \frac{\log m}{m}}, \end{aligned}$$

where  $c$  is an absolute constant.

It is straightforward to obtain the following Rademacher complexity estimates corresponding to neural network class  $\mathcal{F}$ .

**Corollary 3.1.** Let the activation  $\varrho$  be set as  $\varrho = \tanh$ . Let the neural network class  $\mathcal{F}$  be  $\mathcal{F} = \mathcal{N}_\varrho(L_{\mathcal{F}}, S_{\mathcal{F}}, R_{\mathcal{F}})$ . It follows that

$$\begin{aligned} \sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{F}) &\leq c \sqrt{S_{\mathcal{F}}^3 R_{\mathcal{F}}^2 L_{\mathcal{F}} \log(S_{\mathcal{F}} R_{\mathcal{F}}) \frac{\log m}{m}}, \\ \max_{j \in [d]} \sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{F}) &\leq c \sqrt{S_{\mathcal{F}}^{2L+1} R_{\mathcal{F}}^{2L} L_{\mathcal{F}} \log(S_{\mathcal{F}} R_{\mathcal{F}}) \frac{\log m}{m}}, \end{aligned}$$

where  $c$  is an absolute constant.

**Remark 3.4.** It is worth noting, based on Lemmas 3.1, 3.2, and Corollary 3.1, that the generalization error, signifying the discrepancy between the population risk (expectation) and the empirical risk (finite average), is influenced not only by the sample size but also by the size of neural networks. In contrast, the generalization bounds presented in [63, Theorem 3.3] merely quantify the convergence with respect to the sample size, without considering the impact of the size of neural networks. As a result, the findings in [63, Theorem 3.3] fail to provide a theoretical understanding of how the size of neural networks affects the error.

**Remark 3.5** (Trade-Offs Between Approximation and Generalization). It is apparent from Lemma 3.2 and Corollary 3.1 that as the depth, the number of non-zero weights, and the bound on weights become larger, the generalization error will become larger, while Lemma 1.1 states that the approximation error will decrease. Therefore, there exists a trade-off between the approximation and generalization errors.

With the aid of Theorem 3.1, Lemmas 1.1, 3.1, 3.2 and Corollary 3.1, we finally achieve the following convergence rates of reconstructions of source density in elliptic equations.

**Theorem 3.2** (Convergence Rates of Population Risk). *Let  $\delta \in (0, 1)$ ,  $\alpha \geq 1$  and  $\varrho = \tanh$ . Suppose that  $(u^\dagger, f^\dagger) \in H^{\alpha+2}(\Omega) \times H^{\alpha+1}(\Omega)$ . Set  $\mathcal{U} = N_\varrho(L_{\mathcal{U}}, S_{\mathcal{U}}, R_{\mathcal{U}})$  and  $\mathcal{F} = N_\varrho(L_{\mathcal{F}}, S_{\mathcal{F}}, R_{\mathcal{F}})$ , where*

$$\begin{aligned} L_{\mathcal{U}} &= \mathcal{O}(1), \quad S_{\mathcal{U}} = \mathcal{O}(\varepsilon^{-d/(\alpha-\mu)}), \\ R_{\mathcal{U}} &= \mathcal{O}(\varepsilon^{-(9d+4+4\mu)/(2(\alpha-\mu))-2}), \\ L_{\mathcal{F}} &= \mathcal{O}(1), \quad S_{\mathcal{F}} = \mathcal{O}(\varepsilon^{-d/(\alpha-\mu)}), \\ R_{\mathcal{F}} &= \mathcal{O}(\varepsilon^{-(9d+2+4\mu)/(2(\alpha-\mu))-2}). \end{aligned}$$

Suppose that the number of samples  $m$  is larger than  $\mathcal{O}(\varepsilon^{-4-c((d+1)/(\alpha-\mu))\log(d+2)})$ . Then the following inequality holds for each  $\lambda \in (0, 1)$  with probability at least  $1 - (13 + 7d)\tau$ :

$$L_\lambda(\hat{u}_\lambda^\delta, \hat{f}_\lambda^\delta) \leq \mathcal{O}\left(\varepsilon^2 \log^{1/2}\left(\frac{1}{\varepsilon}\right)\right) + \mathcal{O}(\lambda) + \mathcal{O}(\delta^2) + \mathcal{O}\left(\varepsilon^{(d+4(\alpha-\mu))/(2(\alpha-\mu))} \log^{1/2}\left(\frac{1}{\tau}\right)\right).$$

By combining Theorem 3.2 with stability estimates (Theorem 2.1), we obtain the following convergence rates of reconstructions.

**Corollary 3.2** (Convergence Rates of Reconstructions). *Let  $\delta \in (0, 1)$ ,  $\alpha \geq 1$  and  $\varrho = \tanh$ . Suppose that  $(u^\dagger, f^\dagger) \in H^{\alpha+2}(\Omega) \times H^{\alpha+1}(\Omega)$ . Set the regularization parameter as  $\lambda = \mathcal{O}(\delta^2)$ , and set  $\mathcal{U} = N_\varrho(L_{\mathcal{U}}, S_{\mathcal{U}}, R_{\mathcal{U}})$  and  $\mathcal{F} = N_\varrho(L_{\mathcal{F}}, S_{\mathcal{F}}, R_{\mathcal{F}})$ , where the hyper-parameters are given by  $L_{\mathcal{U}} = \mathcal{O}(1)$ ,  $S_{\mathcal{U}} = \mathcal{O}(\delta^{-d/(\alpha-\mu)})$ ,  $R_{\mathcal{U}} = \mathcal{O}(\delta^{-(9d+4+4\mu)/(2(\alpha-\mu))-2})$ , and  $L_{\mathcal{F}} = \mathcal{O}(1)$ ,  $S_{\mathcal{F}} = \mathcal{O}(\delta^{-d/(\alpha-\mu)})$ ,  $R_{\mathcal{F}} = \mathcal{O}(\delta^{-(9d+2+4\mu)/(2(\alpha-\mu))-2})$ . Suppose that the number of samples  $m$  is larger than  $\mathcal{O}(\delta^{-4-c((d+1)/(\alpha-\mu))\log(d+2)})$ . Then for  $\tau > 0$  with probability at least  $1 - (13 + 7d)\tau$ , the following inequalities hold:*

$$\begin{aligned} \|\hat{u}_\lambda^\delta - u^\dagger\|_{H^1(\Omega)} &= \mathcal{O}\left(\delta \log^{1/4}\left(\frac{1}{\delta}\right)\right) + \mathcal{O}\left(\delta^{(d+4(\alpha-\mu))/(4(\alpha-\mu))} \log^{1/4}\left(\frac{1}{\tau}\right)\right), \\ \|\hat{f}_\lambda^\delta - f^\dagger\|_{L^2(\Omega)} &= \mathcal{O}\left(\delta^{1/2} \log^{1/8}\left(\frac{1}{\delta}\right)\right) + \mathcal{O}\left(\delta^{(d+4(\alpha-\mu))/(8(\alpha-\mu))} \log^{1/8}\left(\frac{1}{\tau}\right)\right). \end{aligned}$$

**Remark 3.6** (High-Probability Form). For a fixed  $\tau$ , the last term  $\mathcal{O}(\varepsilon^{(d+4(\alpha-\mu))/(2(\alpha-\mu))})$  in Theorem 3.2 is strictly smaller than the first one  $\mathcal{O}(\varepsilon^2)$ . Let  $\iota = \log(1/\tau)$ , then with probability  $1 - (13 + 7d)\exp(\iota)$  the following estimate holds,  $L_\lambda(\hat{u}_\lambda^\delta, \hat{f}_\lambda^\delta) \leq \mathcal{O}(\varepsilon^2) + \mathcal{O}(\lambda) + \mathcal{O}(\delta^2) + \mathcal{O}(\varepsilon^{(d+4(\alpha-\mu))/(2(\alpha-\mu))}\iota^{1/2})$ . This shows the probability that this estimate does not hold decays exponentially as the last term increases. This analysis also holds for Corollary 3.2.

**Remark 3.7** (Overcome CoD Under High-Regularity). In Theorem 3.2 and Corollary 3.2, for a fixed regularity index  $\alpha$ , the number of non-zero weights and the bound of weights depend exponentially on the dimension  $d$ , namely the curse of dimensionality (CoD). However, if the ground truth  $(u^\dagger, f^\dagger)$  have high-regularity, that is,  $\alpha = \Omega(d)$ , the number of non-zero weights and the bound of weights could be dimension-independent. It is worth noting that in the literature, other dimension-free rates are based on the assumption that the target function is in Barron-type class [20, 46, 49]. According to [11, (15) in Section IX], functions with derivatives of sufficiently high order belong to Barron class. More precisely, if the partial derivatives of  $g(x)$  of order  $\alpha = \lfloor d/2 \rfloor + 2$  are continuous, then  $g$  is in Barron class. In this sense, the analyses we give follow a similar path as previous results in the literature.

#### 4. Numerical Experiments

In this section, we present numerical reconstructions using method proposed in this paper. The accuracy of reconstructions are measured by the relative  $L^2$ -errors  $\text{err}(u)$ ,  $\text{err}(\partial_j u)$  and  $\text{err}(f)$  defined as

$$\text{err}(u) = \frac{\|u - u^\dagger\|_2}{\|u^\dagger\|_2}, \quad \text{err}(\partial_j u) = \frac{\|\partial_j u - \partial_j u^\dagger\|_2}{\|\partial_j u^\dagger\|_2}, \quad \text{err}(f) = \frac{\|f - f^\dagger\|_2}{\|f^\dagger\|_2}.$$

Given a ground truth source density  $f^\dagger$  and a potential function  $V$ , the ground truth solution  $u^\dagger$  is generated by solving the following elliptic equation with zero Dirichlet boundary condition:

$$\begin{aligned} -\Delta u + Vu &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{4.1}$$

Then the boundary flux can obtained by  $g(x) = \partial_n u^\dagger(x)$  for  $x \in \partial\Omega$ . And the noisy measurement data set  $\{(x_i, y_i^\delta, z_i^\delta)\}_{i=1}^m$  can be generated from (1.2).

All the numerical experiments in this paper have been done on the supercomputing system with RAM 128GB ECC 2400MHz DDR4, Intel Xeon E5-2640v4 2.4GHz (CPU) and NVIDIA Tesla V100 16GB NVLink (GPU). The partial differential equation (4.1) is solved by the PDE toolbox of MATLAB 2023b, and our method was implemented with Python 3.9 on PyTorch 1.13 [54].

We approximate the unknown solution and source density using two neural networks, both of which have two hidden layers, each layer containing 64 units. In this paper, we use the Adam [41] optimizer with a learning rate  $1.0 \times 10^{-4}$ . We use 100000 sample points and set the batch size as 2048. In all experiments, we set the number of epochs as 50000.

In the first test, we consider an elliptic system with an unknown smooth source density.

