SCIENCE CHINA
Mathematics

• ARTICLES • https://doi.org/Science China Mathematics Manuscript for review

Generalization Error Estimates of A Machine Learning Method for Solving High-Dimensional Schrödinger Eigenvalue Problems

Yixiao Guo[1,2], Pingbing Ming[1,2,*] & Hao Yu[1,2]

[1] *SKLMS, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China;*
[2] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, China*

Email: guoyixiao@lsec.cc.ac.cn, mpb@lsec.cc.ac.cn, yuhao@amss.ac.cn

Received ; accepted

Abstract We propose a machine learning method for computing the eigenvalues and eigenfunctions of the Schrödinger operator on a d-dimensional hypercube with Dirichlet boundary conditions. The eigenpairs lie deep within the spectrum. The cut-off function technique is employed to construct trial functions that precisely satisfy the Dirichlet boundary condition. This approach outperforms the standard boundary penalty method, as demonstrated by numerical examples. Assuming that the eigenfunctions belong to a new spectral Barron space, we derive a dimension-free convergence rate $\mathcal{O}(n^{-1/4})$ for the generalization error bound, with all constants in the error bounds being explicit and growing at most polynomially. This assumption is validated by proving a new regularity shift result for the eigenfunctions when the potential belongs to an appropriate spectral Barron space. Moreover, we extend the generalization error bound to the normalized penalty method, which is widely used in practice.

Keywords generalization error, Schrödinger operator, Dirichlet boundary condition, spectral Barron space, curse of dimensionality

MSC(2020) 68T07, 65N12, 65D15, 35J10, 60G50

Citation: Guo Y, Ming P, Yu H. Generalization Error Estimates of A Machine Learning Method for Solving High-Dimensional Schrödinger Eigenvalue Problems. Sci China Math, 2022, 65, https://doi.org/Science China Mathematics Manuscript for review

1 Introduction

The high-dimensional Schrödinger eigenvalue problem plays a crucial role in various fields, such as computational chemistry, condensed matter physics, and quantum computing [48, 54]. While traditional numerical methods have achieved significant success in solving low-dimensional partial differential equations (PDEs) and eigenvalue problems, the curse of dimensionality (CoD) remains a major challenge, as computational costs increase exponentially with dimensionality. Recently, machine learning has emerged as a promising approach to mitigate the CoD. Significant progress has been made in applying

* Corresponding author

deep learning-based methods to solve PDEs [19, 21, 32, 56, 60, 73] and the Schrödinger eigenvalue problems [11, 13, 15, 23, 33–35, 44, 66], among many others.

This study aims to develop a high-precision machine learning method for solving Schrödinger eigenvalue problems and to analyze its generalization error bound. One of the main challenges in using neural networks to solve PDEs and eigenvalue problems is accurately handling the essential boundary conditions. One common approach to address this challenge is to incorporate a boundary penalty term into the loss function [21, 38]. However, recent studies [18, 41, 50, 62, 65] have demonstrated that inaccurately imposing boundary conditions can adversely affect network training and accuracy. The authors in [38, 70] have shown that the error caused by the boundary penalty is inversely proportional to the penalty factor. To mitigate this issue, we adopt the approach proposed in [9, 42, 43, 62], which utilizes the product of neural network outputs and cut-off functions. This ensures that the *ansatz* functions precisely satisfy the boundary conditions.

To solve the Schrödinger eigenvalue problem, we use the Rayleigh quotient as the loss function. However, this loss function is not Lipschitz continuous, primarily due to the denominator, which involves the square of the L^2 norm of the trial functions. This introduces new challenges in deriving generalization bounds. To overcome this issue, we leverage concentration inequalities for ratio type empirical processes [26, 27] and bounds for expected values of sup-norms of empirical processes [25]. These inequalities have proven to be crucial in bounding the generalization error of learning algorithms.

To derive the generalization error bound, we will work within the Barron-type spaces, first introduced by Barron in his seminal work [6], as Barron functions achieve dimension-free approximation rates. Such spaces have been further developed in recent studies [20, 45, 47, 59]. In this work, we introduce a new spectral Barron space, denoted as $\mathfrak{B}^s(\Omega)$, defined on the unit hypercube Ω, which is particularly suitable for studying Dirichlet boundary value problems. This space, referred to as the sine spectral Barron space, can be viewed as a variant of the cosine spectral Barron space $\mathfrak{C}^s(\Omega)$ proposed in [46, 47]. We establish a new regularity theory for the Dirichlet eigenvalue problem for Schrödinger operators in $\mathfrak{B}^s(\Omega)$. Notably, the functions within $\mathfrak{B}^s(\Omega)$ also admit dimension-free approximation rates with two-layer neural networks.

The following is an informal version of the main generalization theorem; cf. Theorem 3.4.

Theorem 1.1. *Under Assumptions 2.1 and 3.3, let* $\mathcal{F} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(\|u^*\|_{\mathfrak{B}^s(\Omega)})$ *with* $\tau = \sqrt{m}$. *Let* u_n^m *be a minimizer of the empirical loss* $L_{k,n}$ *over* $\mathcal{F}$. *If* n *and* m *are large enough, then with probability at least* $1-\delta$, *there holds*

$$L_k(u_n^m) - \lambda_k \lesssim \left[\sqrt{\frac{m\,(k+\ln(m/\delta))}{n}} + \frac{1}{\sqrt{m}}\right] \|u^*\|_{\mathfrak{B}^s(\Omega)}.$$

1.1 Our contributions

Our contributions are summarized as follows:

1. We introduce a new spectral Barron space $\mathfrak{B}^s(\Omega)$ on $\Omega = (0,1)^d$, particularly suited for investigating Dirichlet boundary value problems, which can be viewed as a homogeneous version of $\mathfrak{C}^s(\Omega)$ introduced in [46]. We prove a novel regularity estimate in Theorem 3.8 by showing that all eigenfunctions lie in $\mathfrak{B}^{s+2}(\Omega)$ if the potential function belongs to $\mathfrak{C}^s(\Omega)$ with $s \geqslant 0$.

2. We present the cut-off function technique to construct trial functions that satisfy the essential boundary conditions. We show that functions in the sine spectral Barron space can be well approximated in the H^1-norm using two-layer ReLU or Softplus networks multiplied by a cut-off function; see Theorem 3.1 and Theorem 3.2. The approximation rate is $\mathcal{O}(m^{-1/2})$ with m denoting the number of neurons, which is dimension free.

3. We introduce concentration inequalities for ratio-type suprema to handle the Rayleigh quotient and derive an oracle inequality for the empirical loss. An exponential inequality is established to bound the normalized complexity measure in ratio-type concentration inequalities.

4. We prove a priori generalization error bounds for the proposed learning-based method; see Theorem 3.4. These bounds hold for higher-order eigenmodes, not just the ground state, and the convergence rate $\mathcal{O}(n^{-1/4})$ is independent of dimensionality. We also clarify the dependence of the prefactors on the relevant parameters, including the order of the eigenmodes and the dimensionality, with the dependence being polynomial and of lower degree.

5. We demonstrate the effectiveness of incorporating a normalized penalty term by proving, in Theorem 3.9, that the solutions of the normalized penalty method [21] are away from zero with high probability. The generalization error bounds for the normalization penalty method are established in Corollary 3.11. We also characterize the cumulative error in Theorem 3.13 and prove the sharp cumulative rate of generalization error in Proposition 3.14.

6. We test the method numerically in various scenarios, including problems posed on hypercubes up to $d = 10$, unit spheres, three-dimensional rings and problems involving inverse square potentials. The method achieves a relative eigenvalue error as low as $\mathcal{O}(10^{-3})$, or even better, for the first 30 eigenvalues. Furthermore, the numerical results highlight that exactly enforcing the boundary condition improves accuracy by an order of magnitude or more compared with the boundary penalty method.

1.2 Related works

Many attempts have been made to understand why learning-based methods can overcome the CoD. In [47, 49,58,70], the authors established a priori generalization error bounds for solving elliptic PDEs using two-layer neural networks, demonstrating a dimension-independent convergence rate. Similar generalization bounds have been proven for Black-Scholes PDEs and high-dimensional nonlinear heat equations, as shown in [10,28–30,37]. Despite these advancements, the analysis of high-dimensional eigenvalue problems remains limited, with notable exceptions in [46] and [38].

In [46], the authors proposed a machine learning method for solving the Neumann eigenvalue problem associated with the Schrödinger operator. They introduced a spectral Barron space $\mathfrak{C}^s(\Omega)$ and demonstrated that functions within this space may be well approximated in the H^1-norm using two-layer ReLU or Softplus networks. Moreover, they established the existence of the ground state in $\mathfrak{C}^{s+2}(\Omega)$ when the potential functions reside in $\mathfrak{C}^s(\Omega)$ with $s \geqslant 0$. In the present work, we extend this approach by using two-layer ReLU or Softplus networks, multiplied by a cut-off function, to approximate functions within $\mathfrak{B}^s(\Omega)$. The approximation rate in [46] is further improved by eliminating the logarithmic term from the approximation bound for the Softplus network. In addition, we prove that all eigenfunctions lie in $\mathfrak{B}^{s+2}(\Omega)$ if the potential function belongs to $\mathfrak{C}^s(\Omega)$ with $s \geqslant 0$.

The authors in [38] also employed neural networks to solve the Dirichlet eigenvalue problem for the Schrödinger operator posed on a bounded C^m domain Ω. They assumed that $V \in C^{m-1}(\Omega)$ with $m > \max\{2, d/2 - 2\}$, and used a loss function that includes a boundary penalty, a normalization penalty, and an orthogonal penalty. They demonstrated a convergence rate $\mathcal{O}(n^{-1/16})$ for the generalization error. In contrast, by using the Rayleigh quotient and trial functions that precisely satisfy the boundary conditions, our loss function includes only the orthogonal penalty, resulting in improved numerical accuracy. Additionally, we prove a better accumulative rate for the generalization errors.

The main limitations of our work are twofold. First, in Assumption 2.1, we assume that V is bounded both above and below. This assumption excludes commonly used singular potentials, although the algorithm remains applicable to these singular potentials, such as the inverse square potential used in our numerical tests. The second limitation concerns the regularity assumption on V, namely $V \in \mathfrak{C}^s$ with $s \geqslant 1$. This indicates that V is relatively smooth, as demonstrated in [46] and [45].

1.3 Notations

Let $H^1(\Omega)$ be the standard Sobolev space [1] with the norm $\|\cdot\|_{H^1(\Omega)}$, while $H_0^1(\Omega)$ is the closure of $C_0^\infty(\Omega)$ in $H^1(\Omega)$. For a function set $\mathcal{F}$, we denote $\mathcal{F}_{>r} := \{f \in \mathcal{F} : \|f\|_{L^2(\Omega)} > r\}$. The constant C may differ from line to line.

The remainder of the paper is organized as follows. First, we introduce the basic settings and show stability estimates in §2. Our main theoretical results are summarized in §3, followed by the numerical results in §4. We then focus on proving the theoretical findings. In §5, we derive an oracle inequality for the generalization error. The approximation results are presented in §6, while the statistical error bound is established in §7. The proof of Theorem 3.4 is provided in §8. Regularity estimates are derived in §9. Stability estimates and proofs for certain technical results are deferred to Appendix A–Appendix D. Results concerning the penalty method are presented in §Appendix E, the accumulative error is analyzed in §Appendix F, and a priori bound for the eigenmodes is proven in §Appendix G.

2 Basic settings

Let $\Omega = (0,1)^d$ be the unit hypercube. Consider the Dirichlet eigenvalue problem

$$\mathcal{H}u := -\Delta u + Vu = \lambda u \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{2.1}$$

where $\mathcal{H}$ is the Schrödinger operator and V is a potential function. Throughout this paper, we make the following assumption on V.

Assumption 2.1. *There exist positive constants $V_{\min}$ and $V_{\max}$ such that $V_{\min} \leqslant V(x) \leqslant V_{\max}$ for every $x \in \Omega$.*

Given Assumption 2.1 and the fact that Ω is a connected, open, bounded domain, $(-\Delta + V)^{-1}$ is a compact, self-adjoint operator on $L^2(\Omega)$. Consequently, $\mathcal{H}$ has a purely discrete spectrum $\{\lambda_j\}_{j=1}^{\infty}$, with ∞ as the only accumulation point. The eigenfunctions of $\mathcal{H}$ form an orthonormal basis on $L^2(\Omega)$ [22]. We list all the eigenvalues in an ascending order, with multiplicities, as $0 < \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \cdots \uparrow \infty$, and denote the first $k-1$ normalized orthogonal eigenfunctions by $\{\psi_j\}_{j=1}^{k-1}$. The k-th smallest eigenvalue is obtained by minimizing the Rayleigh quotient

$$\lambda_k = \min_{E} \max_{u \in E\setminus\{0\}} \frac{\langle u, \mathcal{H}u\rangle_{H^1\times H^{-1}}}{\|u\|^2_{L^2(\Omega)}},$$

where the minimum is taken over all k-dimensional subspace $E \subset H_0^1(\Omega)$, and $\langle\cdot,\cdot\rangle_{H^1\times H^{-1}}$ denotes the dual product on $H^1(\Omega)\times H^{-1}(\Omega)$. Given the first $k-1$ eigenpairs $(\lambda_1,\psi_1), \cdots, (\lambda_{k-1},\psi_{k-1})$, the k-th eigenvalue λ_k may be characterized as

$$\lambda_k = \min_{u \in E^{(k-1)}} \frac{\langle u, \mathcal{H}u\rangle_{H^1\times H^{-1}}}{\|u\|^2_{L^2(\Omega)}}, \tag{2.2}$$

where $E^{(k-1)} = \big\{u \in H_0^1(\Omega)\setminus\{0\} \ \big|\ u \perp \psi_i, 1 \leqslant i \leqslant k-1\big\}$.