**Example 4.1** (Smooth Source Density). Let  $\Omega = (0, 1)^2$  and  $V(x) = 1$  for each  $x \in \Omega$ . The ground truth source density  $f^\dagger$  are given as follows:

$$\begin{aligned}\phi_1(x_1, x_2) &= \exp(-9 \times (x_1 - 0.3)^2 - 25 \times (x_2 - 0.7)^2), \\ \phi_2(x_1, x_2) &= \exp(-25 \times (x_1 - 0.7)^2 - 9 \times (x_2 - 0.3)^2), \\ f^\dagger(x_1, x_2) &= 25\phi_1(x_1, x_2) + 36\phi_2(x_1, x_2).\end{aligned}$$

The Table 1 presents the relative  $L^2$ -error of the reconstructions in Example 4.1 for various noise levels. It is evident from the table that our method maintains a high level of accuracy even as the noise level increases. Notably, acceptable reconstruction accuracy is achieved even with noise levels as high as 50%. This observation is further supported by Fig. 1.

Table 1: The relative  $L^2$ -error of reconstructions under various noise levels in Example 4.1.

$\lambda = 1.0 \times 10^{-8}$	Noise level $\delta$				
	0%	1%	10%	20%	50%
$\text{err}(\widehat{u}_\lambda^\delta)$	$5.25 \times 10^{-4}$	$7.50 \times 10^{-4}$	$1.13 \times 10^{-3}$	$2.15 \times 10^{-3}$	$3.06 \times 10^{-3}$
$\text{err}(\partial_1 \widehat{u}_\lambda^\delta)$	$3.50 \times 10^{-3}$	$3.90 \times 10^{-3}$	$4.15 \times 10^{-3}$	$5.42 \times 10^{-3}$	$7.50 \times 10^{-3}$
$\text{err}(\partial_2 \widehat{u}_\lambda^\delta)$	$3.31 \times 10^{-3}$	$3.92 \times 10^{-3}$	$3.97 \times 10^{-3}$	$5.78 \times 10^{-3}$	$7.62 \times 10^{-3}$
$\text{err}(\widehat{f}_\lambda^\delta)$	$1.14 \times 10^{-2}$	$1.39 \times 10^{-2}$	$1.47 \times 10^{-2}$	$1.86 \times 10^{-2}$	$2.80 \times 10^{-2}$

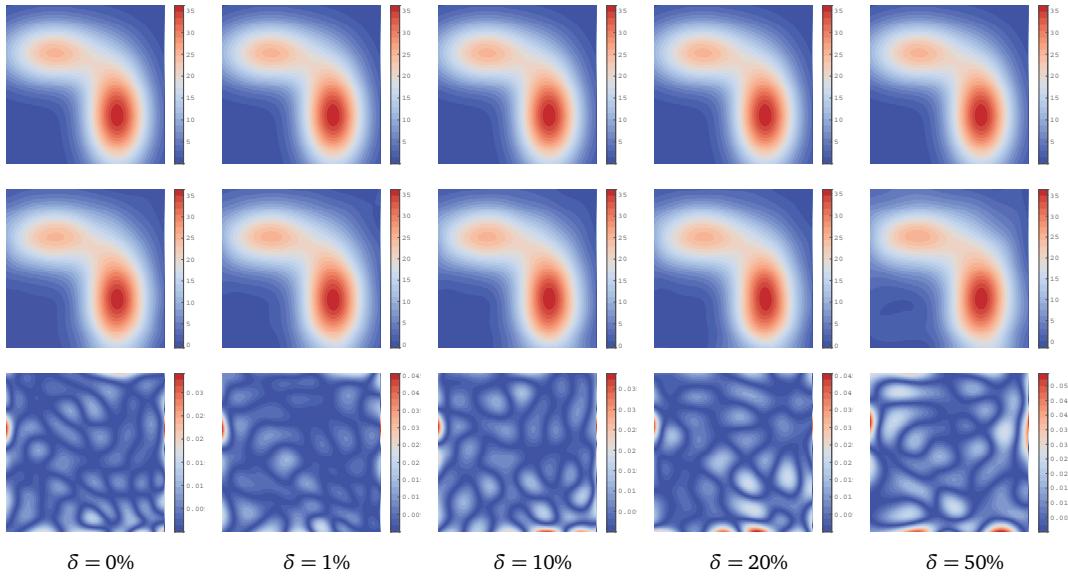


Figure 1: The ground truth source density  $f^\dagger$  (top), the recovered source density  $\widehat{f}_\lambda^\delta$  (middle), and the relative point-wise absolute error of the reconstruction  $|\widehat{f}_\lambda^\delta - f^\dagger| / \|f^\dagger\|_\infty$  (bottom) with regularization parameter  $\lambda = 1.0 \times 10^{-8}$  under different noise levels in Example 4.1.

Furthermore, both Table 1 and Fig. 2 demonstrate that the reconstructed solution  $\hat{u}_\lambda^\delta$  approximates the ground truth solution  $u^\dagger$  not only in terms of the  $L^2$ -norm but also in the  $H^1$ -semi-norm. This finding supports our stability estimates (Theorem 2.1) and the analysis of convergence rates (Corollary 3.2).

To further investigate the impact of explicit regularization on the reconstruction, we present numerical results with different regularization parameters  $\lambda$  under the noise level 50%, as displayed in Table 2 and Fig. 3. The experimental results illustrate that the re-

Table 2: The relative  $L^2$ -error of reconstructions with various regularization parameters in Example 4.1.

$\delta = 50\%$	Regularization parameters $\lambda$				
	$1.0 \times 10^{-5}$	$1.0 \times 10^{-6}$	$1.0 \times 10^{-7}$	$1.0 \times 10^{-8}$	0.0
$\text{err}(\hat{u}_\lambda^\delta)$	$1.10 \times 10^{-2}$	$3.63 \times 10^{-3}$	$7.96 \times 10^{-3}$	$8.77 \times 10^{-3}$	$4.78 \times 10^{-3}$
$\text{err}(\partial_1 \hat{u}_\lambda^\delta)$	$2.72 \times 10^{-2}$	$7.68 \times 10^{-3}$	$6.94 \times 10^{-3}$	$7.11 \times 10^{-3}$	$7.97 \times 10^{-3}$
$\text{err}(\partial_2 \hat{u}_\lambda^\delta)$	$2.33 \times 10^{-2}$	$7.74 \times 10^{-3}$	$7.09 \times 10^{-3}$	$7.20 \times 10^{-3}$	$7.92 \times 10^{-3}$
$\text{err}(\hat{f}_\lambda^\delta)$	$6.58 \times 10^{-2}$	$2.68 \times 10^{-2}$	$2.63 \times 10^{-2}$	$2.62 \times 10^{-2}$	$2.99 \times 10^{-2}$

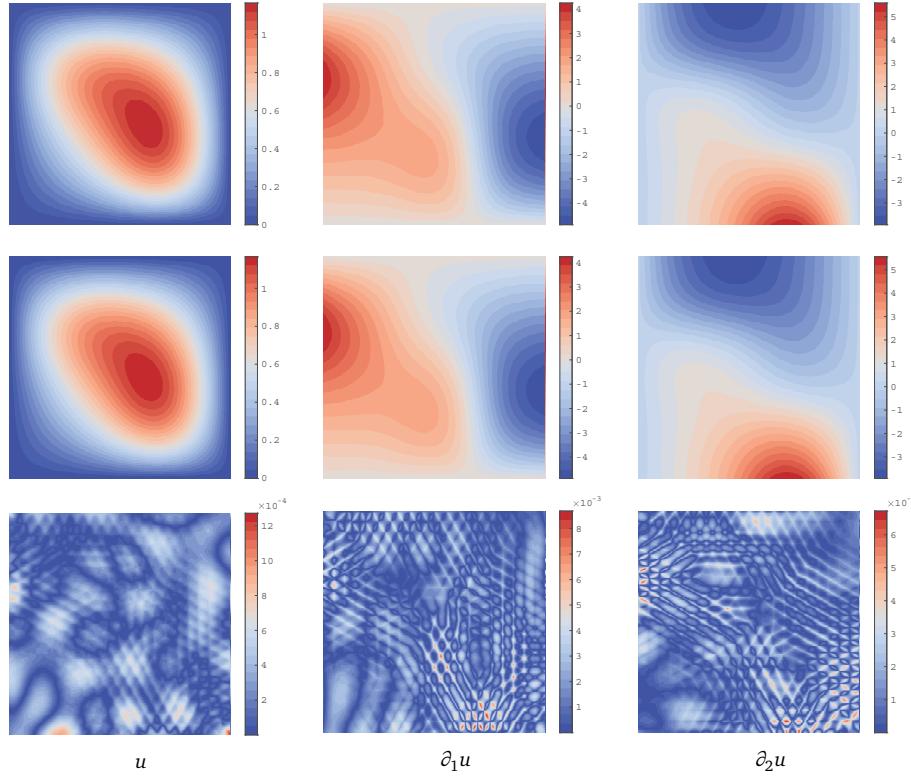


Figure 2: The ground truth solution  $(u^\dagger, \partial_1 u^\dagger, \partial_2 u^\dagger)$  (top), the recovered solution  $(\hat{u}_\lambda^\delta, \partial_1 \hat{u}_\lambda^\delta, \partial_2 \hat{u}_\lambda^\delta)$  (middle), and the relative point-wise absolute error of the reconstruction  $(|\hat{u}_\lambda^\delta - u^\dagger|/\|u^\dagger\|_\infty, |\partial_1 \hat{u}_\lambda^\delta - \partial_1 u^\dagger|/\|\partial_1 u^\dagger\|_\infty, |\partial_2 \hat{u}_\lambda^\delta - \partial_2 u^\dagger|/\|\partial_2 u^\dagger\|_\infty)$  (bottom) under the noise level  $\delta = 10\%$  in Example 4.1.

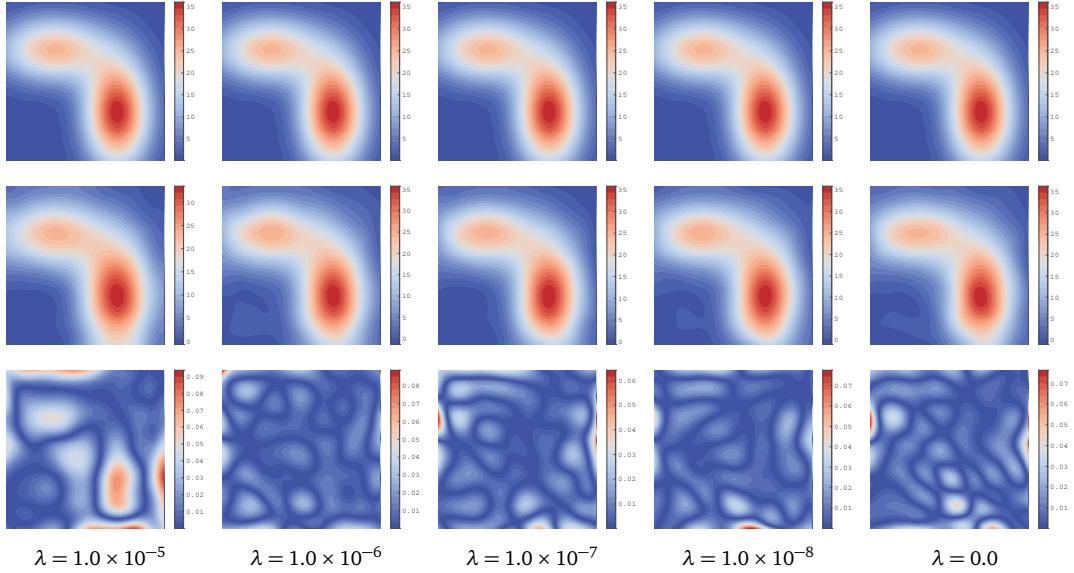


Figure 3: The ground truth source density  $f^\dagger$  (top), the recovered source density  $\hat{f}_\lambda^\delta$  (middle), and the relative point-wise absolute error of the reconstruction  $|\hat{f}_\lambda^\delta - f^\dagger|/\|f^\dagger\|_\infty$  (bottom) under noise level  $\delta = 50\%$  with different regularization parameters  $\lambda$  in Example 4.1.

construction accuracy is insensitive to the selection of regularization parameter. Notably, the method exhibits high accuracy within the range of regularization parameters from  $1.0 \times 10^{-6}$  to  $1.0 \times 10^{-8}$ .

The second test consider the non-smooth source density.

**Example 4.2** (Non-Smooth Source Density). Let  $\Omega = (0, 1)^2$  and  $V(x) = 1$  for each  $x \in \Omega$ . The ground truth source density are given as follows:

$$\phi(x_1, x_2) = \exp(-16 \times (x - 0.6)^2 - 16.0 \times (y - 0.4)^2),$$

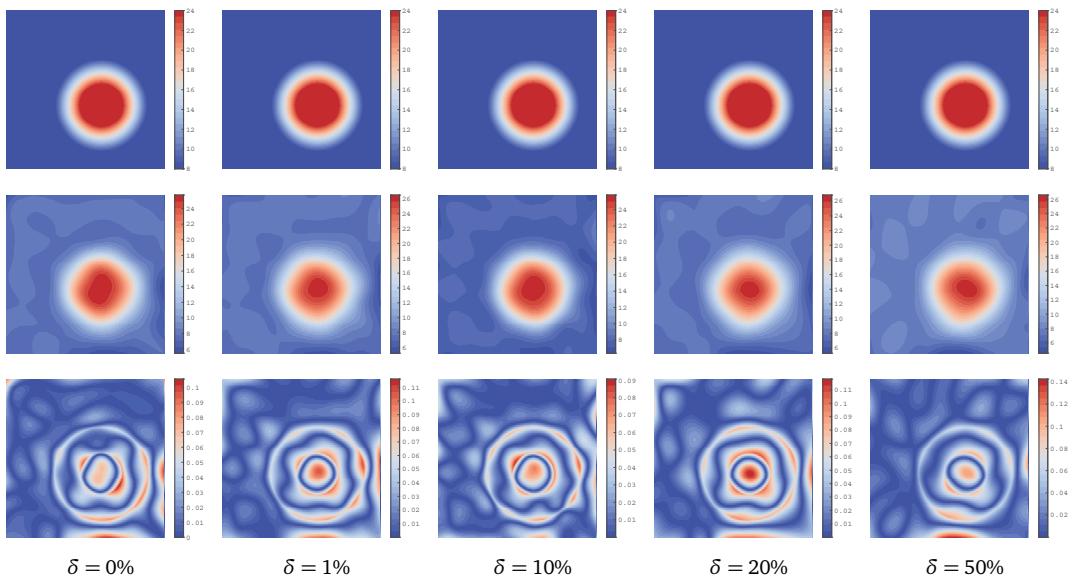
$$f^\dagger(x_1, x_2) = \begin{cases} 8, & \phi(x_1, x_2) \in (-\infty, 0.25], \\ 32\phi(x_1, x_2), & \phi(x_1, x_2) \in (0.25, 0.75], \\ 24, & \phi(x_1, x_2) \in (0.75, +\infty). \end{cases}$$

It is worth noting that the ground truth source density in Example 4.2 is not smooth, which does not align with the conditions outlined in the theoretical results (Theorem 3.2 and Corollary 3.2) presented in Section 3. However, the excellent performance of our method in non-smooth situations is evident from Table 3 and Fig. 4.

Similar to Example 4.2, our method demonstrates robustness against data noise, as high reconstruction accuracy is maintained even with noise levels of up to 50%. Furthermore, it is important to note that the reconstructed error is primarily concentrated at the discontinuities of the source density.

Table 3: The relative  $L^2$ -error of reconstructions under various noise levels in Example 4.2.

$\lambda = 1.0 \times 10^{-8}$	Noise level $\delta$				
	0%	1%	10%	20%	50%
$\text{err}(\widehat{u}_\lambda^\delta)$	$2.22 \times 10^{-3}$	$1.87 \times 10^{-3}$	$2.34 \times 10^{-3}$	$3.27 \times 10^{-3}$	$4.88 \times 10^{-3}$
$\text{err}(\partial_1 \widehat{u}_\lambda^\delta)$	$1.20 \times 10^{-2}$	$1.01 \times 10^{-2}$	$9.96 \times 10^{-3}$	$1.11 \times 10^{-2}$	$1.38 \times 10^{-2}$
$\text{err}(\partial_2 \widehat{u}_\lambda^\delta)$	$1.02 \times 10^{-2}$	$9.35 \times 10^{-3}$	$1.18 \times 10^{-2}$	$1.33 \times 10^{-2}$	$1.70 \times 10^{-2}$
$\text{err}(\widehat{f}_\lambda^\delta)$	$5.71 \times 10^{-2}$	$4.96 \times 10^{-2}$	$5.41 \times 10^{-2}$	$5.90 \times 10^{-2}$	$6.95 \times 10^{-2}$

Figure 4: The ground truth source density  $f^\dagger$  (top), the recovered source density  $\widehat{f}_\lambda^\delta$  (middle), and the relative point-wise absolute error of the reconstruction  $|\widehat{f}_\lambda^\delta - f^\dagger|/\|f^\dagger\|_\infty$  (bottom) under different noise levels in Example 4.2.

## 5. Proofs of Lemmas and Theorems

### 5.1. Proofs of stability estimates

*Proof.* [Proof of Theorem 2.1] It is obvious from (1.2) and (2.2) that  $\|u - u^\dagger\|_{H^1(\Omega)}^2 \leq L_\lambda(u, f)$ , which implies the first inequality of stability estimates. We next consider the stability estimate for the source density. Let  $\psi \in H^1(\Omega)$  and  $(u, f) \in H^2(\Omega) \times H^1(\Omega)$ . Applying the trace theorem, we have that  $\|T\psi\|_{L^2(\partial\Omega)} \leq C\|\psi\|_{H^1(\Omega)}$  for some positive constant  $C$  only depending on  $\Omega$ . As a consequence, it follows that

$$\begin{aligned}
 & \langle f - f^\dagger, \psi \rangle_{(H^1(\Omega))^*, H^1(\Omega)} = (f - f^\dagger, \psi)_{L^2(\Omega)} \\
 &= (f + \Delta u - Vu, \psi)_{L^2(\Omega)} + (\Delta(u^\dagger - u), \psi)_{L^2(\Omega)} + (V(u - u^\dagger), \psi)_{L^2(\Omega)} \\
 &= (f + \Delta u - Vu, \psi)_{L^2(\Omega)} + (\nabla(u - u^\dagger), \nabla\psi)_{L^2(\Omega)}
 \end{aligned}$$

$$\begin{aligned} & + (g - \partial_n u, T\psi)_{L^2(\partial\Omega)} + (V(u - u^\dagger), \psi)_{L^2(\Omega)} \\ & \leq C (\|f + \Delta u - Vu\|_{L^2(\Omega)} + \|u - u^\dagger\|_{H^1(\Omega)} + \|\partial_n u - g\|_{L^2(\partial\Omega)}) \|\psi\|_{H^1(\Omega)}, \end{aligned}$$

where  $C$  is a positive constant only depending on  $\Omega$  and  $\|V\|_{L^\infty(\Omega)}$ . Here the first equality holds from (1.1), the second equality used Green's formula, and the inequality is due to Cauchy-Schwarz inequality and the trace theorem. Then according to the definition of the dual norm, we find that

$$\begin{aligned} \|f - f^\dagger\|_{(H^1(\Omega))^*} & \leq C (\|f + \Delta u - Vu\|_{L^2(\Omega)} + \|u - u^\dagger\|_{H^1(\Omega)} + \|\partial_n u - g\|_{L^2(\partial\Omega)}) \\ & \leq CL_\lambda^{1/2}(u, f) + 2C\delta, \end{aligned}$$

where the second inequality holds from the triangular inequality and (1.3). On the other hand, it follows from (2.2) that

$$\|f\|_{H^1(\Omega)} \leq \lambda^{-1/2} L_\lambda^{1/2}(u, f).$$

Then we obtain that

$$\begin{aligned} \|f - f^\dagger\|_{L^2(\Omega)}^2 & \leq \|f - f^\dagger\|_{H^1(\Omega)} \|f - f^\dagger\|_{(H^1(\Omega))^*} \\ & \leq C \|f^\dagger\|_{H^1(\Omega)} (1 + \lambda^{-1/2} L_\lambda^{1/2}(u, f)) (L_\lambda^{1/2}(u, f) + \delta), \end{aligned}$$

which completes the proof.  $\square$

## 5.2. Proof of the oracle inequality

Define the population excess risk of  $(u, f)$  as

$$\mathcal{E}_\lambda(u, f) = \|u - u^\dagger\|_{H^1(\Omega)}^2 + \lambda \|f\|_{H^1(\Omega)}^2 + R_{\text{int}}(u, f) + R_{\text{bdy}}(u).$$

It is apparent that  $L_\lambda(u, f) = \mathcal{E}_\lambda(u, f) + 2(1+d)\delta^2$ . For ease of notations, we define the functional  $G_\lambda(\cdot, \cdot)$  as

$$G_\lambda(u, f) = \lambda \|f\|_{H^1(\Omega)}^2 + R_{\text{int}}(u, f) + R_{\text{bdy}}(u).$$

Further, define its empirical counterpart  $\widehat{G}_\lambda(u, f)$  based on  $\{x_i^\Omega\}_{i=1}^m$  and  $\{x_i^\Gamma\}_{i=1}^m$  by

$$\widehat{G}_\lambda(u, f) = \lambda \frac{|\Omega|}{m} \sum_{i=1}^m \left\{ f(x_i^\Omega)^2 + \|\nabla f(x_i^\Omega)\|_2^2 \right\} + \widehat{R}_{\text{int}}(u, f) + \widehat{R}_{\text{bdy}}(u).$$

Throughout this section, we introduce the following notations:

$$\begin{aligned} B_{\mathcal{U}}^{(0)} &= \sup_{u \in \mathcal{U} \cup \{u^\dagger\}} \|u\|_{L^\infty(\Omega)}, & B_{\mathcal{F}}^{(0)} &= \sup_{f \in \mathcal{F} \cup \{f^\dagger\}} \|f\|_{L^\infty(\Omega)}, \\ B_{\mathcal{U}}^{(1)} &= \max_{j \in [d]} \sup_{u \in \mathcal{U} \cup \{u^\dagger\}} \|\partial_j u\|_{L^\infty(\Omega)}, & B_{\mathcal{F}}^{(1)} &= \max_{j \in [d]} \sup_{f \in \mathcal{F} \cup \{f^\dagger\}} \|\partial_j f\|_{L^\infty(\Omega)}, \\ B_{\mathcal{U}}^{(2)} &= \max_{j \in [d]} \sup_{u \in \mathcal{U} \cup \{u^\dagger\}} \|\partial_j^2 u\|_{L^\infty(\Omega)}. \end{aligned}$$

*Proof of Theorem 3.1.* We divide the rest of the proof into two steps.