It is intuitive to seek an approximate solution to Problem (2.2) within a hypothesis class $\mathcal{F} \subset H_0^1(\Omega)$ that is parameterized by neural networks. To achieve this, we introduce a loss function that incorporates orthogonal penalty terms to enforce the orthogonal constraints. Specifically, the loss function for computing the k-th eigenfunction is given by

$$L_k(u) = \frac{\langle u, \mathcal{H}u\rangle_{H^1\times H^{-1}}}{\|u\|^2_{L^2(\Omega)}} + \beta \sum_{j=1}^{k-1} \frac{\langle u, \psi_j\rangle^2}{\|u\|^2_{L^2(\Omega)}}, \tag{2.3}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product on $L^2(\Omega)$, and β is a penalty parameter that should be chosen such that $\beta > \lambda_k - \lambda_1$. For the ground state (i.e., when $k = 1$), the loss function simplifies to the Rayleigh quotient, and the orthogonal penalty term is no longer needed. In practice, the Monte Carlo method is employed to compute the high-dimensional integral in the loss function, and an approximate solution is obtained through empirical risk minimization. Denote by $\mathcal{P}_\Omega$ the uniform probability distribution on Ω

and X, X_1, X_2, $\cdots$ are *i.i.d.* (independent identically distributed) random variables according to $\mathcal{P}_\Omega$. The population loss $L_k(u)$ is rewritten as

$$L_k(u) = \frac{\mathcal{E}_V(u) + \mathcal{E}_P(u)}{\mathcal{E}_2(u)}, \tag{2.4}$$

where

$$\mathcal{E}_V(u) := \mathbf{E}\left[|\nabla u(X)|^2 + V(X)|u(X)|^2\right],$$
$$\mathcal{E}_P(u) := \beta \sum_{j=1}^{k-1} \left(\mathbf{E}\left[u(X)\psi_j(X)\right]\right)^2, \qquad \mathcal{E}_2(u) := \mathbf{E}\left[|u(X)|^2\right].$$

The population loss $L_k(u)$ is approximated by the empirical loss

$$L_{k,n}(u) = \frac{\mathcal{E}_{n,V}(u) + \mathcal{E}_{n,P}(u)}{\mathcal{E}_{n,2}(u)}, \tag{2.5}$$

where

$$\mathcal{E}_{n,V} := \frac{1}{n}\sum_{i=1}^{n} \left(|\nabla u(X_i)|^2 + V(X_i)|u(X_i)|^2\right),$$
$$\mathcal{E}_{n,P} := \beta \sum_{j=1}^{k-1} \left(\frac{1}{n}\sum_{i=1}^{n} u(X_i)\psi_j(X_i)\right)^2, \qquad \mathcal{E}_{n,2} := \frac{1}{n}\sum_{i=1}^{n} |u(X_i)|^2.$$

Let $\widehat{u}_n$ be a minimizer of $L_{k,n}(u)$ within $\mathcal{F}$, i.e., $\widehat{u}_n = \arg\min_{u\in\mathcal{F}} L_{k,n}(u)$, and we approximate λ_k by

$$\widehat{\lambda}_{k,n} = \frac{\langle \widehat{u}_n, \mathcal{H}\widehat{u}_n\rangle}{\langle \widehat{u}_n, \widehat{u}_n\rangle}.$$

Let U_k be the true solution space for the k-th eigenfunction in $H_0^1(\Omega)$, i.e.,

$$U_k := \operatorname{span}\{\psi_1, \psi_2, \ldots, \psi_{k-1}\}^{\perp} \cap \ker(\mathcal{H} - \lambda_k I) \subset H_0^1(\Omega). \tag{2.6}$$

Any non-zero function is a minimizer of $L_k(u)$ if and only if it lies in U_k. Our goal is to estimate $|\widehat{\lambda}_{k,n} - \lambda_k|$ and the offset of the direction of $\widehat{u}_n$ from the subspace U_k, which is commonly referred as the generalization error.

The following proposition shows that $|\widehat{\lambda}_{k,n} - \lambda_k|$ may be bounded by the energy excess $L_k(\widehat{u}_n) - \lambda_k$.

Proposition 2.2. *Under Assumption 2.1, for any nonzero $u \in H^1(\Omega)$ and $\beta > \lambda_k - \lambda_1$,*

$$\left|\frac{\langle u, \mathcal{H}u\rangle}{\langle u, u\rangle} - \lambda_k\right| \leqslant \max\left\{\frac{\lambda_k - \lambda_1}{\beta + \lambda_1 - \lambda_k}, 1\right\}\left(L_k(u) - \lambda_k\right).$$

To quantify the offset of a direction u from the subspace U_k, we define $P^{\perp}$ as the orthogonal projection operator from $L^2(\Omega)$ to $U_k^{\perp}$, the orthogonal complement of U_k. Let $\lambda_{k'}$ be the first eigenvalue of $\mathcal{H}$ that is strictly greater than λ_k, i.e., $k' \geqslant k+1$, $\lambda_{k'} > \lambda_k$ and $\lambda_{k'-1} = \lambda_k$. The following proposition shows that $\|P^{\perp}\widehat{u}_n\|_{H^1(\Omega)}$ may also be bounded by the energy excess $L_k(\widehat{u}_n) - \lambda_k$.

Proposition 2.3. *Under Assumption 2.1, for any $u \in H^1(\Omega)$ and $\beta > \lambda_k - \lambda_1$,*

$$\left\|P^{\perp}u\right\|_{L^2(\Omega)}^2 \leqslant \frac{L_k(u) - \lambda_k}{\min\{\beta + \lambda_1 - \lambda_k, \lambda_{k'} - \lambda_k\}}\|u\|_{L^2(\Omega)}^2, \tag{2.7a}$$

$$\left\|\nabla\left(P^{\perp}u\right)\right\|_{L^2(\Omega)}^2 \leqslant \left(L_k(u) - \lambda_k\right)\left(\frac{\lambda_k - V_{\min}}{\min\{\beta + \lambda_1 - \lambda_k, \lambda_{k'} - \lambda_k\}} + 1\right)\|u\|_{L^2(\Omega)}^2. \tag{2.7b}$$

The above two results generalize [46, Proposition 2.1] to multiple eigenvalues. Proposition 2.2 suggests that $\beta + \lambda_1 - \lambda_k$ serves as a metric for evaluating the stability of the approximate eigenvalues, while Proposition 2.3 demonstrates that both the gap $\lambda_{k'} - \lambda_k$ and the factor $\beta + \lambda_1 - \lambda_k$ influence the stability of the approximate eigenfunctions. We postpone the proof of Propositions 2.2 and 2.3 to Appendix A.

3 Main Results

Before stating the main result, we introduce a new function space. Let $\{\hat{u}(k)\}_{k\in\mathbb{N}_+^d}$ denote the Fourier coefficients of $u \in L^1(\Omega)$ against the basis $\{\Phi_k\}_{k\in\mathbb{N}_+^d}$ given by

$$\{\Phi_k\}_{k\in\mathbb{N}_+^d} := \left\{\prod_{i=1}^d \sin(\pi k_i x_i) \mid k = (k_1, k_2, \dots, k_d), k_i \in \mathbb{N}_+\right\}.$$

For $s \geqslant 0$, the sine spectral Barron space $\mathfrak{B}^s(\Omega)$ is defined by

$$\mathfrak{B}^s(\Omega) := \left\{u \in L^1(\Omega) \;\mid\; \|u\|_{\mathfrak{B}^s(\Omega)} < \infty\right\}, \tag{3.1}$$

which is equipped with the spectral Barron norm

$$\|u\|_{\mathfrak{B}^s(\Omega)} = \sum_{k\in\mathbb{N}_+^d} (1 + \pi^s |k|_1^s)\, |\hat{u}(k)|, \tag{3.2}$$

where $|k|_1$ is the ℓ^1-norm of a vector k. $\mathfrak{B}^s(\Omega)$ is a Banach space, as it can be viewed as a weighted ℓ^1 space $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$, of the sine coefficients defined on the lattice $\mathbb{N}_+^d$ with the weight $W_s(k) = 1 + \pi^s|k|_1^s$. Moreover, it is straightforward to verify that the functions in $\mathfrak{B}^s(\Omega)$ are continuous and vanish on the boundary.

Our approximation result shows that functions in $\mathfrak{B}^s(\Omega)$ may be well approximated by $\varphi v(x;\theta)$ in the H^1-norm using a two-layer neural network with ReLU or Softplus activation functions, where v is a neural network function and φ is an approximating distance function, and we may take φ as

$$\varphi(x) = \left[\sum_{i=1}^d \frac{1}{\sin(\pi x_i)}\right]^{-1} = \frac{\prod_{i=1}^d \sin(\pi x_i)}{\sum_{i=1}^d \prod_{\substack{1\leqslant j\leqslant d\\ j\neq i}} \sin(\pi x_j)}, \qquad x \in \Omega.$$

Given an activation function ϕ, the number of the neurons m and a positive constant B, define

$$\mathcal{F}_{\phi,m}(B) := \left\{c + \sum_{i=1}^m \gamma_i \phi(w_i \cdot x - t_i) \mid |c| \leqslant B, |w_i|_1 = 1, |t_i| \leqslant 1, \sum_{i=1}^m |\gamma_i| \leqslant 4B\right\}.$$

The first approximation result concerns the approximation of the hypothesis space $\varphi\mathcal{F}_{\mathrm{ReLU},m}(B)$[1)] with $\mathrm{ReLU}(x) = \max\{x, 0\}$.

Theorem 3.1. *For $u \in \mathfrak{B}^s(\Omega)$ with $s \geqslant 3$, there exists $v_m \in \mathcal{F}_{\mathrm{ReLU},m}(\|u\|_{\mathfrak{B}^s(\Omega)})$ such that*

$$\|u - \varphi v_m\|_{H^1(\Omega)} \leqslant \frac{28\|u\|_{\mathfrak{B}^s(\Omega)}}{\sqrt{m}}.$$

Next, we replace the hypothesis space by $\varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ with the Softplus activation function

$$\mathrm{SP}_\tau(z) := \tau^{-1}\,\mathrm{SP}(\tau z) = \tau^{-1}\ln\left(1 + \mathrm{e}^{\tau z}\right),$$

where $\tau > 0$ is a scaling parameter, and $\mathrm{SP}_\tau \to \mathrm{ReLU}$ as $\tau \to 0$. The rescaled Softplus function may be viewed as a smooth approximation of ReLU, and its smoothness is particularly useful for bounding the complexities of function classes that involve the derivatives of the neural network functions. The following theorem shows that $\varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ admits a similar approximation bound as $\varphi\mathcal{F}_{\mathrm{ReLU},m}(B)$.

Theorem 3.2. *For $u \in \mathfrak{B}^s(\Omega)$ with $s \geqslant 3$, there exists $v_m \in \mathcal{F}_{\mathrm{SP}_\tau,m}\left(\|u\|_{\mathfrak{B}^s(\Omega)}\right)$ such that*

$$\|u - \varphi v_m\|_{H^1(\Omega)} \leqslant \frac{64\|u\|_{\mathfrak{B}^s(\Omega)}}{\sqrt{m}}.$$

1) In this paper, let $\varphi\mathcal{F} := \{\varphi v : v \in \mathcal{F}\}$, where $\mathcal{F}$ is a function set.

Motivated by Theorem 3.1 and Theorem 3.2, we make the following regularity assumption.

Assumption 3.3. *There exists a normalized eigenfunction $u^* \in \mathfrak{B}^s(\Omega)$ for some $s \geqslant 3$ lying in the subspace U_k defined in (2.6).*

Denote $\bar{\mu}_1 = 0$ and $\bar{\mu}_k = \max_{1\leqslant j\leqslant k-1} \|\psi_j\|_{L^\infty(\Omega)}$ for $k \geqslant 2$. The following generalization error bound is the main result of this work.

Theorem 3.4. *Under Assumptions 2.1 and 3.3, let $\mathcal{F} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ with $B = \|u^*\|_{\mathfrak{B}^s(\Omega)}$ and $\tau = 9\sqrt{m}$. For $r \in (0,1/2)$ and let u_n^m be a minimizer of the empirical loss $L_{k,n}$ within $\mathcal{F}_{>r}$. Given $\delta \in (0,1/3)$, assume that n and m are large enough so that $64B/\sqrt{m} \leqslant 1/2$ and*

$$\Upsilon_1(n,m,B,r,\delta) := \frac{B}{r}\sqrt{\frac{1+V_{\max}}{n}}\left(\sqrt{m\ln\frac{B\,(1+\sqrt{m}/d)\,(1+V_{\max})}{rd}}+\sqrt{\frac{\ln(1/\delta)}{d}}\right) \leqslant C, \tag{3.3a}$$

$$\Upsilon_2(n,m,k,B,\bar{\mu}_k,r,\delta) := \sqrt{\frac{k\bar{\mu}_k B}{nr}}\left[\sqrt{m\ln\left(\frac{\bar{\mu}_k B}{rd}\right)}+\sqrt{\frac{\ln(k/\delta)}{d}}\right] \leqslant 1, \tag{3.3b}$$

where C is an absolute constant. Then with probability at least $1-3\delta$,

$$L_k\left(u_n^m\right) - \lambda_k \leqslant C\big[\lambda_k\Upsilon_1(n,m,B,r,\delta) + \beta\Upsilon_2(n,m,k,B,\bar{\mu}_k,r,\delta) + (V_{\max}+\beta+\lambda_k)\,B/\sqrt{m}\big], \tag{3.4}$$

where C is an absolute constant.