**Step 1.** Relate the population excess risk with its empirical counterpart. Define a function  $\phi_u^{(0)}$  for each sample  $\{x_i^\Omega\}_{i=1}^m \subseteq \Omega$  as

$$\phi_u^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) = \sup_{u \in \mathcal{U}} \left( \|u - y^\delta\|_{L^2(\Omega)}^2 - \frac{|\Omega|}{m} \sum_{i=1}^m (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 \right).$$

For a point  $x_k^{\Omega,\prime} \in \Omega$  such that  $x_k^{\Omega,\prime} \neq x_k^\Omega$ , we have

$$\begin{aligned} & \phi_u^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) - \phi_u^{(0)}\left(\{x_i^\Omega\}_{i=1}^{k-1} \cup \{x_k^{\Omega,\prime}\} \cup \{x_i^\Omega\}_{i=k+1}^m\right) \\ & \leq \sup_{u \in \mathcal{U}} \left( \frac{|\Omega|}{m} (u(x_k^\Omega) - y^\delta(x_k^\Omega))^2 - \frac{|\Omega|}{m} (u(x_k^{\Omega,\prime}) - y^\delta(x_k^{\Omega,\prime}))^2 \right) \leq \frac{4|\Omega|}{m} (B_{\mathcal{U}}^{(0)} + \delta)^2, \end{aligned}$$

then it follows from McDiarmid's inequality (Lemma A.1) that for each  $\tau \in (0, 1)$ , the following inequality holds with probability at least  $1 - \tau$ :

$$\phi_u^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) \leq \mathbb{E}_{\{x_i\}_{i=1}^m} [\phi_u^{(0)}(\{x_i^\Omega\}_{i=1}^m)] + 4|\Omega| (B_{\mathcal{U}}^{(0)} + \delta)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.1)$$

We next bound the expectation in the right-hand side by the technique of symmetrization. Let  $\{\tilde{x}_i^\Omega\}_{i=1}^m$  be a set of independent copies of  $x^\Omega$  and be independent of  $\{x_i^\Omega\}_{i=1}^m$ . Further, suppose that  $\{\sigma_i\}_{i=1}^m$  is a set of i.i.d. Rademacher variables independent of  $\{x_i^\Omega\}_{i=1}^m$  and  $\{\tilde{x}_i^\Omega\}_{i=1}^m$ . Then it follows from Jensen's inequality that

$$\begin{aligned} & \mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \sup_{u \in \mathcal{U}} \left( \mathbb{E}_{x^\Omega} [(u(x^\Omega) - y^\delta(x^\Omega))^2] - \frac{1}{m} \sum_{i=1}^m (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 \right) \\ & = \mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \sup_{u \in \mathcal{U}} \left( \mathbb{E}_{\{\tilde{x}_i\}_{i=1}^m} \frac{1}{m} \sum_{i=1}^m (u(\tilde{x}_i^\Omega) - y^\delta(\tilde{x}_i^\Omega))^2 - \frac{1}{m} \sum_{i=1}^m (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 \right) \\ & \leq \mathbb{E}_{\{x_i^\Omega, \tilde{x}_i^\Omega\}_{i=1}^m} \sup_{u \in \mathcal{U}} \left( \frac{1}{m} \sum_{i=1}^m (u(\tilde{x}_i^\Omega) - y^\delta(\tilde{x}_i^\Omega))^2 - (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 \right) \\ & = \mathbb{E}_{\{x_i^\Omega, \tilde{x}_i^\Omega, \sigma_i\}_{i=1}^m} \sup_{u \in \mathcal{U}} \left( \frac{1}{m} \sum_{i=1}^m \sigma_i \left( (u(\tilde{x}_i^\Omega) - y^\delta(\tilde{x}_i^\Omega))^2 - (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 \right) \right) \\ & = 2\mathbb{E}_{\{x_i^\Omega, \sigma_i\}_{i=1}^m} \sup_{u \in \mathcal{U}} \left( \frac{1}{m} \sum_{i=1}^m \sigma_i (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 \right) \\ & = 2\mathfrak{R}_m(\{x \mapsto (u(x) - y^\delta(x))^2 : x \in \Omega, u \in \mathcal{U}\}). \end{aligned}$$

Combining this with Ledoux-Talagrand contraction inequality (Lemma A.3), we have

$$\mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} [\phi_u^{(0)}(\{x_i^\Omega\}_{i=1}^m)] \leq 8|\Omega| (B_{\mathcal{U}}^{(0)} + \delta) \mathfrak{R}_m(\mathcal{U} - y^\delta) = 8|\Omega| (B_{\mathcal{U}}^{(0)} + \delta) \mathfrak{R}_m(\mathcal{U}). \quad (5.2)$$

Notice that changing one point on  $\{x_i^\Omega\}_{i=1}^m$  changes  $\widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U})$  by at most  $2B_{\mathcal{U}}^{(0)}/m$ , that is, for each  $k \in [m]$

$$\begin{aligned} & \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) - \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^{k-1} \cup \{x_k^{\Omega,\prime}\} \cup \{x_i^\Omega\}_{i=k+1}^m}(\mathcal{U}) \\ & \leq \mathbb{E}_{\sigma_k} \left[ \sup_{u \in \mathcal{U}} \left( \frac{1}{m} \sigma_k u(x_k^\Omega) - \frac{1}{m} \sigma_i u(x_k^{\Omega,\prime}) \right) \right] \leq \frac{2}{m} B_{\mathcal{U}}^{(0)}. \end{aligned}$$

Then according to McDiarmid's inequality (Lemma A.1), with probability at least  $1 - \tau$  the following holds:

$$\mathfrak{R}_m(\mathcal{U}) \leq \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) + 2B_{\mathcal{U}}^{(0)} \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.3)$$

According to (5.1) to (5.3), for each  $\tau \in (0, 1)$ , the following inequality holds with probability at least  $1 - 2\tau$ :

$$\phi_u^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) \leq 8|\Omega| \left( B_{\mathcal{U}}^{(0)} + \delta \right) \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) + 20|\Omega| \left( B_{\mathcal{U}}^{(0)} + \delta \right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.4)$$

By the same arguments, we have that the following inequality holds with probability at least  $1 - 2\tau$ :

$$\begin{aligned} & \|\partial_j u - z_j^\delta\|_{L^2(\Omega)}^2 - \frac{|\Omega|}{m} \sum_{i=1}^m (\partial_j u(x_i^\Omega) - z_j^\delta(x_i^\Omega))^2 \\ & \leq 8|\Omega| \left( B_{\mathcal{U}}^{(1)} + \delta \right) \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{U}) + 20|\Omega| \left( B_{\mathcal{U}}^{(1)} + \delta \right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \end{aligned}$$

Then summarizing with respect to  $j \in [d]$  and using (5.4) yield that the following holds with probability at least  $1 - 2(1+d)\tau$ :

$$\begin{aligned} & \left( \|u - y^\delta\|_{L^2(\Omega)}^2 - \|\nabla u - z^\delta\|_{L^2(\Omega)}^2 \right) \\ & - \frac{|\Omega|}{m} \sum_{i=1}^m \left( (u(x_i^\Omega) - y^\delta(x_i^\Omega))^2 + \|\nabla u(x_i^\Omega) - z^\delta(x_i^\Omega)\|_2^2 \right) \\ & \leq 8|\Omega| \left\{ \left( B_{\mathcal{U}}^{(0)} + \delta \right) \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) + d \left( B_{\mathcal{U}}^{(1)} + \delta \right) \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{U}) \right\} \\ & + 20|\Omega| \left\{ \left( B_{\mathcal{U}}^{(0)} + \delta \right)^2 + d \left( B_{\mathcal{U}}^{(1)} + \delta \right)^2 \right\} \sqrt{\frac{\log(1/\tau)}{2m}}. \end{aligned} \quad (5.5)$$

In order to estimate the generalization error corresponding to the regularization term, we define functions  $\phi_f^{(0)}, \phi_f^{(1)} : \Omega^m \rightarrow \mathbb{R}$  as

$$\begin{aligned} \phi_f^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) &= \sup_{f \in \mathcal{F}} \left( \|f\|_{L^2(\Omega)}^2 - \frac{|\Omega|}{m} \sum_{i=1}^m f(x_i^\Omega)^2 \right), \\ \phi_f^{(1)}\left(\{x_i^\Omega\}_{i=1}^m\right) &= \sup_{f \in \mathcal{F}} \left( \|\partial_j f\|_{L^2(\Omega)}^2 - \frac{|\Omega|}{m} \sum_{i=1}^m \partial_j f(x_i^\Omega)^2 \right), \end{aligned}$$

and notice that

$$\begin{aligned}\phi_f^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) - \phi_f^{(0)}\left(\{x_i^\Omega\}_{i=1}^{k-1} \cup \{x_k^{\Omega,\prime}\} \cup \{x_i^\Omega\}_{i=k+1}^m\right) &\leq \frac{|\Omega|}{m} \left(B_{\mathcal{F}}^{(0)}\right)^2, \\ \phi_f^{(1)}\left(\{x_i^\Omega\}_{i=1}^m\right) - \phi_f^{(1)}\left(\{x_i^\Omega\}_{i=1}^{k-1} \cup \{x_k^{\Omega,\prime}\} \cup \{x_i^\Omega\}_{i=k+1}^m\right) &\leq \frac{|\Omega|}{m} \left(B_{\mathcal{F}}^{(1)}\right)^2.\end{aligned}$$

As a consequent, we have that for each  $\tau \in (0, 1)$ , the following inequality holds from McDiarmid's inequality (Lemma A.1) with probability at least  $1 - \tau$ :

$$\phi_f^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) \leq \mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \left[ \phi_f^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) \right] + |\Omega| \left(B_{\mathcal{F}}^{(0)}\right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}, \quad (5.6)$$

$$\phi_f^{(1)}\left(\{x_i^\Omega\}_{i=1}^m\right) \leq \mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \left[ \phi_f^{(1)}\left(\{x_i^\Omega\}_{i=1}^m\right) \right] + |\Omega| \left(B_{\mathcal{F}}^{(1)}\right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.7)$$

Moreover, it follows from the technique of symmetrization that

$$\mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \left[ \phi_f^{(0)}\left(\{x_i^\Omega\}_{i=1}^m\right) \right] \leq 2|\Omega| \mathfrak{R}_m \left( \{x \mapsto f(x)^2 : x \in \Omega, f \in \mathcal{F}\} \right) \leq 4|\Omega| B_{\mathcal{F}}^{(0)} \mathfrak{R}_m(\mathcal{F}),$$

$$\mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \left[ \phi_f^{(1)}\left(\{x_i^\Omega\}_{i=1}^m\right) \right] \leq 2|\Omega| \mathfrak{R}_m \left( \{x \mapsto \partial_j f(x)^2 : x \in \Omega, f \in \mathcal{F}\} \right) \leq 4|\Omega| B_{\mathcal{F}}^{(1)} \mathfrak{R}_m(\partial_j \mathcal{F}),$$

where  $\partial_j \mathcal{F} = \{x \mapsto \partial_j f(x) : x \in \Omega, f \in \mathcal{F}\}$  and we used Ledoux-Talagrand contraction inequality (Lemma A.3). By combining above two inequalities with (5.6)-(5.7) and using McDiarmid's inequality (Lemma A.1) like (5.3), for each  $\tau \in (0, 1)$ , the following inequality holds with probability at least  $1 - 2(1+d)\tau$ :

$$\begin{aligned}\|f\|_{H^1(\Omega)}^2 - \frac{|\Omega|}{m} \sum_{i=1}^m \left\{ f(x_i^\Omega)^2 + \|\nabla f(x_i^\Omega)\|_2^2 \right\} \\ \leq 4|\Omega| \left\{ B_{\mathcal{F}}^{(0)} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{F}) + dB_{\mathcal{F}}^{(1)} \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_j \mathcal{F}) \right\} \\ + 9|\Omega| \left\{ \left(B_{\mathcal{F}}^{(0)}\right)^2 + d \left(B_{\mathcal{F}}^{(1)}\right)^2 \right\} \sqrt{\frac{\log(1/\tau)}{2m}}.\end{aligned} \quad (5.8)$$

It remains to investigate two physics-informed terms. Define a function  $\phi_{\text{int}} : \Omega^m \rightarrow \mathbb{R}$  by

$$\phi_{\text{int}}\left(\{x_i^\Omega\}_{i=1}^m\right) = \sup_{(u,f) \in \mathcal{U} \times \mathcal{F}} (R_{\text{int}}(u,f) - \widehat{R}_{\text{int}}(u,f)).$$

Denote by  $\bar{V} = \|V\|_{L^\infty(\Omega)}$ . It is straightforward to verify that

$$\phi_{\text{int}}\left(\{x_i^\Omega\}_{i=1}^m\right) - \phi_{\text{int}}\left(\{x_i^\Omega\}_{i=1}^{k-1} \cup \{x_k^{\Omega,\prime}\} \cup \{x_i^\Omega\}_{i=k+1}^m\right) \leq \frac{|\Omega|}{m} \left(B_{\mathcal{F}}^{(0)} + dB_{\mathcal{U}}^{(2)} + \bar{V} B_{\mathcal{U}}^{(0)}\right)^2,$$

which implies from McDiarmid's inequality (Lemma A.1) that for each  $\tau \in (0, 1)$ , the following inequality holds with probability at least  $1 - \tau$ :

$$\phi_{\text{int}}\left(\{x_i^\Omega\}_{i=1}^m\right) \leq \mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \left[ \phi_{\text{int}}\left(\{x_i^\Omega\}_{i=1}^m\right) \right] + |\Omega| \left(B_{\mathcal{F}}^{(0)} + dB_{\mathcal{U}}^{(2)} + \bar{V} B_{\mathcal{U}}^{(0)}\right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.9)$$

Denote  $h(x) = f(x) + \Delta u(x) - V(x)u(x)$ . Then  $R_{\text{int}}(u, f) = \mathbb{E}[h(x)]$ . By the similar argument mentioned above, we have that

$$\begin{aligned} & \mathbb{E}_{\{x_i^\Omega\}_{i=1}^m} \sup_{(u, f) \in \mathcal{U} \times \mathcal{F}} \left( \mathbb{E}_{x^\Omega} [h(x^\Omega)^2] - \frac{1}{m} \sum_{i=1}^m h(x_i^\Omega)^2 \right) \\ & \leq \mathbb{E}_{\{x_i^\Omega, \tilde{x}_i^\Omega\}_{i=1}^m} \sup_{(u, f) \in \mathcal{U} \times \mathcal{F}} \left( \frac{1}{m} \sum_{i=1}^m h(\tilde{x}_i^\Omega)^2 - \frac{1}{m} \sum_{i=1}^m h(x_i^\Omega)^2 \right) \\ & = \mathbb{E}_{\{x_i^\Omega, \tilde{x}_i^\Omega, \sigma_i\}_{i=1}^m} \sup_{(u, f) \in \mathcal{U} \times \mathcal{F}} \left( \frac{1}{m} \sum_{i=1}^m \sigma_i (h(\tilde{x}_i^\Omega)^2 - h(x_i^\Omega)^2) \right) \\ & = 2\mathfrak{R}_m \left( \{x \mapsto h(x)^2 = (f(x) + \Delta u(x) - V(x)u(x))^2 : x \in \Omega, u \in \mathcal{U}, f \in \mathcal{F}\} \right) \\ & \leq 4 \left( B_{\mathcal{F}}^{(0)} + dB_{\mathcal{U}}^{(2)} + \bar{V}B_{\mathcal{U}}^{(0)} \right) \left\{ \mathfrak{R}_m(\mathcal{F}) + d \max_{j \in [d]} \mathfrak{R}_m(\partial_{jj}^2 \mathcal{U}) + \bar{V} \mathfrak{R}_m(\mathcal{U}) \right\}, \end{aligned}$$

where

$$\partial_{jj}^2 \mathcal{U} = \{x \mapsto \partial_{jj}^2 u(x) : x \in \Omega, u \in \mathcal{U}\}.$$

Using McDiarmid's inequality (Lemma A.1), we have that the following inequalities hold with probability at least  $1 - \tau$ :

$$\mathfrak{R}_m(\partial_{jj}^2 \mathcal{U}) \leq \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_{jj}^2 \mathcal{U}) + 2B_{\mathcal{U}}^{(2)} \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.10)$$

Combining (5.9) with (5.10), we have that for each  $\tau \in (0, 1)$ , the following inequality holds with probability at least  $1 - (3 + d)\tau$ :

$$\begin{aligned} & R_{\text{int}}(u, f) - \widehat{R}_{\text{int}}(u, f) \\ & \leq 4|\Omega| \left( B_{\mathcal{F}}^{(0)} + dB_{\mathcal{U}}^{(2)} + \bar{V}B_{\mathcal{U}}^{(0)} \right) \left\{ \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{F}) + d \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\partial_{jj}^2 \mathcal{U}) + \bar{V} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) \right\} \\ & \quad + 9|\Omega| \left( B_{\mathcal{F}}^{(0)} + dB_{\mathcal{U}}^{(2)} + \bar{V}B_{\mathcal{U}}^{(0)} \right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \end{aligned} \quad (5.11)$$

Define a function  $\phi_{\text{bdy}} : (\partial\Omega)^m \rightarrow \mathbb{R}$  by

$$\phi_{\text{bdy}} \left( \{x_i^\Gamma\}_{i=1}^m \right) = \sup_{u \in \mathcal{U}} \left( \|\partial_n u - g\|_{L^2(\partial\Omega)}^2 - \frac{|\partial\Omega|}{m} \sum_{i=1}^m (\partial_n u(x_i^\Gamma)) - g(x_i^\Gamma)^2 \right).$$

Denote  $\bar{g} = \|g\|_{L^\infty(\partial\Omega)}$ . It is apparent that

$$\phi_{\text{bdy}} \left( \{x_i^\Gamma\}_{i=1}^m \right) - \phi_{\text{bdy}} \left( \{x_i^\Gamma\}_{i=1}^{k-1} \cup \{x_k^\Gamma\} \cup \{x_i^\Gamma\}_{i=k+1}^m \right) \leq \frac{|\partial\Omega|}{m} \left( \sqrt{dB_{\mathcal{U}}^{(1)}} + \bar{g} \right)^2,$$

consequently, for each  $\tau \in (0, 1)$ , the following holds from McDiarmid's inequality (Lemma A.1) with probability at least  $1 - \tau$ :

$$\phi_{\text{bdy}} \left( \{x_i^\Gamma\}_{i=1}^m \right) \leq \mathbb{E}_{\{x_i^\Gamma\}_{i=1}^m} [\phi_{\text{bdy}} \left( \{x_i^\Gamma\}_{i=1}^m \right)] + |\partial\Omega| \left( \sqrt{dB_{\mathcal{U}}^{(1)}} + \bar{g} \right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \quad (5.12)$$

On the other hand, we find that the following inequality holds with probability at least  $1 - \tau$ :

$$\begin{aligned} & \mathbb{E}_{\{x_i^\Gamma\}_{i=1}^m} \left[ \phi_{\text{bdy}} \left( \{x_i^\Gamma\}_{i=1}^m \right) \right] \\ & \leq 2|\partial\Omega| \mathfrak{R}_m \left( \{x \mapsto (\partial_n u(x) - g(x))^2 : x \in \partial\Omega, u \in \mathcal{U}\} \right) \\ & \leq 4|\partial\Omega| \left( \sqrt{d}B_{\mathcal{U}}^{(1)} + \bar{g} \right) \mathfrak{R}_m(\partial_n \mathcal{U}) \\ & \leq 4|\partial\Omega| \left( \sqrt{d}B_{\mathcal{U}}^{(1)} + \bar{g} \right) \widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_n \mathcal{U}) + 8|\partial\Omega| \left( \sqrt{d}B_{\mathcal{U}}^{(1)} + \bar{g} \right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}, \end{aligned}$$

where

$$\begin{aligned} \partial_n \mathcal{U} &= \{x \mapsto \partial_n u(x) : x \in \partial\Omega, u \in \mathcal{U}\}, \\ T\mathcal{U} &= \{x \mapsto Tu(x) : x \in \partial\Omega, u \in \mathcal{U}\}. \end{aligned}$$

Combining this with (5.12) gives that for each  $\tau \in (0, 1)$  the following inequality holds with probability at least  $1 - \tau$ :

$$\begin{aligned} R_{\text{bdy}}(u) - \widehat{R}_{\text{bdy}}(u) &\leq 4|\partial\Omega| \left( \sqrt{d}B_{\mathcal{U}}^{(1)} + \bar{g} \right) \widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_n \mathcal{U}) \\ &\quad + 9|\partial\Omega| \left( \sqrt{d}B_{\mathcal{U}}^{(1)} + \bar{g} \right)^2 \sqrt{\frac{\log(1/\tau)}{2m}}. \end{aligned} \quad (5.13)$$

By using (5.5), (5.8), (5.11) and (5.13), we have that for each  $\tau \in (0, 1)$ , the following holds with probability at least  $1 - (9 + 5d)\tau$ :

$$\begin{aligned} L_\lambda(\widehat{u}_\lambda^\delta, \widehat{f}_\lambda^\delta) &\leq \inf_{(\bar{u}, \bar{f}) \in \mathcal{U} \times \mathcal{F}} L_\lambda(\bar{u}, \bar{f}) + C\mathcal{E}_{\text{gen}}(\mathcal{U}, \mathcal{F}, n) \\ &\leq \inf_{(\bar{u}, \bar{f}) \in \mathcal{U} \times \mathcal{F}} \mathcal{E}_\lambda(\bar{u}, \bar{f}) + 2(1+d)\delta^2 + C\mathcal{E}_{\text{gen}}(\mathcal{U}, \mathcal{F}, n). \end{aligned} \quad (5.14)$$