In particular, with the choice of $m = \mathcal{O}(\sqrt{n/k})$ and n large enough, there exists $\tilde{C} > 0$ such that with probability at least $1-3\delta$,

$$L_k\left(u_n^m\right) - \lambda_k \leqslant \tilde{C}\left[\left(\frac{k}{n}\right)^{1/4}\left(\sqrt{\frac{\ln(n/k)}{k}}+1\right)+\sqrt{\frac{k\ln(k/\delta)}{n}}\right].$$

In (3.4), $\lambda_k\Upsilon_1$ and $\beta\Upsilon_2$ correspond to the statistical errors of the Rayleigh quotient and the orthogonal penalty term, respectively, and $(V_{\max}+\beta+\lambda_k)\,B/\sqrt{m}$ stands for the approximation error. The convergence rate $\mathcal{O}(n^{-1/4})$ is dimension-free. The dependence of $\tilde{C}$ on all parameters is explicit and at most a polynomial of low degree. We have established a high probability bound for the generalization error, which immediately implies an expectation bound.

To prove Theorem 3.4, we first introduce concentration inequalities for ratio-type suprema to address the Rayleigh quotient, and then derive an oracle inequality in §5, which decomposes the generalization error into the sum of the approximation error and the statistical error. The concentration inequality controls each term in statistical error by the expectation of suprema of empirical processes and probability tail terms.

In §6, we estimate the approximation error. To prove Theorem 3.1, we tackle the constraint approximation by first showing that $\mathfrak{B}^s(\Omega)$ lies in the convex hull of φ times spectral Barron functions of lower order, as presented in §6.2. In §6.3, we construct network to approximate trigonometric functions, demonstrating that the convex hull of $\varphi\mathcal{F}_{\mathrm{ReLU},m}$ is larger and contains $\mathfrak{B}^s(\Omega)$. Theorem 3.1 then follows from Maurey's lemma. Theorem 3.2 is derived by replacing ReLU with Softplus in the same construction, carefully controlling the error caused by the replacement. Using Theorem 3.2 and the continuity of L_k with respect to the H^1 norm, we bound the approximation error in Theorem 6.7.

To estimate the statistical error, we develop tools to bound the expectation of suprema of empirical processes by covering numbers of corresponding VC class in §7.2 and derive the covering number bounds for certain function classes in §7.1. Combining all these estimates, we obtain the statistical error bound in Theorem 7.9. Finally, Theorem 3.4 follows from the approximation results in § 6 and the statistical error bounds in § 7.

Remark 3.5. Under Assumption 2.1, there exist constants c_1, c_2 and c_3 such that for $k = 1,2,\cdots$,

$$c_1 dk^{2/d} + V_{\min} \leqslant \lambda_k \leqslant c_2 dk^{2/d} + V_{\max},$$

and

$$\|\psi_k\|_{L^\infty(\Omega)} \leqslant \left(c_3 k^{2/d} + \frac{e\,(V_{\max} - V_{\min})}{\pi d}\right)^{d/4}.$$

We refer to Appendix G for a proof.

Remark 3.6. Condition (3.3b) may be removed provided that Υ_2 in (3.4) is replaced by $\Upsilon_2(1+\Upsilon_2)^3$.

Substituting Theorem 3.4 into Proposition 2.2 and Proposition 2.3, we obtain

Corollary 3.7. *Under the same assumptions of Theorem 3.4, with $m = \mathcal{O}(\sqrt{n/k})$, there exist constants $\tilde{C}_1$ depending only on B, d, $V_{\max}$, λ_k, $\bar{\mu}_k$, β, r^{-1}, $(\beta+\lambda_1-\lambda_k)^{-1}$ polynomially, and $\tilde{C}_2$ depending only on the same constants as $\tilde{C}_1$ and $(\lambda_{k'}-\lambda_k)^{-1}$ polynomially such that with probability at least $1-3\delta$,*

$$\left|\frac{\langle u_n^m, \mathcal{H}u_n^m\rangle}{\langle u_n^m, u_n^m\rangle} - \lambda_k\right| \leqslant \tilde{C}_1\left[\left(\frac{k}{n}\right)^{1/4}\left(\sqrt{\frac{\ln(n/k)}{k}}+1\right) + \sqrt{\frac{k\ln(k/\delta)}{n}}\right],$$

$$\|P^\perp u_n^m\|_{H^1(\Omega)}^2 \leqslant \tilde{C}_2\left[\left(\frac{k}{n}\right)^{1/4}\left(\sqrt{\frac{\ln(n/k)}{k}}+1\right) + \sqrt{\frac{k\ln(k/\delta)}{n}}\right].$$

Next, we prove a regularity result of the eigenfunctions in the sine Barron space, which validates Assumption 3.3. We firstly recall the spectral Barron space defined in [46,47]. Let $\{\check{w}(k)\}_{k\in\mathbb{N}_0^d}$ represent the Fourier coefficients of a function $w \in L^1(\Omega)$ against the basis

$$\{\Psi_k\}_{k\in\mathbb{N}_0^d} := \left\{\prod_{i=1}^d \cos(\pi k_i x_i) \mid k_i \in \mathbb{N}_0\right\}.$$

For $s \geqslant 0$, the cosine spectral Barron space $\mathfrak{C}^s(\Omega)$ is defined by

$$\mathfrak{C}^s(\Omega) := \left\{w \in L^1(\Omega) \;\middle|\; \|w\|_{\mathfrak{C}^s(\Omega)} < \infty\right\}, \tag{3.5}$$

which is equipped with the spectral Barron norm

$$\|w\|_{\mathfrak{C}^s(\Omega)} = \sum_{k\in\mathbb{N}_0^d} (1+\pi^s|k|_1^s)\,|\check{w}(k)|.$$

Theorem 3.8. *If $V \in \mathfrak{C}^s(\Omega)$ with $s \geqslant 0$, then any eigenfunction of Problem (2.1) lies in $\mathfrak{B}^{s+2}(\Omega)$.*

Theorem 3.8 establishes that Assumption 3.3 holds if $V \in \mathfrak{C}^1(\Omega)$. To prove Theorem 3.8, we first show that the inverse of the Schrödinger operator $\mathcal{H}^{-1} : \mathfrak{B}^s(\Omega) \to \mathfrak{B}^{s+2}(\Omega)$ is bounded. We then derive regularity estimates for the eigenfunctions using a bootstrap argument. A detailed proof may be found in §9.

In Theorem 3.4, we assume that the L^2-norms of the approximated eigenfunctions are bounded below by r, where $r \in (0, 1/2)$. This assumption is reasonable, because, in practice, a very small L^2-norm can lead to numerical instability when computing the Rayleigh quotient. From a theoretical perspective, the L^2- normalization $u/\|u\|_{L^2(\Omega)}$ of all functions u in $\varphi\mathcal{F}_{\mathrm{SP}_\tau,m}$ is not uniformly bounded, which may cause the statistical error to blow up. Our results show that the statistical error depends linearly on $1/r$.

In practice, we treat the minimization of $L_{k,n}$ in $\mathcal{F}_{>r}$ as solving an optimization problem in $\mathcal{F}$, subject with the constraint $\|u\|_{L^2(\Omega)} > r$. As shown in Figure 1, the L^2-norm of the network function gradually increases as $L_{k,n}$ is minimized using stochastic gradient descent methods, due to the scaling invariance of $L_{k,n}$. The gradient descent methods implicitly regularize the L^2-norm of the solution when minimizing a scaling-invariant loss function, enabling us to automatically obtain solutions that satisfy the constraint $\|u\|_{L^2(\Omega)} > r$.

Adding a normalized penalty term is a natural approach to solve the constrained optimization problem. In [21,38], the authors introduced a normalized penalty term $\gamma\,(\mathcal{E}_2(u)-1)^2$ in the loss function. We shall analyze this method in § 3.1.

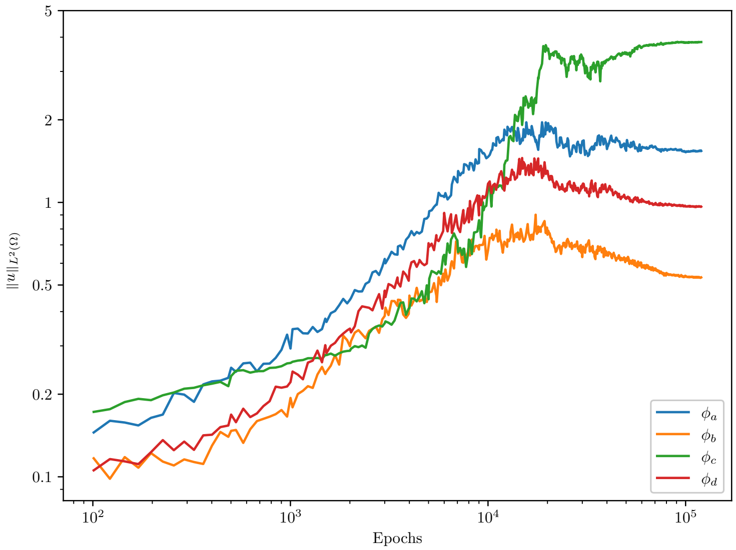


Figure 1 $\|u\|_{L^2(\Omega)}$ for different cut-off functions. Details of the numerical experiments are provided in Section 4. (Color online)

3.1 Extensions to the normalization penalty method

The normalization penalty method in [21] employs the population loss

$$\mathscr{L}_k(u) := L_k(u) + \gamma\left(\mathcal{E}_2(u) - 1\right)^2$$

and the empirical loss $\mathscr{L}_{k,n}(u) := L_{k,n}(u) + \gamma\left(\mathcal{E}_{n,2}(u) - 1\right)^2$, where $\gamma > 0$ is a penalty parameter. Let u_n be a minimizer of $\mathscr{L}_{k,n}(u)$ within $\mathcal{F}$. Firstly, we show that $\|u_n\|_{L^2(\Omega)} \geqslant 1/2$ with high probability for sufficiently large γ.

Theorem 3.9. *Under Assumptions 2.1 and 3.3, let $\gamma \geqslant 4\lambda_k$ and $u_n = \arg\min_{u\in\mathcal{F}} \mathscr{L}_{k,n}(u)$ where $\mathcal{F} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ with $B = \|u^*\|_{\mathfrak{B}^s(\Omega)}$ and $\tau = 9\sqrt{m}$. Given $\delta \in (0, 1/4)$, assume that n and m are large enough such that $C\left(1 + V_{\max} + \beta/\gamma\right) B/\sqrt{m} \leqslant 1$ and*

$$\frac{CB}{d}\left(\frac{B}{d} + 1\right)\sqrt{\frac{d\,(1 + \ln B)\,m + \ln(1/\delta)}{n}} \leqslant 1, \tag{3.6a}$$

$$C\left(1 + V_{\max} + \beta/\gamma\right) B^2\sqrt{\frac{\ln(1/\delta)}{n}} + \frac{C\beta\bar{\mu}_k B}{\gamma d}\sqrt{\frac{k\ln(k/\delta)}{n}} \leqslant 1, \tag{3.6b}$$

where C is an absolute constant. Then with probability at least $1 - 4\delta$,

$$\mathcal{E}_2(u_n) \geqslant 1/4.$$

Remark 3.10. The assumption $\gamma \geqslant 4\lambda_k$ in Theorem 3.9 may be relaxed. Pursuing the proof, one can prove that $\|u_n\|_{L^2(\Omega)} \geqslant r$ with high probability when n, m are large enough, as long as $\gamma > \lambda_k/(1 - r^2)^2$.

The condition (3.6) is weaker than (3.3) to certain degree, because the left-hand sides of (3.3a), (3.3b) are $\mathcal{O}(\sqrt{m\ln m/n})$, $\mathcal{O}(\sqrt{m/n})$ while the left-hand sides of (3.6a), (3.6b) are $\mathcal{O}(\sqrt{m/n})$, $\mathcal{O}(1/\sqrt{n})$.