**Step 2.** Approximation error. In this part, we will prove the following inequality:

$$\inf_{(\bar{u}, \bar{f}) \in \mathcal{U} \times \mathcal{F}} \mathcal{E}_\lambda(\bar{u}, \bar{f}) \leq C \inf_{u \in \mathcal{U}} \|u - u^\dagger\|_{H^2(\Omega)}^2 + (1+2\lambda) \inf_{f \in \mathcal{F}} \|f - f^\dagger\|_{H^1(\Omega)}^2 + 2\lambda \|f^\dagger\|_{H^1(\Omega)}^2, \quad (5.15)$$

combining which with (5.14) yields the desired result. Since  $(u^\dagger, f^\dagger)$  satisfies (1.1), it follows that

$$\begin{aligned} R_{\text{int}}(u, f) &= \|(f - f^\dagger) + \Delta(u - u^\dagger) - V(u - u^\dagger)\|_{L^2(\Omega)}^2 \\ &\leq \|f - f^\dagger\|_{L^2(\Omega)}^2 + (1 + \bar{V}) \|u - u^\dagger\|_{H^2(\Omega)}^2, \end{aligned}$$

and using the trace theorem gives that

$$R_{\text{bdy}}(u, f) = \|\partial_n(u - u^\dagger)\|_{L^2(\partial\Omega)} \leq C \|u - u^\dagger\|_{H^1(\Omega)}^2,$$

where  $C$  is a constant only depending on  $\Omega$ . On the other hand, we have that  $\|f\|_{H^1(\Omega)} \leq \|f - f^\dagger\|_{H^1(\Omega)} + \|f^\dagger\|_{H^1(\Omega)}$ . Combining this with above two inequalities and taking infimum with respect to  $(\bar{u}, \bar{f}) \in \mathcal{U} \times \mathcal{F}$  implies (5.15). This completes the proof.  $\square$

### 5.3. Proof of convergence rates

The proof of [15, Lemma 3.6].

**Lemma 5.1.** *Let  $\varrho = \tanh$  and  $\mathcal{G} = \mathcal{N}_\varrho(L, S, R)$ . Then it follows that*

$$B_{\mathcal{G}}^{(0)} \leq RS, \quad B_{\mathcal{G}}^{(1)} \leq R^L S^L, \quad B_{\mathcal{G}}^{(0)} \leq 2R^{2L} S^{2L},$$

and

$$\begin{aligned} \sup_{x \in \mathcal{X}} |g(x, \theta) - g(x; \theta')| &\leq \kappa^{(0)} \|\theta - \theta'\|_\infty, \\ \sup_{x \in \mathcal{X}} \max_{j \in [d]} |\partial_j g(x, \theta) - \partial_j g(x; \theta')| &\leq \kappa^{(1)} \|\theta - \theta'\|_\infty, \\ \sup_{x \in \mathcal{X}} \max_{j \in [d]} |\partial_{jj}^2 g(x, \theta) - \partial_{jj}^2 g(x; \theta')| &\leq \kappa^{(2)} \|\theta - \theta'\|_\infty, \end{aligned}$$

where  $\kappa^{(0)} = R^L S^L$ ,  $\kappa^{(1)} = 2R^{2L} S^{2L}$  and  $\kappa^{(2)} = 10R^{3L} S^{3L}$ .

*Proof of Lemma 3.1.* A direct conclusion of Lemma 5.1.  $\square$

*Proof of Lemma 3.2.* We first consider the Rademacher complexity of the function class  $\mathcal{U}$ . According to Lemma A.6, we have that

$$\begin{aligned} \sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) &\leq \epsilon + B_{\mathcal{U}}^{(0)} \sqrt{\frac{2H(\epsilon, \mathcal{U}, L^\infty(\Omega))}{m}} \\ &\leq \epsilon + B_{\mathcal{U}}^{(0)} \sqrt{\frac{2H(\epsilon/\kappa_{\mathcal{U}}^{(0)}, \Theta, \ell_\infty)}{m}} \\ &\leq \epsilon + B_{\mathcal{U}}^{(0)} \sqrt{\frac{2S_{\mathcal{U}} \log(3R_{\mathcal{U}} \kappa_{\mathcal{U}}^{(0)}/\epsilon)}{m}}, \end{aligned}$$

where the second inequality is due to Lemma A.7 and the last one holds from Lemma A.8. Combining this with Lemma 5.1 and setting  $\epsilon = B_{\mathcal{U}}^{(0)}/\sqrt{m}$  yield

$$\sup_{\{x_i^\Omega\}_{i=1}^m} \widehat{\mathfrak{R}}_{\{x_i^\Omega\}_{i=1}^m}(\mathcal{U}) \leq 3S_{\mathcal{U}} R_{\mathcal{U}} \sqrt{\frac{S_{\mathcal{U}} (L_{\mathcal{U}} \log(S_{\mathcal{U}} R_{\mathcal{U}}) + \log m)}{m}}.$$

By a same argument, we can obtain Rademacher complexities of other function classes. One more thing we need to note is that for each fixed  $\{x_i^\Gamma\}_{i=1}^m$ , it holds from Cauchy-Scharwz inequality that

$$\widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_n \mathcal{U}) \leq \|n\|_1 \max_{j \in [d]} \widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_j \mathcal{U}) \leq \sqrt{d} \widehat{\mathfrak{R}}_{\{x_i^\Gamma\}_{i=1}^m}(\partial_j \mathcal{U}).$$

This completes the proof.  $\square$

*Proof of Theorem 3.2.* According to Lemma 1.1, it holds that

$$\begin{aligned} \inf_{u \in \mathcal{U}} \|u - u^\dagger\|_{H^2(\Omega)} &\leq |\Omega| \|u^\dagger\|_{H^{\alpha+2}(\Omega)}^2 \varepsilon^2, \\ \inf_{f \in \mathcal{F}} \|f - f^\dagger\|_{H^1(\Omega)} &\leq |\Omega| \|f^\dagger\|_{H^{\alpha+1}(\Omega)}^2 \varepsilon^2, \end{aligned} \quad (5.16)$$

where

$$\mathcal{U} = \mathcal{N}_\varrho(L_{\mathcal{U}}, S_{\mathcal{U}}, R_{\mathcal{U}}) \quad \text{and} \quad \mathcal{F} = \mathcal{N}_\varrho(L_{\mathcal{F}}, S_{\mathcal{F}}, R_{\mathcal{F}})$$

with

$$\begin{aligned} L_{\mathcal{U}} &= c \log(d+2), \quad S_{\mathcal{U}} = C \varepsilon^{-d/(\alpha-\mu)}, \quad R_{\mathcal{U}} = C \varepsilon^{-(9d+4+4\mu)/(2(\alpha-\mu))-2}, \\ L_{\mathcal{F}} &= c \log(d+1), \quad S_{\mathcal{F}} = C \varepsilon^{-d/(\alpha-\mu)}, \quad R_{\mathcal{F}} = C \varepsilon^{-(9d+2+4\mu)/(2(\alpha-\mu))-2}. \end{aligned} \quad (5.17)$$

Here  $c, \mu > 0$  are absolute constants and the constant  $C$  depends on  $\alpha, d, \Omega$  and  $\mu$ . On the other hand, we find from Lemma 3.1, 3.2 and Corollary 3.1 that

$$\begin{aligned} \mathcal{E}_{\text{gen}}(\mathcal{U}, \mathcal{F}, m) &\leq C \left\{ R_{\mathcal{U}}^{4L} S_{\mathcal{U}}^{4L} + R_{\mathcal{F}}^{2L} S_{\mathcal{F}}^{2L} + \delta^2 (\log m)^2 \right\} \sqrt{S_{\mathcal{U}} L_{\mathcal{U}} \log(S_{\mathcal{U}} R_{\mathcal{U}}) \frac{\log m}{m}} \\ &\quad + C \left\{ R_{\mathcal{U}}^{4L} S_{\mathcal{U}}^{4L} + R_{\mathcal{F}}^{2L} S_{\mathcal{F}}^{2L} + \delta^2 (\log m)^2 \right\} \sqrt{S_{\mathcal{F}} L_{\mathcal{F}} \log(S_{\mathcal{F}} R_{\mathcal{F}}) \frac{\log m}{m}} \\ &\quad + C \left\{ R_{\mathcal{U}}^{4L} S_{\mathcal{U}}^{4L} + R_{\mathcal{F}}^{2L} S_{\mathcal{F}}^{2L} + \delta^2 \right\} \sqrt{\frac{\log(1/\tau)}{2m}}, \end{aligned} \quad (5.18)$$

where  $C$  is a positive constant only depending on  $d, \Omega, \|V\|_{L^\infty(\Omega)}$  and  $\|g\|_{L^\infty(\partial\Omega)}$ . Plugging (5.17) in (5.18) yields that

$$\begin{aligned} \mathcal{E}_{\text{gen}}(\mathcal{U}, \mathcal{F}, m) &\leq C \left\{ \varepsilon^{-c((d+1)/(\alpha-\mu)) \log(d+2)} + \delta^2 (\log m)^2 \right\} \\ &\quad \times \left( \varepsilon^{-d/(2(\alpha-\mu))} \sqrt{\frac{\log m}{m}} + \sqrt{\frac{\log(1/\tau)}{2m}} \right). \end{aligned} \quad (5.19)$$

Combining (5.16) with (5.19) yields the desired result.  $\square$

## 6. Conclusions and Discussions

We have proposed a method based on Physics-Informed Neural Networks (PINNs) for identifying sources in elliptic equations. By employing neural networks to approximate the unknown source term and solution, and minimizing the empirical risk using a stochastic gradient descent (SGD)-type algorithm, we are able to recover the source density and solution. To assess the effectiveness of our method experimentally, we apply it to various examples with different noise levels and regularization parameters. These experiments highlight the remarkable robustness of our method when encountering data noise. Furthermore, we provide theoretical analysis by presenting convergence rates of the reconstruction

in relation to the noise level. This analysis yields valuable insights a priori for determining appropriate choices for regularization parameters, the number of samples, and the size of the neural networks.

The proposed method is applicable to high-dimensional problems and can be integrated with adaptive sampling methods, as described in [33]. Moreover, the approach of combining Tikhonov regularization with PINNs can be extended to address parameter identification problems involving partial differential equations other than the one considered in this study. The analytical strategies developed in this work hold great promise for establishing convergence rates in future investigations of various inverse problems.

## Appendix A. Supplementary Definitions and Lemmas

**Definition A.1** (Gaussian Complexity). *Let  $\mathcal{X} \subseteq \mathbb{R}^d$  ( $d \geq 1$ ) be a bounded domain and let  $\mu$  be a measure on  $\mathcal{X}$ . Suppose that  $\mathcal{G}$  is a family of functions mapping from  $\mathcal{X}$  to  $\mathbb{R}$  and  $\{x_i\}_{i=1}^m$  is a set of samples i.i.d. drawn from  $\mu$ . Let  $\sigma = \{\sigma_i\}_{i=1}^m$  be a set of i.i.d. standard Gaussian variables and independent of  $\{x_i\}_{i=1}^m$ . Then the empirical Gaussian complexity of  $\mathcal{G}$  with respect to the sample set  $\{x_i\}_{i=1}^m$  is defined as*

$$\widehat{\mathfrak{G}}_{\{x_i\}_{i=1}^m}(\mathcal{G}) = \mathbb{E}_\sigma \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^m \sigma_i g(x_i) \right],$$

where  $\mathbb{E}_\sigma[\cdot] = \mathbb{E}[\cdot | \{x_i\}_{i=1}^m]$ .