As a direct application of our method, we obtain the generalization error bound of the normalization penalty method. Let $\Upsilon(n, m, k, B, \bar{\mu}_k, \beta, r, \delta)$ denote the error bound in the right-hand side of (3.4), i.e.,

$$\Upsilon(n, m, k, B, \bar{\mu}_k, \beta, r, \delta) := C\left[\lambda_k\Upsilon_1(n, m, B, r, \delta) + \beta\Upsilon_2(n, m, k, B, \bar{\mu}_k, r, \delta) + \left(V_{\max} + \beta + \lambda_k\right) B/\sqrt{m}\right].$$

Corollary 3.11. *Under the same assumptions of Theorem 3.9, for $\delta \in (0, 1/7)$, assume further that n and m are large enough so that (3.3a) and (3.3b) hold with $r = 0.49$. Then, there exists C such that*

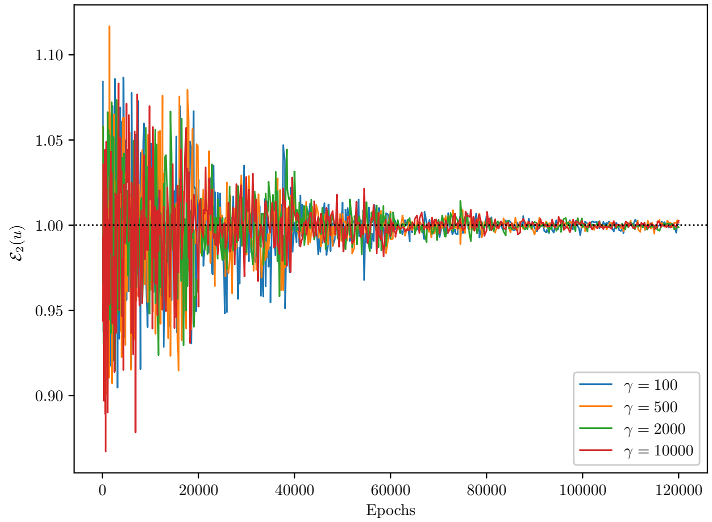


Figure 2 $\mathcal{E}_2(u)$ for different cut-off functions. Details about the numerical experiments are provided in Section 4. (Color online)

with probability at least $1-7\delta$,

$$\begin{aligned} L_k(u_n) - \lambda_k \leqslant\ & \Upsilon(n,m,k,B,\bar{\mu}_k,\beta,1,\delta) \\ & + C\gamma\left[\left(\frac{B^2}{d^2}+1\right)\sqrt{\frac{d\,(1+\ln B)\,m+\ln(1/\delta)}{n}}+\frac{B^2}{m}\right]. \end{aligned} \tag{3.7}$$

The second term in the right-hand side of (3.7) corresponds to the statistical and approximation error due to the normalization penalty term. There is a trade-off with the normalization penalty method: while a larger value of γ ensures that u_n remains away from zero, but results in a larger generalization error. To prove Theorem 3.9, we derive a new oracle inequality to bound $|\mathcal{E}_2(u_n)-1|$ utilizing Hoeffding's inequality. We then prove the generalization bound using the Rademacher complexity. The Rademacher complexity is bounded by applying Dudley's theorem and leveraging the covering number bounds discussed in § 7.

The poof of Corollary 3.11 relies on the fact $\|u_n\|_{L^2(\Omega)} \geqslant 1/2$, allowing the estimates used in the proof of Theorem 3.4 with $r=0.49$ to remain applicable. We refer to Appendix E for the proof of Theorem 3.9 and Corollary 3.11. Numerical results in Figure 2 show that $\mathcal{E}_2(u_n)$ exceeds 1/4 with high probability.

Remark 3.12 (The choice of hyperparameters)**.** By Proposition 2.2 and Proposition 2.3, it follows that $\beta-\lambda_k+\lambda_1>0$ should not be excessively small. For the normalization penalty method, γ should be greater than a constant multiple of λ_k. From (3.4) and (3.7), a reasonable choice of both β and γ are $\mathcal{O}(\lambda_k)$.

3.2 Analysis of the accumulative error

In practice, we replace the exact eigenfunctions in the orthogonal penalty term with the approximate eigenfunctions obtained in the previous $k-1$ steps, which introduces an accumulative error. Theorem 3.13 details the impact of using approximate eigenfunctions on the generalization bound for the k-th step, while Proposition 3.14 provides the rate at which the accumulated generalization errors grow.

Let $\mathfrak{u}_{\theta j}$ denote the j-th approximate eigenfunction parameterized by the neural network. We use the L^2-normalization of $\mathfrak{u}_{\theta j}$ as an approximation of ψ_j and denote $\bar{\nu}_k = \max_{1\leqslant j\leqslant k-1}\|\mathfrak{u}_{\theta j}\|_{L^\infty(\Omega)}/\|\mathfrak{u}_{\theta j}\|_{L^2(\Omega)}$. Consider the loss function

$$\widetilde{L}_k(u) = \frac{\mathcal{E}_V(u)}{\mathcal{E}_2(u)} + \beta_k\sum_{j=1}^{k-1}\frac{\mathscr{P}_j(u)}{\mathcal{E}_2(u)\mathcal{E}_2(\mathfrak{u}_{\theta j})}$$

and the empirical loss

$$\widetilde{L}_{k,n}(u) = \frac{\mathcal{E}_{n,V}(u)}{\mathcal{E}_{n,2}(u)} + \beta_k \sum_{j=1}^{k-1} \frac{\mathscr{P}_{n,j}(u)}{\mathcal{E}_{n,2}(u)\mathcal{E}_{n,2}\left(\mathfrak{u}_{\theta j}\right)},$$

where $\mathscr{P}_j(u) := \langle u, \mathfrak{u}_{\theta j}\rangle^2$ and $\mathfrak{P}_{n,j}(u) := \left(n^{-1}\sum_{i=1}^n u(X_i)\mathfrak{u}_{\theta j}(X_i)\right)^2$.

Theorem 3.13. *Under Assumptions 2.1 and 3.3, let $\mathcal{F} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ with $B = \|u^*\|_{\mathfrak{B}^s(\Omega)}$ and $\tau = 9\sqrt{m}$. Let $r \in (0,1/2)$, $\|\mathfrak{u}_{\theta j}\|_{L^2(\Omega)} \geqslant r$ for all j and $\mathfrak{u}_{\theta k} = \arg\min_{u\in\mathcal{F}_{>r}} \widetilde{L}_{k,n}(u)$. Given $\delta \in (0,1/3)$, assume that n and m are large enough so that $64B/\sqrt{m} \leqslant 1/2$, (3.3a) and (3.3b) with $\bar{\mu}_k$ replaced by $\bar{\nu}_k$ hold true. Then, with probability at least $1-3\delta$,*

$$L_k\left(\mathfrak{u}_{\theta k}\right) - \lambda_k \leqslant \Upsilon(n,m,k,B,\bar{\nu}_k,\beta_k,r,\delta) + 8\beta_k \sum_{j=1}^{k-1} \sqrt{\frac{L_j(\mathfrak{u}_{\theta j}) - \lambda_j}{\min\left\{\beta_j + \lambda_1 - \lambda_j, \lambda_{j'} - \lambda_j\right\}}}. \tag{3.8}$$

The first term in the right-hand side of (3.8) is the generalization error at k-th step and the second term represents the accumulative error, where $\bar{\nu}_k$ plays a similar role as $\bar{\mu}_k$ in Theorem 3.4. Proposition 3.14 shows that the generalization error grows quadratically as the order of the eigenfunction.

Proposition 3.14. *Assume that for all $k \geqslant 1$,*

$$L_k\left(\mathfrak{u}_{\theta k}\right) - \lambda_k \leqslant \Delta_k + 8\beta_k \sum_{j=1}^{k-1} \sqrt{\frac{L_j(\mathfrak{u}_{\theta j}) - \lambda_j}{\min\left\{\beta_j + \lambda_1 - \lambda_j, \lambda_{j'} - \lambda_j\right\}}}.$$

Let $\tau_k = \max_{1\leqslant j\leqslant k} \Delta_j/\beta_j$, $\rho_0 = 0$ and $\rho_k = \max_{1\leqslant j\leqslant k} 4\sqrt{\beta_j/\min\left\{\beta_j + \lambda_1 - \lambda_j, \lambda_{j'} - \lambda_j\right\}}$. Then,

$$L_k\left(\mathfrak{u}_{\theta k}\right) - \lambda_k \leqslant \beta_k\Big((k-1)\rho_{k-1} + \sqrt{\tau_k}\Big)^2. \tag{3.9}$$

When $k=1$, the bound (3.9) simplifies to $L_k\left(\mathfrak{u}_{\theta k}\right) - \lambda_k \leqslant \Delta_k$, with no accumulated error.

To prove Theorem 3.13, we shall derive a uniform bound for $\left|\widetilde{L}_k(u) - L_k(u)\right|$ and establish a new oracle inequality to address the penalty term. Proposition 3.14 is proved by an induction argument. Using a similar argument, one can show that the quadratic growth rate of accumulative error with respect to k is sharp. We defer the poof of Theorem 3.13 and Proposition 3.14 to Appendix F. As with Corollary 3.7, the error bounds in Corollary 3.11, Theorem 3.13 and Proposition 3.14 can also be expressed in terms of the eigenvalues and the H^1-norm of eigenfunctions. For simplicity, We do not delve into the details here.

Remark 3.15 (Generalizability to more general neural networks)**.** For simplicity, we focus on shallow ReLU and Softplus networks to prove the theoretical results. However, our findings also extend to more general activation functions σ as long as σ is smooth and $\left|(\mathrm{ReLU} - \sigma)^{(l)}(z)\right|$ decays at least exponentially as $|z| \to \infty$ for $l = 0,1$. This condition ensures that the rescaled version $\sigma_\tau(\cdot) := \tau^{-1}\sigma(\tau\cdot)$ satisfies similar properties to those in Lemma B.5. Consequently, all our results hold with only different absolute constants. For example, our proof applies to GeLU, SiLU and Mish because the distances between ReLU and these functions decay exponentially, including the derivatives of these distances.

Our analysis may also be extended to deep neural networks (DNNs). Regarding approximation error, our proof shows that we only need to approximate the spectral Barron functions one order lower than the eigenfunction using DNNs. Recent results in [45] provide further insights into this matter. For statistical error, the covering number of shallow networks should be replaced by that of DNNs. The VC dimension and the covering number of DNNs are well-established and can be found in [7]. While our focus here is on the key challenges related to constrained approximation and generalization error bounds, the extension to DNNs will be explored in future work. In all our numerical experiments, we use DNNs to leverage their superior representation power.

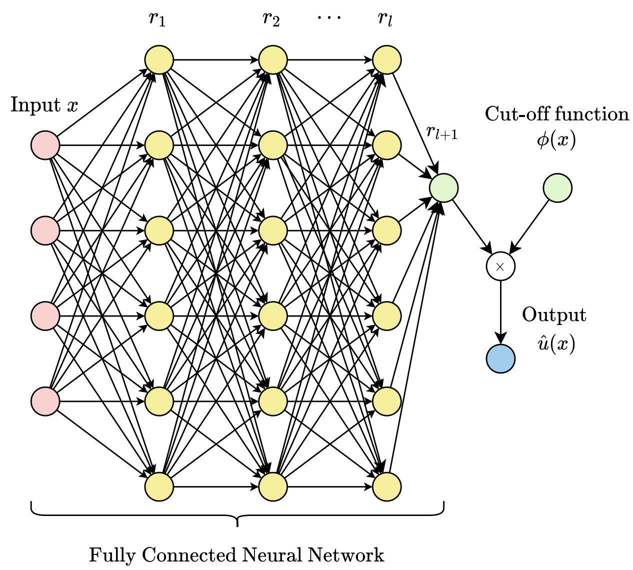


Figure 3 The neural network architecture combined with cut-off funcions. (Color online)

4 Numerical results

4.1 Deep Learning Model

In this section, we show several numerical experiments for (2.1) to demonstrate the effectiveness of our method. Although the previous theoretical analysis mainly focused on shallow networks, which only contain one hidden layer, we will use neural networks with more hidden layers for the numerical experiments, as is common in most related works. The main reason for this choice is that training DNN tends to be more efficient, and their numerical performance is usually superior to that of shallow networks.

Figure 3 shows the neural network architecture we used. The left part of our model consists of a fully connected neural network with several hidden layers, each having the same width. We denote the number of hidden layers by l and the width of each layer by m. In our experiments, we take $l = 3$ and $m = 40$, resulting in a network with approximately $3,500$ parameters. The activation function used is $\sigma = \tanh$, and we vectorize $\sigma(x)$ as $\tilde{\sigma}(x)$, i.e., $\tilde{\sigma}(x) = (\sigma(x_1), \sigma(x_2), \ldots, \sigma(x_m))$.

The input layer maps the coordinates of the sampling points, from $\mathbb{R}^d$ to $\mathbb{R}^m$. The output of the first layer is given by $r_1 = \tilde{\sigma}(W_1 x + b_1)$, where $W_1 \in \mathbb{R}^{m\times d}$ and $b_1 \in \mathbb{R}^m$. Subsequent hidden layers also contain similar transformations, mapping values from $\mathbb{R}^m$ to $\mathbb{R}^m$, and the output of the i-th layer is represented by

$$r_i = \tilde{\sigma}(W_i r_{i-1} + b_i), \quad 2 \leqslant i \leqslant l,$$

where $W_i \in \mathbb{R}^{m\times m}$ and $b_i \in \mathbb{R}^m$. The final layer on the left part is an output node, yielding a value given by $r_{l+1} = W_{l+1} r_l + b_{l+1}$, where $W_{l+1} \in \mathbb{R}^m$, $b_{l+1} \in \mathbb{R}$. The final output function $u(x)$ is obtained by multiplying the output of the left part network with a cut-off function $u(x) = r_{l+1}(x)\phi(x)$. As mentioned earlier, the cut-off function plays a key role in the architecture by ensuring the output function satisfies the homogeneous Dirichlet boundary condition. Similar architectures appear in [31, 36, 39].

The complete set of parameters in our architecture is defined as $\theta\colon = \{W_1, \ldots, W_{l+1}, b_1, \ldots, b_{l+1}\}$. In each epoch, the loss function is computed using (2.5), and parameters are updated using the adaptive moment estimation (ADAM) optimizer. This process is iterated over multiple epochs until the loss function decreases to a sufficiently small value, indicating that the approximate eigenmodes have been achieved. Initially, the learning rate is set to 5×10^{-3}, and $1,000$ points are used to compute the empirical loss. To balance accuracy and efficiency, the learning rate is reduced to a quarter of its current value, and the number of sampling points is doubled every $20,000$ epochs. Each eigenmode is derived through parameter updates over $120,000$ epochs. Besides, when solving for the k-th eigenvalue ($k \geqslant 2$), we take $\beta = 4\lambda_{k-1}$. This choice is sufficiently large to find the subsequent correct eigenvalue in these cases.

4.2 Regular potential

In the first test, we use the potential

$$V(x_1, \cdots, x_d) = \frac{1}{d}\sum_{i=1}^{d} \cos(\pi x_i + \pi), \tag{4.1}$$

and test our method in $\Omega = (-1, 1)^d$. Since this potential function is essentially decoupled, we compute the reference eigenmodes using the spectral method, as described in [33].

We test four different cut-off functions,

$$\phi_a = \prod_{i=1}^{d}(1 - x_i^2), \quad \phi_b = \left(\sum_{i=1}^{d} \frac{1}{1 - x_i^2}\right)^{-1}, \quad \phi_c = \prod_{i=1}^{d} \cos\left(\frac{\pi}{2}x_i\right), \quad \phi_d = \left(\sum_{i=1}^{d} \frac{1}{\cos(\frac{\pi}{2}x_i)}\right)^{-1},$$

and the results are summarized in Table 1. All results indicate that our method provides satisfactory solutions. Each of the four cut-off functions produce solutions with errors less than 6×10^{-4} for the first eigenvalue and less than 5×10^{-3} for at least the first 30 eigenvalues. Notably, the performance varies slightly across different cut-off functions, with specific functions, such as ϕ_c, achieving errors of less than 1×10^{-3} for the first 30 eigenvalues.

We also compare our results with the standard penalty method, which imposes the boundary conditions in a soft manner. The loss function we used is

$$\frac{\langle u, \mathcal{H}u\rangle_{H^1\times H^{-1}}}{\|u\|^2_{L^2(\Omega)}} + \beta\sum_{j=1}^{k-1} \frac{\langle u, \psi_j\rangle^2}{\|u\|^2_{L^2(\Omega)}} + \gamma\frac{\|u\|^2_{L^2(\partial\Omega)}}{\|u\|^2_{L^2(\Omega)}}. \tag{4.2}$$

It includes an additional boundary penalty term compared with (2.3). On the one hand, a small γ may introduce considerable model error. On the other hand, using a large γ can enhance numerical accuracy, but training will become more difficult and inefficient due to the rough loss landscape. To ensure a fair comparison, we use the boundary penalty method with different hyperparameter γ.

For comparison, we set the cut-off function be the identity function in the network architecture and keep the rest of the configurations the same as in the previous test. Table 2 demonstrates that the boundary penalty method performs much worse than our method, with errors one or more orders of magnitude larger than ours. Even with the optimal choice of the hyperparameter, $\gamma = 500$, the numerical accuracy deteriorate rapidly, and the error for the tenth eigenvalue exceeds 1×10^{-2}. While taking other hyperparameter γ, the error is larger than 1×10^{-2} even for the first eigenvalue.

Next, we set $d = 10$ and evaluate our method in a higher-dimension scenario. In this case, the smallest eigenvalue is unique, and the second to the eleventh eigenvalues are equal, followed by 45 equal eigenvalues. Therefore, we choose to calculate only the first 15 eigenvalues. The cut-off functions ϕ_a and ϕ_c are employed, while also using the boundary penalty method with a proper γ. As shown in Table 3, our method outperforms the boundary penalty method. All calculations yield errors less than 1×10^{-2} for the first 15 eigenvalues, which is significantly less than that of the boundary penalty method. Notably, the cut-off function ϕ_c seems to be the best for $d = 10$.

We summarize the previous results in Figure 4, which demonstrates that the proposed methods all significantly outperform the penalty method in both cases, although the performance of different cut-off functions may vary.

Table 1 Estimates of the eigenvalues with potential (4.1) and $d = 5$, using different cut-off functions.

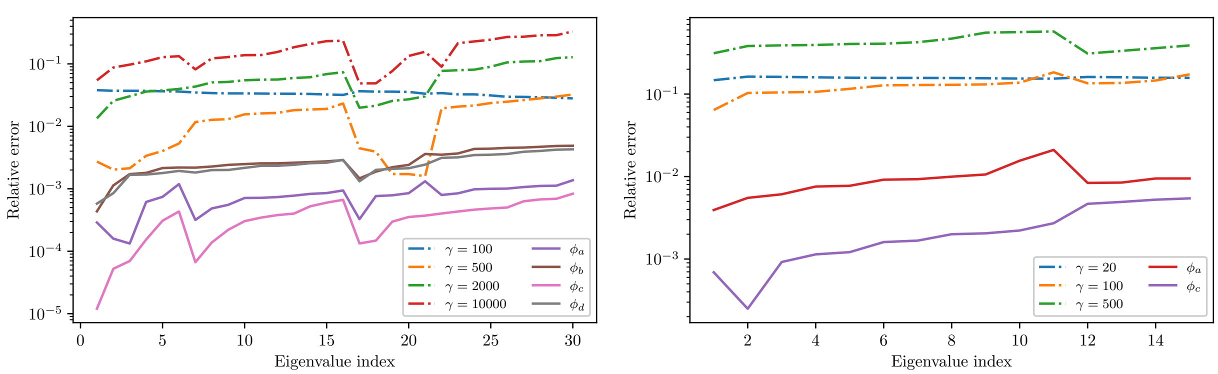
		$k = 1$	$k = 2$	$k = 3$	$k = 5$	$k = 10$	$k = 15$	$k = 30$
	Exact	11.8345	19.3369	19.3369	19.3369	26.8392	26.8392	34.3416
ϕ_a	Result	11.8379	19.3338	19.3394	19.3512	26.8583	26.8621	34.3886
	Rel. error	2.87×10^{-4}	1.60×10^{-4}	1.29×10^{-4}	7.40×10^{-4}	7.12×10^{-4}	8.53×10^{-4}	1.37×10^{-3}
ϕ_b	Result	11.8396	19.3588	19.3700	19.3784	26.9059	26.9130	34.5095
	Rel. error	4.31×10^{-4}	1.13×10^{-3}	1.71×10^{-3}	2.15×10^{-3}	2.49×10^{-3}	2.75×10^{-3}	4.89×10^{-3}
ϕ_c	Result	11.8343	19.3358	19.3382	19.3428	26.8474	26.8553	34.3702
	Rel. error	1.69×10^{-5}	5.69×10^{-5}	6.72×10^{-5}	3.05×10^{-4}	3.06×10^{-4}	6.00×10^{-4}	8.33×10^{-4}
ϕ_d	Result	11.8413	19.3533	19.3692	19.3714	26.8972	26.9095	34.4887
	Rel. error	5.75×10^{-4}	8.48×10^{-4}	1.67×10^{-3}	1.78×10^{-3}	2.16×10^{-3}	2.62×10^{-3}	4.28×10^{-3}

Table 2 Estimates of the eigenvalues with potential (4.1) and $d = 5$, using the boundary penalty method with different parameters γ.

		$k = 1$	$k = 2$	$k = 3$	$k = 5$	$k = 10$	$k = 15$	$k = 30$
γ	Exact	11.8345	19.3369	19.3369	19.3369	26.8392	26.8392	34.3416
100	Result	11.3854	18.6194	18.6198	18.6323	25.9356	25.9711	33.3787
	Rel. error	3.79×10^{-2}	3.71×10^{-2}	3.71×10^{-2}	3.64×10^{-2}	3.37×10^{-2}	3.23×10^{-2}	2.80×10^{-2}
500	Result	11.8023	19.3761	19.3781	19.4148	27.2571	27.3496	35.4504
	Rel. error	2.72×10^{-3}	2.03×10^{-3}	2.13×10^{-3}	4.03×10^{-3}	1.56×10^{-2}	1.90×10^{-2}	3.23×10^{-2}
2000	Result	11.9934	19.8330	19.9228	20.0627	28.3052	28.6824	38.7467
	Rel. error	1.34×10^{-2}	2.57×10^{-2}	3.03×10^{-2}	3.75×10^{-2}	5.46×10^{-2}	6.87×10^{-2}	1.28×10^{-1}
10000	Result	12.4805	21.0279	21.2185	21.8146	30.5426	33.0487	45.7008
	Rel. error	5.46×10^{-2}	8.74×10^{-2}	9.73×10^{-2}	1.28×10^{-1}	1.38×10^{-1}	2.31×10^{-1}	3.31×10^{-1}

Table 3 Estimates of the eigenvalues with potential (4.1) and $d = 10$.

		$k = 1$	$k = 2$	$k = 3$	$k = 5$	$k = 11$	$k = 12$	$k = 15$
	Exact	24.1728	31.6250	31.6250	31.6250	31.6250	39.0772	39.0772
cut-off	Result	24.2677	31.7994	31.8178	31.8693	31.9868	39.4044	39.4476
(ϕ_a)	Rel. error	3.93×10^{-3}	5.51×10^{-3}	6.10×10^{-3}	7.72×10^{-3}	1.14×10^{-2}	8.37×10^{-3}	9.48×10^{-3}
cut-off	Result	24.1895	31.6329	31.6541	31.6633	31.711	39.2598	39.2898
(ϕ_c)	Rel. error	6.91×10^{-4}	2.50×10^{-4}	9.20×10^{-4}	1.21×10^{-3}	2.72×10^{-3}	4.67×10^{-3}	5.44×10^{-3}
penalty	Result	20.6123	26.4801	26.5147	26.6332	26.7601	32.7747	32.9277
($\gamma = 20$)	Rel. error	1.47×10^{-1}	1.63×10^{-1}	1.62×10^{-1}	1.58×10^{-1}	1.54×10^{-1}	1.61×10^{-1}	1.87×10^{-1}
penalty	Result	25.7217	34.8861	34.9384	35.2846	37.4258	44.3579	45.8481
($\gamma = 100$)	Rel. error	6.41×10^{-2}	1.03×10^{-1}	1.05×10^{-1}	1.16×10^{-1}	1.83×10^{-1}	1.35×10^{-1}	1.73×10^{-1}
penalty	Result	31.7256	43.7200	43.9436	44.4301	49.7948	51.1777	54.2697
($\gamma = 500$)	Rel. error	3.12×10^{-1}	3.82×10^{-1}	3.90×10^{-1}	4.05×10^{-1}	5.75×10^{-1}	3.10×10^{-1}	3.89×10^{-1}



(a) $d = 5$ (b) $d = 10$

Figure 4 Relative error of the eigenvalues with potential (4.1). (Color online)

4.3 Inverse square potential

In the last test, we use the inverse square potential

$$V(x,y,z)=\frac{c^2}{x^2+y^2+z^2}, \tag{4.3}$$

and solve the three-dimensional Schrödinger eigenvalue problem. We compare our result with [61].

Let the domain be a unit ball, we take $c=1/3$ and use the cut-off function $\phi(x,y,z)=1-x^2-y^2-z^2$. As presented in Table 4, the relative difference between these two methods for the first five eigenvalues is less than 4×10^{-4}. For the tenth eigenvalue, the highest eigenvalue provided in [61], the relative difference remains around 1×10^{-3}.

We also test the method in a 3-dimensional ring, i.e., $\Omega=\{(x,y,z)\in\mathbb{R}^3 \mid \frac{1}{2}\leqslant\sqrt{x^2+y^2+z^2}\leqslant 1\}$ and take $c=1/2$. The cut-off function is

$$\phi(x,y,z)=\Big(1-x^2-y^2-z^2\Big)\Big(x^2+y^2+z^2-\frac{1}{4}\Big). \tag{4.4}$$

We report the results in Table 5, which keep the difference of the first nine eigenvalues less than 2×10^{-3}.

Table 4 Estimates of the eigenvalues with potential (4.3) in a unit ball.

	$k=1$	$k=2$	$k=3$	$k=5$	$k=10$
Our	10.7873	20.6167	20.6184	33.5391	41.4362
[61]	10.7836	20.6206	20.6206	33.5352	41.3859
Rel. diff.	3.43×10^{-4}	1.89×10^{-4}	1.07×10^{-4}	1.16×10^{-4}	1.22×10^{-3}

Table 5 Estimates of the eigenvalues with potential (4.3) in a three-dimensional ring.

	$k=1$	$k=2$	$k=3$	$k=5$	$k=9$
Our	40.0149	43.7195	43.7281	51.1062	51.1355
[61]	39.9433	43.6545	43.6545	51.0341	51.0341
Rel. diff.	1.79×10^{-3}	1.49×10^{-3}	1.69×10^{-3}	1.41×10^{-3}	1.99×10^{-3}

5 Oracle inequality for the generalization error

In this section, we introduce an oracle inequality for the empirical loss. As a preparation, we firstly introduce concentration inequalities for ratio-type suprema of empirical processes. The study of ratio type empirical processes has a long history that goes back to the 1970s-1980s when certain classical function classes $\{\mathbf{1}_{(-\infty,t]}: t\in\mathbb{R}\}$ have been explored in great detail [67] and Alexander extended this theory to ratio type empirical processes indexed by VC classes of sets [2,3] in the late 1980s. Thereafter, there has been a great deal of work on the development of ratio type inequalities, primarily, in more specialized contexts of nonparametric statistics [24,52] and learning theory [8].