**Definition A.2** (Covering Number and Metric Entropy). *Let  $(\mathcal{H}, \rho(\cdot, \cdot))$  be a pseudo-metric space and  $\mathcal{G} \subseteq \mathcal{H}$ . A set  $\mathcal{G}_\epsilon \subseteq \mathcal{H}$  is called a  $\rho(\cdot, \cdot)$   $\epsilon$ -cover of  $\mathcal{G}$  if for each  $g \in \mathcal{G}$ , there exists  $g_\epsilon \in \mathcal{G}_\epsilon$ , such that  $\rho(g, g_\epsilon) \leq \epsilon$ . Then the  $\epsilon$ -covering number of  $\mathcal{G}$  is defined by*

$$N(\epsilon, \mathcal{G}, \rho(\cdot, \cdot)) = \min \{|\mathcal{G}_\epsilon| : \mathcal{G}_\epsilon \text{ is a } \rho(\cdot, \cdot) \text{ } \tau\text{-cover of } \mathcal{G}\}.$$

Moreover, the  $\rho(\cdot, \cdot)$   $\epsilon$ -metric entropy of  $\mathcal{G}$  is defined by  $H(\epsilon, \mathcal{G}, \rho(\cdot, \cdot)) = \log N(\epsilon, \mathcal{G}, \rho(\cdot, \cdot))$ .

**Lemma A.1** (McDiarmid's Inequality, cf. Mohri & Rostamizadeh [51, Theorem D.8]). *Let  $\{x_i\}_{i=1}^m \subseteq \mathcal{X}$  be a set of  $m$  independent random variables and assume that there exists  $\{c_i\}_{i=1}^m \subseteq \mathbb{R}_+$  such that  $g : \mathcal{X}^m \rightarrow \mathbb{R}$  satisfies the following condition:*

$$|f(z_1, \dots, z_k, \dots, z_m) - f(z_1, \dots, z'_k, \dots, z_m)| \leq c_i,$$

for each  $k \in [m]$  and each points  $z_1, \dots, z_m, z'_k \in \mathcal{X}$ . Then for each  $\delta > 0$ , the following inequalities hold:

$$\begin{aligned} \Pr(f(x_1, \dots, x_m) - \mathbb{E}[f(x_1, \dots, x_m)] \geq \delta) &\leq \exp\left(-\frac{2\delta^2}{\sum_{i=1}^m c_i^2}\right), \\ \Pr(f(x_1, \dots, x_m) - \mathbb{E}[f(x_1, \dots, x_m)] \leq -\delta) &\leq \exp\left(-\frac{2\delta^2}{\sum_{i=1}^m c_i^2}\right). \end{aligned}$$

**Lemma A.2** (cf. Giné & Nickl [23, Theorem 2.1.20]). *Let  $\{X_t : t \in T\}$  be a separable centred Gaussian process such that  $\Pr(\sup_{t \in T} |X_t| < \infty) > 0$ . Then  $\sigma := \sup_{t \in T} (\mathbb{E} X_t^2)^{1/2} < \infty$  and  $\mathbb{E}[\sup_{t \in T} |X_t|] < \infty$ . Furthermore, the following inequality holds:*

$$\Pr\left(\left|\sup_{t \in T} |X_t| - \mathbb{E}[\sup_{t \in T} |X_t|]\right| > \delta\right) \leq 2 \exp\left(-\frac{\delta^2}{2\sigma^2}\right).$$

**Lemma A.3** (Ledoux-Talagrand Contraction Inequality, cf. Mohri & Rostamizadeh [51, Lemma 5.7]). *Let  $\{\phi_i\}_{i=1}^m$  be  $m$   $L$ -Lipschitz functions from  $\mathbb{R}$  to  $\mathbb{R}$ , and  $\{\sigma_i\}_{i=1}^m$  be Rademacher random variables. Then for any hypothesis set  $\mathcal{G}$  of real-valued functions, the following inequality holds:*

$$\mathbb{E}_{\{\sigma_i\}_{i=1}^m} \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^m \sigma_i \phi_i \circ g(x_i) \right] \leq L \mathbb{E}_{\{\sigma_i\}_{i=1}^m} \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^m \sigma_i g(x_i) \right] = L \widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G}).$$

**Lemma A.4** (cf. Bartlett & Mendelson [12, Lemma 1]). *Let  $\mathcal{X} \subseteq \mathbb{R}^d$  and  $\{x_i\}_{i=1}^m \subseteq \mathcal{X}$ . Let  $\mathcal{G}$  be a finite set of functions mapping from  $\mathcal{X}$  to  $\mathbb{R}$ . Then it follows that  $\widehat{\mathfrak{G}}_{\{x_i\}_{i=1}^m}(\mathcal{G}) \leq 2\sqrt{\log m} \widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G})$ .*

**Lemma A.5** (Massart's Lemma, cf. Mohri & Rostamizadeh [51, Theorem 3.7]). *Let  $\mathcal{X} \subseteq \mathbb{R}^d$  and  $\{x_i\}_{i=1}^m \subseteq \mathcal{X}$ . Let  $\mathcal{G}$  be a finite set of functions mapping from  $\mathcal{X}$  to  $\mathbb{R}$ . Suppose that  $\|g\|_{L^\infty(\mathcal{X})} \leq B_\mathcal{G}$  for each  $g \in \mathcal{G}$ . Then it follows that*

$$\widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G}) \leq B_\mathcal{G} \sqrt{\frac{2 \log |\mathcal{G}|}{m}}.$$

**Lemma A.6** (Dudley Inequality). *Let  $\mathcal{X} \subseteq \mathbb{R}^d$  and  $\{x_i\}_{i=1}^m \subseteq \mathcal{X}$ . Let  $\mathcal{G}$  be a set of functions mapping from  $\mathcal{X}$  to  $\mathbb{R}$ . Suppose that  $\|g\|_{L^\infty(\mathcal{X})} \leq B_\mathcal{G}$  for each  $g \in \mathcal{G}$ . Then it follows that*

$$\widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G}) \leq \epsilon + B_\mathcal{G} \sqrt{\frac{2H(\epsilon, \mathcal{G}, L^\infty(\mathcal{X}))}{m}}.$$

*Proof.* Let  $\mathcal{G}_\epsilon$  be an  $L^\infty(\mathcal{X})$   $\epsilon$ -cover of  $\mathcal{G}$ , that is, for each  $g \in \mathcal{G}$ , there exists  $g_\epsilon \in \mathcal{G}_\epsilon$ , such that  $\|g - g_\epsilon\|_{L^\infty(\mathcal{X})} \leq \epsilon$ . Further, suppose that  $|\mathcal{G}_\epsilon| = N(\epsilon, \mathcal{G}, L^\infty(\mathcal{X}))$ . Then it follows from the convexity of suprema that

$$\begin{aligned} \widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G}) &\leq \mathbb{E}_\sigma \left[ \sup_{g \in \mathcal{G}} \frac{1}{m} \sum_{i=1}^m \sigma_i (g(x_i) - g_\epsilon(x_i)) \right] + \mathbb{E}_\sigma \left[ \sup_{g_\epsilon \in \mathcal{G}_\epsilon} \frac{1}{m} \sum_{i=1}^m \sigma_i g_\epsilon(x_i) \right] \\ &\leq \epsilon + \widehat{\mathfrak{R}}_{\{x_i\}_{i=1}^m}(\mathcal{G}_\epsilon), \end{aligned}$$

where the last inequality holds from Hölder's inequality. According to Massart's lemma (Lemma A.5), we completes the proof.  $\square$

The following lemma shows that the covering numbers of parameterized classes of functions that are Lipschitz in the parameter can be controlled by the covering numbers of the parameter space.

**Lemma A.7** (Metric Entropy of Parametric Classes). *Let  $\mathcal{G}$  be a parameterized class of functions, that is,  $\mathcal{G} = \{x \mapsto g(x; \theta) : x \in \mathcal{X}, \theta \in \Theta\}$ . Suppose that there exists a positive constant  $\kappa$ , such that  $|g(x; \theta) - g(x; \theta')| \leq \kappa \|\theta - \theta'\|_\infty$ . Then it follows that*

$$H(\epsilon, \mathcal{G}, L^\infty(\mathcal{X})) \leq H(\epsilon/\kappa, \Theta, \ell_\infty).$$

*Proof.* Let  $\Theta_\delta$  be an  $\ell_\infty$   $\delta$ -cover of  $\Theta$  with  $|\Theta_\delta| = N(\delta, \Theta, \ell_\infty)$ . Then for each  $\theta \in \Theta$ , there exists  $\theta_\delta \in \Theta_\delta$  such that  $\|\theta - \theta_\delta\|_\infty \leq \delta$ . Denote by  $\mathcal{G}_\delta = \{x \mapsto g(x; \theta) : x \in \mathcal{X}, \theta \in \Theta_\delta\}$ . It is apparent that  $|\mathcal{G}_\delta| = |\Theta_\delta|$ . Then for each  $\theta \in \Theta$ , there exists  $\theta_\delta \in \Theta_\delta$ , such that

$$|g(x; \theta) - g(x; \theta_\delta)| \leq \kappa \delta,$$

which means that  $\mathcal{G}_\delta$  is an  $L^\infty(\mathcal{X})$   $\kappa \delta$ -cover of  $\mathcal{G}$ , and thus  $N(\kappa \delta, \mathcal{G}, L^\infty(\mathcal{X})) \leq |\mathcal{G}_\delta|$ . Setting  $\epsilon = \kappa \delta$  and taking logarithm on both sides of the inequality yields the desired result.  $\square$

**Lemma A.8** (Metric Entropy of Finite-Dimensional Norm-Balls, [59, Lemma 5.7]). *Let  $B_R$  be a ball with respect to the metric  $\|\cdot\|$  with radius  $R$ , that is,  $B_R = \{x \in \mathbb{R}^S : \|x\| \leq R\}$ . Then it holds that  $H(\epsilon, B_R, \|\cdot\|) \leq S \log(3R/\epsilon)$ .*

### Acknowledgments

This work is supported by the National Key Research and Development Program of China (Grant No. 2020YFA0714200), by the National Nature Science Foundation of China (Grant Nos. 12125103, 12071362, 12371441), and by the Fundamental Research Funds for the Central Universities. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

### References

- [1] V. Akcelik, G. Biros, O. Ghattas, K.R. Long and B. van Bloemen Waanders, *A variational finite element method for source inversion for convective-diffusive transport*, Finite Elem. Anal. Des. **39**, 683–705 (2003).
- [2] M.A. Anastasio, J. Zhang, D. Modgil and P.J.L. Rivière, *Application of inverse source concepts to photoacoustic tomography*, Inverse Problems **23**, S21 (2007).
- [3] M. Anthony and P.L. Bartlett, *Neural Network Learning: Theoretical Foundations*, Cambridge University Press (1999).
- [4] J. Atmadja and A.C. Bagtzoglou, *State of the art report on mathematical methods for groundwater pollution source identification*, Environ. Forensics **2**, 205–214 (2001).
- [5] A.E. Badia and T. Ha-Duong, *On an inverse source problem for the heat equation. Application to a pollution detection problem*, J. Inverse Ill-Posed Probl. **10**, 585–599 (2002).
- [6] A.E. Badia and A.E. Hajj, *Identification of dislocations in materials from boundary measurements*, SIAM J. Appl. Math. **73**, 84–103 (2013).
- [7] A.E. Badia, A.E. Hajj, M. Jazar and H. Moustafa, *Lipschitz stability estimates for an inverse*

- source problem in an elliptic equation from interior measurements*, Appl. Anal. **95**, 1873–1890 (2016).
- [8] A.E. Badia and T. Nara, *An inverse source problem for Helmholtz's equation from the Cauchy data with a single wave number*, Inverse Problems **27**, 105001 (2011).
  - [9] G. Bao, X. Ye, Y. Zang and H. Zhou, *Numerical solution of inverse problems by weak adversarial networks*, Inverse Problems **36**, 115003 (2020).
  - [10] M. Barati Moghaddam, M. Mazaheri and J. Mohammad Vali Samani, *Inverse modeling of contaminant transport for pollution source identification in surface and groundwaters: A review*, Groundwater Sustainable Dev. **15**, 100651 (2021).
  - [11] A. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory **39**, 930–945 (1993).
  - [12] P.L. Bartlett and S. Mendelson, *Rademacher and Gaussian complexities: Risk bounds and structural results*, J. Mach. Learn. Res. **3**, 463–482 (2003).
  - [13] B. Bauer and M. Kohler, *On deep learning as a remedy for the curse of dimensionality in nonparametric regression*, Ann. Statist. **47**, 2261–2285 (2019).
  - [14] M. Benning and M. Burger, *Modern regularization methods for inverse problems*, Acta Numer. **27**, 1–111 (2018).
  - [15] S. Cen, B. Jin, Q. Quan and Z. Zhou, *Hybrid neural-network FEM approximation of diffusion coefficient in elliptic and parabolic problems*, IMA J. Numer. Anal., drad073 (2023).
  - [16] M. Chen, H. Jiang, W. Liao and T. Zhao, *Nonparametric regression on low-dimensional manifolds using deep ReLU networks: Function approximation and statistical recovery*, Inf. Inference **11**, 1203–1253 (2022).
  - [17] A.R. Costall, B.D. Harris, B. Teo, R. Schaa, F.M. Wagner and J.P. Pigois, *Groundwater throughflow and seawater intrusion in high quality coastal aquifers*, Sci. Rep. **10**, (2020).
  - [18] Z. Ding, C. Duan, Y. Jiao and J.Z. Yang, *Semi-supervised deep Sobolev regression: Estimation, variable selection and beyond*, arXiv:2401.04535 (2024).
  - [19] C. Duan, Y. Jiao, X. Lu and J.Z. Yang, *Current density impedance imaging with PINNs*, arXiv:2306.13881 (2023).
  - [20] W. E, C. Ma and L. Wu, *The Barron space and the flow-induced function spaces for neural network models*, Constr. Approx. **55**, 369–406 (2022).
  - [21] H.W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and Its Applications, **375**, Springer Dordrecht (2000).
  - [22] M.H. Farrell, T. Liang and S. Misra, *Deep neural networks for estimation and inference*, Econometrica **89**, 181–213 (2021).
  - [23] E. Giné and R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press (2015).
  - [24] I. Gühring and M. Raslan, *Approximation rates for neural networks with encodable weights in smoothness spaces*, Neural Networks **134**, 107–130 (2021).
  - [25] L. Guo, H. Wu, X. Yu and T. Zhou, *Monte Carlo fPINNs: Deep learning method for forward and inverse problems involving high dimensional fractional partial differential equations*, Comput. Methods Appl. Mech. Engrg. **400**, 115523 (2022).
  - [26] M. Hämäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila and O.V. Lounasmaa, *Magnetoencephalography – Theory, instrumentation, and applications to noninvasive studies of the working human brain*, Rev. Mod. Phys. **65**, 413–497 (1993).
  - [27] A. Hamdi, *Identification of a time-varying point source in a system of two coupled linear diffusion-advection-reaction equations: Application to surface water pollution*, Inverse Problems **25**, 115009 (2009).
  - [28] Q. Hu, S. Shu and J. Zou, *A new variational approach for inverse source problems*, Numer. Math.

- Theory Methods Appl. **12**, 331–347 (2018).
- [29] V. Isakov, *Inverse Problems for Partial Differential Equations*, Applied Mathematical Sciences, Vol. 127, (2018).
  - [30] K. Ito and B. Jin, *Inverse Problems: Tikhonov Theory and Algorithms*, World Scientific (2014).
  - [31] Y. Jiao, Y. Lai, D. Li, X. Lu, F. Wang, Y. Wang and J.Z. Yang, *A rate of convergence of physics informed neural networks for the linear second order elliptic PDEs*, Commun. Comput. Phys. **31**, 1272–1295 (2022).
  - [32] Y. Jiao, Y. Lai, Y. Lo, Y. Wang and Y. Yang, *Error analysis of deep Ritz methods for elliptic equations*, Anal. Appl. (Singap.) **22**(1), 57–87 (2024).
  - [33] Y. Jiao, D. Li, X. Lu, J.Z. Yang and C. Yuan, *GAS: A Gaussian mixture distribution-based adaptive sampling method for PINNs*, arXiv:2303.15849 (2023).
  - [34] Y. Jiao, G. Shen, Y. Lin and J. Huang, *Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors*, Ann. Statist. **51**, 691–716 (2023).
  - [35] B. Jin, X. Li and X. Lu, *Imaging conductivity from current density magnitude using neural networks*, Inverse Problems **38**, 075003 (2022).
  - [36] B. Jin, X. Li, Q. Quan and Z. Zhou, *Conductivity imaging from internal measurements with mixed least-squares deep neural networks*, SIAM J. Imaging Sci. **17**, 147–187 (2024).
  - [37] B. Jin, R. Sau, L. Yin and Z. Zhou, *Solving elliptic optimal control problems using physics informed neural networks*, arXiv:2308.11925 (2023).
  - [38] K. Kaiboriboon, H.O. Lüders, M. Hamaneh, J. Turnbull and S.D. Lhatoo, *EEG source imaging in epilepsy-practicalities and pitfalls*, Nat. Rev. Neurol. **8**, 498–507 (2012).
  - [39] Y.L. Keung and J. Zou, *Numerical identifications of parameters in parabolic systems*, Inverse Problems **14**, 83 (1998).
  - [40] Y.L. Keung and J. Zou, *An efficient linear solver for nonlinear parameter identification problems*, SIAM J. Sci. Comput. **22**, 1511–1526 (2001).
  - [41] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, in: 3rd International Conference on Learning Representations, ICLR 2015, Y. Bengio and Y. LeCun (Eds), Conference Track Proceedings (2015).
  - [42] M. Kohler, A. Krzyzak and S. Langer, *Estimation of a function of low local dimensionality by deep neural networks*, IEEE Trans. Inform. Theory **68**, 4032–4042 (2022).
  - [43] M. Kohler and S. Langer, *On the rate of convergence of fully connected deep neural network regression estimates*, Ann. Statist. **49**, 2231–2249 (2021).
  - [44] R.V. Kohn and B.D. Lowe, *A variational method for parameter identification*, ESAIM Math. Model. Numer. Anal. **22**, 119–158 (1988).
  - [45] X. Liu and Z. Zhai, *Inverse modeling methods for indoor airborne pollutant tracking: Literature review and fundamentals*, Indoor Air **17**, 419–438 (2007).
  - [46] J. Lu and Y. Lu, *A priori generalization error analysis of two-layer neural networks for solving high dimensional Schrödinger eigenvalue problems*, Comm. Amer. Math. Soc. **3**, 1–21 (2022).
  - [47] L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, S.G. Johnson, *Physics-informed neural networks with hard constraints for inverse design*, SIAM J. Sci. Comput. **43**, B1105–B1132 (2021).
  - [48] Y. Lu, H. Chen, J. Lu, L. Ying and J. Blanchet, *Machine learning for elliptic PDEs: Fast rate generalization bound, neural scaling law and minimax optimality*, in: International Conference on Learning Representations (2022).
  - [49] Y. Lu, J. Lu and M. Wang, *A priori generalization analysis of the deep Ritz method for solving high dimensional elliptic partial differential equations*, in: Proceedings of Thirty Fourth Conference on Learning Theory, M. Belkin and S. Kpotufe (Eds), Proceedings of Machine Learning Research, **134**, 3196–3241 (2021).
  - [50] S. Mishra and R. Molinaro, *Estimates on the generalization error of physics-informed neural*

- networks for approximating a class of inverse problems for PDEs*, IMA J. Numer. Anal. **42**, 981–1022 (2021).
- [51] M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning*, MIT Press (2018).
  - [52] R. Nakada and M. Imaizumi, *Adaptive approximation and generalization of deep neural network with intrinsic dimensionality*, J. Mach. Learn. Res. **21**, 1–38 (2020).
  - [53] G. Pang, L. Lu and G.E. Karniadakis, *fPINNs: Fractional Physics-Informed Neural Networks*, SIAM J. Sci. Comput. **41**, A2603–A2626 (2019).
  - [54] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, *PyTorch: An imperative style, high-performance deep learning library*, in: Proceedings of the 33rd International Conference on Neural Information Processing Systems (2019).
  - [55] M. Raissi, P. Perdikaris and G. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys. **378**, 686–707 (2019).
  - [56] A.A. Samarskii and P.N. Vabishchevich, *Numerical Methods for Solving Inverse Problems of Mathematical Physics*, De Gruyter (2007).
  - [57] J. Schmidt-Hieber, *Nonparametric regression using deep neural networks with ReLU activation function*, Ann. Statist. **48**, 1875–1897 (2020).
  - [58] T.H. Skaggs and Z.J. Kabala, *Recovering the release history of a groundwater contaminant*, Water Resour. Res. **30**, 71–79 (1994).
  - [59] M.J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press (2019).
  - [60] G. Wang, Y. Li and M. Jiang, *Uniqueness theorems in bioluminescence tomography*, Med. Phys. **31**, 2289–2299 (2004).
  - [61] J.C.-F. Wong and P. Yuan, *A FE-based algorithm for the inverse natural convection problem*, Internat. J. Numer. Methods Fluids **68**, 48–82 (2012).
  - [62] H. Zhang and J. Liu, *On the recovery of internal source for an elliptic system by neural network approximation*, J. Inverse Ill-Posed Probl. **31**, 739–761 (2023).
  - [63] H. Zhang and J. Liu, *Solving an inverse source problem by deep neural network method with convergence and error analysis*, Inverse Problems **39**, 075013 (2023).
  - [64] M. Zhang, Q. Li and J. Liu, *On stability and regularization for data-driven solution of parabolic inverse source problems*, J. Comput. Phys. **474**, 111769 (2023).