5.1 Concentration inequalities for normalized empirical processes

Let $\mathscr{F}$ be a class of real valued measurable functions[2)] taking values in $[0,1]$. Let X, X_1, X_2, … be *i.i.d.* random variables with distribution $\mathcal{P}$. We denote by $\mathcal{P}_n:=n^{-1}\sum_{i=1}^n\delta_{X_i}$ the empirical measure based on the sample $(X_1,\dots,X_n)$. Let $\mathcal{P}f=\mathbf{E}f(X)$ and $\mathrm{Var}_{\mathcal{P}}(f)=\mathcal{P}f^2-(\mathcal{P}f)^2$. Suppose that $\sigma_{\mathcal{P}}(f)$ is defined such that

$$\mathrm{Var}_{\mathcal{P}}(f)\leqslant\sigma^2_{\mathcal{P}}(f)\leqslant 1,\quad f\in\mathscr{F}.$$

2) In order to avoid measurability problems, we shall assume that the supremum over the class $\mathscr{F}$ or over any of the subclasses we consider is in fact a countable supremum. In this case we say that the class $\mathscr{F}$ is measurable.

In particular, $\sigma_{\mathcal{P}}(f)$ may be the standard deviation itself or equal to $\sqrt{\mathcal{P}f}$ because f takes values in $[0,1]$. Here, we present concentration inequalities for the supremum of the normalized empirical process

$$\sup_{f\in\mathscr{F},\sigma_{\mathcal{P}}(f)>r}\frac{|\mathcal{P}_nf-\mathcal{P}f|}{\sigma^2_{\mathcal{P}}(f)}$$

for some properly chosen cutoff $r\in(0,1)$. Define the random variable

$$\|\mathcal{P}_n-\mathcal{P}\|_{\mathscr{F}}:=\sup_{f\in\mathscr{F}}|\mathcal{P}_nf-\mathcal{P}f|=\sup_{f\in\mathscr{F}}\left|\frac{1}{n}\sum_{i=1}^n f\left(X_i\right)-\mathbf{E}f(X)\right|,$$

which measures the absolute deviation between the sample average $\mathcal{P}_nf$ and the population average $\mathcal{P}f$, uniformly over the class $\mathscr{F}$. For $0<r<s$, we define

$$\mathscr{F}(r):=\{f\in\mathscr{F}:\sigma_{\mathcal{P}}(f)\leqslant r\}\quad\text{ and }\quad\mathscr{F}(r,s]:=\mathscr{F}(s)\backslash\mathscr{F}(r).$$

For $q>1$ and $r<s\leqslant rq^l$ with $l\in\mathbb{N}$, let $\rho_j:=rq^j$ and define[3)]

$$K^{\mathscr{F}}_{n,q}(r,s]:=\max_{1\leqslant j\leqslant l}\frac{\mathbf{E}\,\|\mathcal{P}_n-\mathcal{P}\|_{\mathscr{F}(\rho_{j-1},\rho_j]}}{\rho^2_{j-1}}. \tag{5.1}$$

We recall a concentration inequality proved in [27, Lemma 2] when $\sigma_{\mathcal{P}}(f)=\sqrt{\mathcal{P}f}$.

Lemma 5.1. *[27, Lemma 2] For $t>0$,*

$$\mathbf{P}\left\{\sup_{f\in\mathscr{F}(r,s]}\left|\frac{\mathcal{P}_nf}{\mathcal{P}f}-1\right|\geqslant K_{n,q}(r,s]+\sqrt{\frac{2t}{nr^2}\left(q^2+2K_{n,q}(r,s]\right)}+\frac{t}{3nr^2}\right\}\leqslant\frac{q^2}{q^2-1}\frac{q}{t}\mathrm{e}^{-\mathrm{t}/\mathrm{q}}.$$

Proceeding along the same line that leads to [27, Lemma 2] , we may extend the above result to the more general $\sigma_{\mathcal{P}}(f)$. The proof is quite straightforward, and we omit the details.

Lemma 5.2. *For $t>0$,*

$$\mathbf{P}\left\{\sup_{f\in\mathscr{F}(r,s]}\frac{|\mathcal{P}_nf-\mathcal{P}f|}{\sigma^2_{\mathcal{P}}(f)}\geqslant K_{n,q}(r,s]+\sqrt{\frac{2t}{nr^2}\left(q^2+2K_{n,q}(r,s]\right)}+\frac{t}{3nr^2}\right\}\leqslant\frac{q^2}{q^2-1}\frac{q}{t}\mathrm{e}^{-\mathrm{t}/\mathrm{q}}.$$

Define $\mathcal{K}_n(\mathscr{F},r):=K^{\mathscr{F}}_{n,\sqrt{2}}(r,1]$. It follows from Lemma 5.2 that we deduce

Lemma 5.3. *For $0<\delta<2/\mathrm{e}$, with probability at least $1-\delta$,*

$$\sup_{f\in\mathscr{F},\sigma_{\mathcal{P}}(f)>r}\frac{|\mathcal{P}_nf-\mathcal{P}f|}{\sigma^2_{\mathcal{P}}(f)}<2\mathcal{K}_n(\mathscr{F},r)+\frac{5}{2}\sqrt{\frac{\ln(2/\delta)}{nr^2}}+\frac{2\ln(2/\delta)}{nr^2}.$$

Proof. Recall that $\sigma_{\mathcal{P}}(f)\leqslant 1$. Taking $s=1$, $q=\sqrt{2}$ and $t=\sqrt{2}\ln(2/\delta)$ in Lemma 5.2, we obtain

$$\frac{q^2}{q^2-1}\frac{q}{t}\mathrm{e}^{-\mathrm{t}/\mathrm{q}}=\frac{\delta}{\ln(2/\delta)}<\delta,$$

and with probability at least $1-\delta$,

$$\begin{aligned}\sup_{f\in\mathscr{F}(r,1]}\frac{|\mathcal{P}_nf-\mathcal{P}f|}{\sigma^2_{\mathcal{P}}(f)}&<\mathcal{K}_n(\mathscr{F},r)+2\sqrt{\frac{t}{nr^2}\mathcal{K}_n(\mathscr{F},r)}+2\sqrt{\frac{t}{nr^2}}+\frac{t}{3nr^2}\\&\leqslant 2\mathcal{K}_n(\mathscr{F},r)+2\sqrt{\frac{\sqrt{2}\ln(2/\delta)}{nr^2}}+\frac{4\sqrt{2}\ln(2/\delta)}{3nr^2}\\&\leqslant 2\mathcal{K}_n(\mathscr{F},r)+\frac{5}{2}\sqrt{\frac{\ln(2/\delta)}{nr^2}}+\frac{2\ln(2/\delta)}{nr^2},\end{aligned}$$

which completes the proof. □

3) When there is no ambiguity, we omit the superscript $\mathscr{F}$ of K.

5.2 Oracle inequality for the generalization error

Let $\mathcal{F}$ be some hypothesis class and $0<r<1/2$. We minimize $L_{k,n}(u)$ over $\mathcal{F}_{>r}=\{u\in\mathcal{F}:\mathcal{E}_2(u)>r^2\}$ with a minimizer $u_n=\arg\min_{u\in\mathcal{F}_{>r}}L_{k,n}(u)$. We aim to bound the energy excess $L_k(u_n)-\lambda_k$. For any $u_{\mathcal{F}}\in\mathcal{F}_{>r}$, we write

$$\begin{aligned}L_k\left(u_n\right)-\lambda_k=&L_k\left(u_n\right)-L_{k,n}\left(u_n\right)+L_{k,n}\left(u_n\right)-L_{k,n}\left(u_{\mathcal{F}}\right)\\&+L_{k,n}\left(u_{\mathcal{F}}\right)-L_k\left(u_{\mathcal{F}}\right)+L_k\left(u_{\mathcal{F}}\right)-\lambda_k.\end{aligned}\tag{5.2}$$

Note that $L_{k,n}\left(u_n\right)-L_{k,n}\left(u_{\mathcal{F}}\right)\leqslant 0$ because u_n is the minimizer of $L_{k,n}\left(u\right)$. Therefore,

$$\begin{aligned}L_k\left(u_n\right)-\lambda_k&\leqslant\Big(L_k(u_n)-L_{k,n}(u_n)\Big)+\Big(L_{k,n}\left(u_{\mathcal{F}}\right)-L_k\left(u_{\mathcal{F}}\right)\Big)+\Big(L_k\left(u_{\mathcal{F}}\right)-\lambda_k\Big)\\&=:T_1+T_2+T_3,\end{aligned}\tag{5.3}$$

where T_1 is the statistical error arising from the random approximation of the integrands, T_2 is the Monte Carlo error and T_3 is the approximation error due to restricting the minimizing $L_k(u)$ over $\mathcal{F}_{>r}$ instead of $H_0^1(\Omega)$.

Bounding T_1: We firstly decompose T_1 as

$$\begin{aligned}T_1&\leqslant\left|\frac{\mathcal{E}_{n,V}\left(u_n\right)}{\mathcal{E}_{n,2}\left(u_n\right)}-\frac{\mathcal{E}_V\left(u_n\right)}{\mathcal{E}_2\left(u_n\right)}\right|+\left|\frac{\mathcal{E}_{n,P}\left(u_n\right)}{\mathcal{E}_{n,2}\left(u_n\right)}-\frac{\mathcal{E}_P\left(u_n\right)}{\mathcal{E}_2\left(u_n\right)}\right|\\&\leqslant\left|\frac{\mathcal{E}_{n,V}\left(u_n\right)}{\mathcal{E}_V\left(u_n\right)}\frac{\mathcal{E}_2\left(u_n\right)}{\mathcal{E}_{n,2}\left(u_n\right)}-1\right|\frac{\mathcal{E}_V\left(u_n\right)}{\mathcal{E}_2\left(u_n\right)}+\frac{\mathcal{E}_P\left(u_n\right)\left|\mathcal{E}_2\left(u_n\right)-\mathcal{E}_{n,2}\left(u_n\right)\right|}{\mathcal{E}_2\left(u_n\right)\mathcal{E}_{n,2}\left(u_n\right)}+\frac{\left|\mathcal{E}_{n,P}\left(u_n\right)-\mathcal{E}_P\left(u_n\right)\right|}{\mathcal{E}_{n,2}\left(u_n\right)}\\&=:T_{11}+T_{12}+T_{13}.\end{aligned}$$

To bound T_{11}, T_{12} and T_{13}, we define

$$\begin{aligned}&\mathcal{G}_1:=\left\{g\ \mid\ g=u^2\text{ where }u\in\mathcal{F}\right\},\\&\mathcal{G}_2:=\left\{g\ \mid\ g=|\nabla u|^2+V|u|^2\text{ where }u\in\mathcal{F}\right\},\\&\mathcal{F}_j:=\{g\ \mid\ g=u\psi_j\text{ where }u\in\mathcal{F}\}\quad\text{for }j=1,2,\ldots,k-1.\end{aligned}\tag{5.4}$$

We assume that the set $\mathcal{F}$ satisfies $\sup_{u\in\mathcal{F}}\|u\|_{L^\infty(\Omega)}\leqslant M_{\mathcal{F}}$ so that $\sup_{g\in\mathcal{G}_1}\|g\|_{L^\infty(\Omega)}\leqslant M_{\mathcal{F}}^2$. Assume further that $\sup_{g\in\mathcal{G}_2}\|g\|_{L^\infty(\Omega)}\leqslant M_{\mathcal{G}_2}$ and $\|\psi_j\|_{L^\infty(\Omega)}\leqslant\mu_j$ for each j. So $\sup_{g\in\mathcal{F}_j}\|g\|_{L^\infty(\Omega)}\leqslant\mu_jM_{\mathcal{F}}$. In what follows, we shall derive the high probability bounds for T_{11}, T_{12} and T_{13} by Lemma 5.3. To this end, we rescale the function classes $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{F}_j$ so that their elements take values in $[0,1]$.

We firstly derive a high probability bound for $\mathcal{E}_{n,2}\left(u_n\right)/\mathcal{E}_2\left(u_n\right)$. For the rescaled set $\mathcal{G}_1/M_{\mathcal{F}}^2$ [4)], we take $\sigma_{\mathcal{P}}(f)=\sqrt{\mathcal{P}f}$ and for $n\in\mathbb{N}$ and $0<\delta<1/3$, define

$$\xi_1(n,r,\delta):=2\mathcal{K}_n(\mathcal{G}_1/M_{\mathcal{F}}^2,r/M_{\mathcal{F}})+\frac{5M_{\mathcal{F}}}{2}\sqrt{\frac{\ln(2/\delta)}{nr^2}}+\frac{2M_{\mathcal{F}}^2\ln(2/\delta)}{nr^2},\tag{5.5}$$

and the event

$$A_1(n,r,\delta):=\left\{\sup_{u\in\mathcal{F},\mathcal{E}_2(u)>r^2}\left|\frac{\mathcal{E}_{n,2}\left(u\right)}{\mathcal{E}_2\left(u\right)}-1\right|<\xi_1(n,r,\delta)\right\}.$$

Applying Lemma 5.3 to $\mathcal{G}_1/M_{\mathcal{F}}^2$, we get

$$\mathbf{P}\left[A_1(n,r,\delta)\right]\geqslant 1-\delta.\tag{5.6}$$

Recall that $\mathcal{E}_2(u_n)>r^2$. So if $\xi_1(n,r,\delta)<1$, then, on the event $A_1(n,r,\delta)$,

$$T_{12}\leqslant\frac{\xi_1(n,r,\delta)}{1-\xi_1(n,r,\delta)}\frac{\mathcal{E}_P\left(u_n\right)}{\mathcal{E}_2\left(u_n\right)}.\tag{5.7}$$

4) In this paper, $a\mathscr{F}+b:=\{af+b:f\in\mathscr{F}\}$ where $\mathscr{F}$ is a set of functions and $a,b\in\mathbb{R}$ are some constants.

To bound T_{11}, it follows from $\mathcal{E}_2(u) > r^2$ that $\mathcal{E}_V(u) > \lambda_1 r^2$. For the rescaled set $\mathcal{G}_2/M_{\mathcal{G}_2}$, we take $\sigma_{\mathcal{P}}(f) = \sqrt{\mathcal{P}f}$ and define

$$\xi_2(n,r,\delta) := 2\mathcal{K}_n(\mathcal{G}_2/M_{\mathcal{G}_2}, r\sqrt{\lambda_1/M_{\mathcal{G}_2}}) + \frac{5}{2}\sqrt{\frac{M_{\mathcal{G}_2}\ln(2/\delta)}{n\lambda_1 r^2}} + \frac{2M_{\mathcal{G}_2}\ln(2/\delta)}{n\lambda_1 r^2}. \tag{5.8}$$

Define the event

$$A_2(n,r,\delta) := \left\{ \sup_{u\in\mathcal{F},\mathcal{E}_V(u)>\lambda_1 r^2} \left| \frac{\mathcal{E}_{n,V}(u)}{\mathcal{E}_V(u)} - 1 \right| < \xi_2(n,r,\delta) \right\}.$$

Applying Lemma 5.3 to $\mathcal{G}_2/M_{\mathcal{G}_2}$, we obtain

$$\mathbf{P}\left[A_2(n,r,\delta)\right] \geqslant 1-\delta. \tag{5.9}$$

Therefore, if $\xi_1(n,r,\delta) < 1$, then, on the event $A_1(n,r,\delta)\cap A_2(n,r,\delta)$, there holds

$$T_{11} \leqslant \left(\frac{1+\xi_2(n,r,\delta)}{1-\xi_1(n,r,\delta)} - 1\right)\frac{\mathcal{E}_V(u_n)}{\mathcal{E}_2(u_n)} = \frac{\xi_1(n,r,\delta)+\xi_2(n,r,\delta)}{1-\xi_1(n,r,\delta)}\frac{\mathcal{E}_V(u_n)}{\mathcal{E}_2(u_n)}. \tag{5.10}$$

To bound T_{13}, we rescale $\mathcal{F}_j$ to $\mathcal{F}_j/(2\mu_j M_{\mathcal{F}}) + 1/2$. For any $g\in\mathcal{F}_j$, $g = u\psi_j$,

$$\operatorname{Var}\left(\frac{g}{2\mu_j M_{\mathcal{F}}} + \frac{1}{2}\right) \leqslant \frac{1}{4}\mathbf{E}\left|\frac{u\psi_j}{\mu_j M_{\mathcal{F}}}\right|^2 \leqslant \frac{1}{4}\mathbf{E}\left|\frac{u\psi_j}{\mu_j M_{\mathcal{F}}}\right| \leqslant \frac{\|u\|_{L^2(\Omega)}}{4\mu_j M_{\mathcal{F}}},$$

where we have used $|u\psi_j| \leqslant \mu_j M_{\mathcal{F}}$ in the second inequality and the fact $\|\psi_j\|_{L^2(\Omega)} = 1$ in the last inequality. Therefore, we may take

$$\sigma^2_{\mathcal{P}}(f) = \frac{\|u\|_{L^2(\Omega)}}{4\mu_j M_{\mathcal{F}}} \quad \text{for any } f := \frac{u\psi_j}{2\mu_j M_{\mathcal{F}}} + \frac{1}{2} \in (2\mu_j M_{\mathcal{F}})^{-1}\mathcal{F}_j + \frac{1}{2}. \tag{5.11}$$

Note that $\mathcal{E}_2(u) > r^2$ implies $\|u\|_{L^2(\Omega)} > r$. For all $1\leqslant j\leqslant k-1$, we define

$$\xi_{3,j}(n,r,\delta) := 2\mathcal{K}_n\left(\frac{\mathcal{F}_j}{2\mu_j M_{\mathcal{F}}} + \frac{1}{2}, \sqrt{\frac{r}{4\mu_j M_{\mathcal{F}}}}\right) + 5\sqrt{\frac{\mu_j M_{\mathcal{F}}\ln(2k/\delta)}{nr}} + \frac{8\mu_j M_{\mathcal{F}}\ln(2k/\delta)}{nr},$$

and $\xi_3(n,r,\delta) := \max_{1\leqslant j\leqslant k-1}\xi_{3,j}(n,r,\delta)$. Notice that for $f\in(2\mu_j M_{\mathcal{F}})^{-1}\mathcal{F}_j + 1/2$,

$$\sup_{\sigma_{\mathcal{P}}(f)>\sqrt{r/4\mu_j M_{\mathcal{F}}}} \frac{|\mathcal{P}_n f - \mathcal{P}f|}{\sigma^2_{\mathcal{P}}(f)} = \sup_{u\in\mathcal{F},\|u\|_{L^2(\Omega)}>r} \frac{2\left|\mathcal{P}_n(u\psi_j) - \langle u,\psi_j\rangle\right|}{\|u\|_{L^2(\Omega)}}.$$

For each $1\leqslant j\leqslant k-1$, we define the events

$$A_{3,j}(n,r,\delta) := \left\{ \sup_{u\in\mathcal{F},\|u\|_{L^2(\Omega)}>r} \frac{2\left|\mathcal{P}_n(u\psi_j) - \langle u,\psi_j\rangle\right|}{\|u\|_{L^2(\Omega)}} < \xi_{3,j}(n,r,\delta) \right\},$$

and $A_3(n,r,\delta) := \bigcap_{j=1}^{k-1} A_{3,j}(n,r,\delta)$. Applying Lemma 5.3 to $(2\mu_j M_{\mathcal{F}})^{-1}\mathcal{F}_j + 1/2$, we get $\mathbf{P}\left[A_{3,j}(n,r,\delta)\right] \geqslant 1-\delta/k$. Hence,

$$\mathbf{P}\left[A_3(n,r,\delta)\right] \geqslant 1 - \frac{\delta}{k}(k-1) \geqslant 1-\delta. \tag{5.12}$$

Using $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ for $a,b\in\mathbb{R}$, on event $A_3(n,r,\delta)$, we obtain

$$\begin{aligned}
\frac{|\mathcal{E}_{n,P}(u_n) - \mathcal{E}_P(u_n)|}{\beta\mathcal{E}_2(u_n)} &\leqslant \sum_{j=1}^{k-1}\left[\frac{|\mathcal{P}_n(u_n\psi_j) - \mathcal{P}(u_n\psi_j)|^2}{\|u_n\|^2_{L^2(\Omega)}} + \frac{2|\langle u_n,\psi_j\rangle|}{\|u_n\|_{L^2(\Omega)}}\frac{|\mathcal{P}_n(u_n\psi_j) - \mathcal{P}(u_n\psi_j)|}{\|u_n\|_{L^2(\Omega)}}\right] \\
&\leqslant \sum_{j=1}^{k-1}\left[\frac{\xi^2_{3,j}}{4} + \xi_{3,j}\frac{|\langle u_n,\psi_j\rangle|}{\|u_n\|_{L^2(\Omega)}}\right] \\
&\leqslant \sum_{j=1}^{k-1}\frac{\xi^2_{3,j}}{4} + \left(\sum_{j=1}^{k-1}\xi^2_{3,j}\right)^{1/2}\left(\sum_{j=1}^{k-1}\frac{|\langle u_n,\psi_j\rangle|^2}{\|u_n\|^2_{L^2(\Omega)}}\right)^{1/2} \\
&\leqslant \frac{k}{4}\xi_3(n,r,\delta)^2 + \sqrt{k}\,\xi_3(n,r,\delta).
\end{aligned} \tag{5.13}$$

Hence, on event $A_1(n,r,\delta)\cap A_3(n,r,\delta)$, if $\xi_1(n,r,\delta)<1$, then

$$T_{13}=\frac{\mathcal{E}_2\left(u_n\right)}{\mathcal{E}_{n,2}\left(u_n\right)}\frac{\left|\mathcal{E}_{n,P}\left(u_n\right)-\mathcal{E}_P\left(u_n\right)\right|}{\mathcal{E}_2\left(u_n\right)}\leqslant\frac{\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right). \tag{5.14}$$

It follows from (5.7), (5.10) and (5.14) that if $\xi_1(n,r,\delta)<1$, then, within event $\bigcap_{i=1}^3 A_i(n,r,\delta)$, we obtain

$$\begin{aligned}T_1&\leqslant\frac{\xi_1+\xi_2}{1-\xi_1}\frac{\mathcal{E}_V\left(u_n\right)}{\mathcal{E}_2\left(u_n\right)}+\frac{\xi_1}{1-\xi_1}\frac{\mathcal{E}_P\left(u_n\right)}{\mathcal{E}_2\left(u_n\right)}+\frac{\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right)\\&\leqslant\frac{\xi_1+\xi_2}{1-\xi_1}L_k\left(u_n\right)+\frac{\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right),\end{aligned} \tag{5.15}$$

while it follows from (5.6), (5.9) and (5.12) that

$$\mathbf{P}\left[A_1(n,r,\delta)\cap A_2(n,r,\delta)\cap A_3(n,r,\delta)\right]\geqslant 1-3\delta. \tag{5.16}$$

Bounding T_2. Similar to bounding T_1, it follows from $\|u_{\mathcal{F}}\|_{L^2(\Omega)}>r$ that if $\xi_1(n,r,\delta)<1$, then, within event $\bigcap_{i=1}^3 A_i(n,r,\delta)$, we get

$$T_2\leqslant\frac{\xi_1+\xi_2}{1-\xi_1}L_k\left(u_{\mathcal{F}}\right)+\frac{\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right). \tag{5.17}$$

Bounding T_3. A combination of the bounds (5.3), (5.15) and (5.17) leads to

$$\begin{aligned}L_k\left(u_n\right)-\lambda_k\leqslant&\frac{\xi_1+\xi_2}{1-\xi_1}\left(L_k\left(u_n\right)-\lambda_k\right)+\frac{1+\xi_2}{1-\xi_1}\left(L_k\left(u_{\mathcal{F}}\right)-\lambda_k\right)+2\lambda_k\frac{\xi_1+\xi_2}{1-\xi_1}\\&+\frac{2\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right).\end{aligned}$$

If $2\xi_1+\xi_2<1$, then

$$L_k\left(u_n\right)-\lambda_k\leqslant\frac{\left(1+\xi_2\right)\left(L_k\left(u_{\mathcal{F}}\right)-\lambda_k\right)+2\lambda_k\left(\xi_1+\xi_2\right)+\beta\left(k\xi_3^2/2+2\sqrt{k}\xi_3\right)}{1-2\xi_1-\xi_2}.$$

Combining the above estimate with (5.16), we obtain

Theorem 5.4. *Let $u_n=\arg\min_{u\in\mathcal{F}_{>r}}L_{k,n}(u)$, $0<\delta<1/3$ and let $\{\xi_i(n,r,\delta)\}_{i=1}^3$ be defined in (5.5), (5.8), (5.2), respectively. Assume that $2\xi_1+\xi_2\leqslant 1/2$ and $u_{\mathcal{F}}\in\mathcal{F}_{>r}$. Then, with probability at least $1-3\delta$,*

$$L_k\left(u_n\right)-\lambda_k\leqslant 4\lambda_k\left(\xi_1+\xi_2\right)+\beta\left(k\xi_3^2+4\sqrt{k}\xi_3\right)+3\left(L_k\left(u_{\mathcal{F}}\right)-\lambda_k\right).$$

6 Approximation theorem for sine spectral Barron functions

In this section, we study the properties of the sine spectral Barron functions on the hypercube as well as the neural network approximation.

6.1 Preliminaries

We start with some preliminary results about the sine functions $\mathfrak{S}=\{\Phi_k\}_{k\in\mathbb{N}_+^d}$. It is clear that the set $\mathfrak{S}$ forms an orthogonal basis of $L^2(\Omega)$ and $H_0^1(\Omega)$. Given $u\in L^2(\Omega)$, let $\{\hat{u}(k)\}_{k\in\mathbb{N}_+^d}$ be the Fourier coefficients of u against the basis $\{\Phi_k\}_{k\in\mathbb{N}_+^d}$, hence

$$u(x)=\sum_{k\in\mathbb{N}_+^d}\hat{u}(k)\Phi_k(x)\quad\text{and}\quad\|u\|_{L^2(\Omega)}^2=\sum_{k\in\mathbb{N}_+^d}2^{-d}|\hat{u}(k)|^2,$$

where we have used $\langle\Phi_k,\Phi_k\rangle_{L^2(\Omega)}=2^{-d}$. A straightforward calculation yields that for $u\in H_0^1(\Omega)$,

$$\|u\|_{H^1(\Omega)}^2=\sum_{k\in\mathbb{N}_+^d}2^{-d}\left(1+\pi^2|k|^2\right)|\hat{u}(k)|^2.$$

Lemma 6.1. *The following embedding hold:*

$$\mathfrak{B}^0(\Omega) \hookrightarrow L^\infty(\Omega) \qquad \textit{and} \qquad \mathfrak{B}^2(\Omega) \hookrightarrow H_0^1(\Omega).$$

We postpone its proof to Appendix B.1.

6.2 Sine Spectral Barron Space and the Neural Network Approximation

The main results in this part are summarized in the following two propositions.

Proposition 6.2. *Assume that $u \in \mathfrak{B}^{s+1}(\Omega)$ for some $s \geqslant 0$. Then, u/φ admits the representation*

$$\frac{u(x)}{\varphi(x)} = \sum_{(k,i)\in\Gamma} \hat{v}(k,i)\cos(k_i\pi x_i) \prod_{\substack{1\leqslant j\leqslant d\\ j\neq i}} \sin(k_j\pi x_j), \tag{6.1}$$

where $\Gamma = \left\{(k,i) \mid k \in \mathbb{N}_0^d,\ 1 \leqslant i \leqslant d,\ (k+e_i) \in \mathbb{N}_+^d\right\}$ and e_i is the i-th canonical basis.

Moreover, the coefficients $\hat{v}(k,i)$ satisfy

$$\sum_{(k,i)\in\Gamma} (1+\pi^s|k|_1^s)\,|\hat{v}(k,i)| \leqslant \|u\|_{\mathfrak{B}^{s+1}(\Omega)}. \tag{6.2}$$

Roughly speaking, the above result indicates that u/φ lies in a spectral Barron space of one order lower than u. Using this proposition, we may prove a preliminary H^1 approximation result for functions in the sine spectral Barron space. Denote by $\mathrm{conv}(G)$ the convex hull of a set G, and denote by $\overline{G}$ the H^1-closure of G. Let

$$\Gamma_1 = \left\{k \in \mathbb{Z}^d \ \mid\ \exists 1 \leqslant i \leqslant d, ((|k_1|, |k_2|, \cdots, |k_d|), i) \in \Gamma\right\}.$$

Note that $k \in \Gamma_1$ if and only if k has at most one zero component.

Proposition 6.3. *For $s \geqslant 0$, define*

$$\mathcal{F}_s(B) := \left\{\frac{\gamma}{1+\pi^s|k|_1^s} f(\pi(k\cdot x+b)) \ \mid\ |\gamma| \leqslant B, b \in \{0,1\}, k \in \Gamma_1\right\},$$

where $f(x) = \cos x$ if d is odd and $f(x) = \sin x$ if d is even. Then, for any $u \in \mathfrak{B}^{s+1}(\Omega)$ with $s \geqslant 1$, $u \in \overline{\mathrm{conv}(\varphi\mathcal{F}_s(B_u))}$ with $B_u = \|u\|_{\mathfrak{B}^{s+1}(\Omega)}$ and there exists v_m which is a convex combination of m functions in $\mathcal{F}_s(B_u)$ such that

$$\|u - \varphi v_m\|_{H^1(\Omega)} \leqslant \sqrt{\frac{6}{m}}B_u.$$

When $d > 1$, the constant $\sqrt{6}$ may be replaced by 2.

We postpone the proof of Proposition 6.2 and Proposition 6.3 to Appendix B.3. We exploit the seminal result for nonlinear approximation known as Maurey's method to prove the approximation bounds.

Lemma 6.4. *[6, 55] If $\bar{f}$ is in the closure of the convex hull of a set G in a Hilbert space, with $\|g\| \leqslant b$ for each $g \in G$, then for every $m \geqslant 1$, there is an f_m in the convex hull of m points in G such that*

$$\left\|\bar{f} - f_m\right\|^2 \leqslant \frac{b^2}{m}.$$

6.3 Reduction to ReLU and Softplus Activation Functions

We have found that if $u \in \mathfrak{B}^{s+1}$ with $s \geqslant 1$, u lies in the bounded set $\overline{\mathrm{conv}(\varphi\mathcal{F}_s(B_u))} \subset H^1(\Omega)$ with $B_u = (1+2/\pi)\,\|u\|_{\mathfrak{B}^{s+1}}$. Define the function classes

$$\begin{aligned}
\mathcal{F}_{\mathrm{ReLU}}(B) &:= \{c + \gamma\,\mathrm{ReLU}(w\cdot x - t) \ \mid\ |c| \leqslant B, |w|_1 = 1, |t| \leqslant 1, |\gamma| \leqslant 4B\},\\
\mathcal{F}_{\mathrm{SP}_\tau}(B) &:= \{c + \gamma\mathrm{SP}_\tau(w\cdot x - t) \ \mid\ |c| \leqslant B, |w|_1 = 1, |t| \leqslant 1, |\gamma| \leqslant 4B\}.
\end{aligned} \tag{6.3}$$

In this subsection, we aim to prove that each function in $\varphi\mathcal{F}_s(B)$ lies in $\overline{\mathrm{conv}(\varphi\mathcal{F}_{\mathrm{ReLU}}(B))}$ and $\overline{\mathrm{conv}(\varphi\mathcal{F}_{\mathrm{SP}_\tau}(B))}$ when $s \geqslant 2$, which together with Lemma 6.4 yields Theorem 3.1 and Theorem 3.2.

By Lemma B.2, to prove $\varphi\mathcal{F}_s(B)$ lies in $\overline{\mathrm{conv}(\varphi\mathcal{F}_{\mathrm{ReLU}}(B))}$ and $\overline{\mathrm{conv}(\varphi\mathcal{F}_{\mathrm{SP}_\tau}(B))}$, one only needs to show $\mathcal{F}_s(B)$ lies in $\overline{\mathrm{conv}(\mathcal{F}_{\mathrm{ReLU}}(B))}$ and $\overline{\mathrm{conv}(\mathcal{F}_{\mathrm{SP}_\tau}(B))}$, respectively. Notice that every function in $\mathcal{F}_s(B)$ is a composition of a function g defined on $[-1,1]$ by

$$g(z) = \begin{cases} \frac{\gamma}{1+\pi^s|k|_1^s}\cos(\pi(|k|_1 z + b)), & d \text{ is odd},\\ \frac{\gamma}{1+\pi^s|k|_1^s}\sin(\pi(|k|_1 z + b)), & d \text{ is even}, \end{cases} \tag{6.4}$$

where $k \in \Gamma_1$, $|\gamma| \leqslant B$, $b \in \{0,1\}$, and a linear function $z = w\cdot x$ with $w = k/|k|_1$ or $z = x_1$ in case $k = 0$. When $s \geqslant 2$, it is clear that $g \in C^2([-1,1])$ and g satisfies

$$\|g^{(r)}\|_{L^\infty([-1,1])} \leqslant |\gamma| \leqslant B, \qquad \text{for } r = 0,1,2.$$

The uniform boundness of $\|g\|_{W^{2,\infty}([-1,1])}$ for all k is key to the proof. In addition, we observe that $g'(0) = 0$ if d is odd and $g'(\frac{1}{2|k|_1}) = 0$ if d is even. Proceeding along the same line as [47, Section 4.3], we prove that functions in $\mathcal{F}_s(B)$ can be well approximated by two-layer ReLU networks. Compared with [47, Lemma 4.5], Lemma 6.5 handles both sine and cosine cases.

Lemma 6.5. *Let $g \in C^2([-1,1])$ with $\left\|g^{(r)}\right\|_{L^\infty([-1,1])} \leqslant B$ for $r = 0,1,2$. Assume that $g'(\rho) = 0$ for some $\rho \in [0,1/2]$. Let $\{z_j\}_{j=0}^{2m}$ be a partition of $[-1,1]$ with $z_0 = -1$, $z_m = \rho$, $z_{2m} = 1$ and $z_{j+1} - z_j = h_1 = (\rho+1)/m$ for each $j = 0,\cdots,m-1$; $z_{j+1} - z_j = h_2 = (1-\rho)/m$ for each $j = m,\cdots,2m-1$. Then there exists a two-layer ReLU network*

$$g_m(z) = c + \sum_{i=1}^{2m} a_i \,\mathrm{ReLU}\,(\epsilon_i z - b_i)\,, \quad z \in [-1,1] \tag{6.5}$$

with $c = g(\rho)$, $b_i \in [-1,1]$ and $\epsilon_i \in \{\pm 1\}, i = 1,\cdots,2m$, such that

$$\|g - g_m\|_{W^{1,\infty}([-1,1])} \leqslant \frac{2B}{m}. \tag{6.6}$$

Moreover, we have $|c| \leqslant B$, $|a_i| \leqslant 2Bh_1$ if $i < m$, $|a_m| \leqslant Bh_1$, $|a_{m+1}| \leqslant Bh_2$ and $|a_i| \leqslant 2Bh_2$ if $i > m+1$ so that $\sum_{i=1}^{2m}|a_i| \leqslant 4B$.

Given Proposition 6.3 and Lemma 6.5, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. By Proposition 6.3, for any $u \in \mathfrak{B}^s(\Omega)$ with $s \geqslant 3$, $u \in \overline{\mathrm{conv}(\varphi\mathcal{F}_{s-1}(B_u))}$ with $B_u = \|u\|_{\mathfrak{B}^s(\Omega)}$. Since $s - 1 \geqslant 2$, a combination of Lemma 6.5 and Lemma B.2 yields that the set $\varphi\mathcal{F}_{s-1}(B_u)$ lies in the H^1-closure of $\mathrm{conv}(\varphi\mathcal{F}_{\mathrm{ReLU}}(B_u))$. Hence, $u \in \overline{\mathrm{conv}(\varphi\mathcal{F}_{\mathrm{ReLU}}(B_u))}$. Notice that when $|w|_1 = 1$, for constants a, $b \geqslant 0$,

$$\begin{aligned}\int_\Omega (a + bw\cdot x)^2\,\mathrm{d}\,x &= a^2 + ab\sum_{i=1}^d w_i + b^2\left(\frac{1}{3}\sum_{i=1}^d w_i^2 + \frac{1}{4}\sum_{i\neq j} w_i w_j\right)\\ &\leqslant a^2 + ab|w|_1 + b^2\left(\frac{1}{4}|w|_1^2 + \frac{1}{12}|w|_2^2\right)\\ &\leqslant a^2 + ab + b^2/3.\end{aligned} \tag{6.7}$$

By (6.7), for each $v \in \mathcal{F}_{\mathrm{ReLU}}(B)$, we get

$$\|v\|_{L^2(\Omega)}^2 \leqslant \int_\Omega [B + 4B(w\cdot x + 1)]^2\,\mathrm{d}\,x \leqslant 51B^2.$$

Since $|v(x)| \leqslant (5 + 4w\cdot x)B$, $\|\nabla v\|_{L^\infty(\Omega)} \leqslant |\gamma| \leqslant 4B$, by Lemma B.1 and the inequality (6.7), we obtain

$$\|\nabla(\varphi v)\|_{L^2(\Omega)}^2 \leqslant \int_\Omega (|\nabla\varphi|\,|v| + \varphi\,|\nabla v|)^2\,\mathrm{d}\,x \leqslant B^2\int_\Omega [\pi(5 + 4w\cdot x) + 4]^2\,\mathrm{d}\,x \leqslant 689B^2.$$

Hence, $\|\varphi v\|_{H^1(\Omega)}^2 \leqslant 740B^2$. Therefore, the H^1-norm of each function in $\varphi\mathcal{F}_{\mathrm{ReLU}}(B)$ can be bounded by $28B$. Theorem 3.1 follows immediately from Lemma 6.4. $\square$

Lemma 6.6 shows that functions in $\mathcal{F}_s(B)$ are also well approximated by two-layer Softplus networks.

Lemma 6.6. *Under the same assumption of Lemma 6.5, there exists*

$$g_{\tau,m}(z) = c + \sum_{i=1}^{2m} a_i \mathrm{SP}_\tau\left(\epsilon_i z - b_i\right), \quad z \in [-1,1] \tag{6.8}$$

with $\tau > 0$, $c = g(\rho)$, $b_i \in [-1,1]$ and $\epsilon_i \in \{\pm 1\}$, $i = 1, \cdots, 2m$ such that

$$\|g - g_{\tau,m}\|_{W^{1,\infty}([-1,1])} \leqslant 4B\tau^{-1}(1+\tau^{-1}). \tag{6.9}$$

Moreover, the bounds for $|c|$, $|a_i|$ and $\sum_{i=1}^{2m}|a_i|$ are valid as in Lemma 6.5.

Now we are ready to prove Theorem 3.2. It follows from (B.8c) that

$$\sup_{f \in \mathcal{F}_{\mathrm{SP}_\tau(B)}} \|f\|_{H^1(\Omega)} \leqslant B + 4B\|\mathrm{SP}_\tau\|_{W^{1,\infty}([-2,2])} \leqslant 13B + 4B\tau^{-1}. \tag{6.10}$$

Proof of Theorem 3.2. According to Proposition 6.3, for any $u \in \mathfrak{B}^s(\Omega)$ with $s \geqslant 3$, $u \in \mathrm{conv}(\varphi\mathcal{F}_{s-1}(B))$ with $B = \|u\|_{\mathfrak{B}^s(\Omega)}$. Note that each function in $\mathcal{F}_{s-1}(B)$ with $s \geqslant 3$ is a composition of the multivariate linear function $z = w \cdot x$ with $|w|_1 = 1$ and the univariate function $g(z)$ defined in (6.4) such that $g'(\rho) = 0$ for some $\rho \in [0, 1/2]$ and $\left\|g^{(r)}\right\|_{L^\infty([-1,1])} \leqslant B$ for $r = 0, 1, 2$. By Lemma 6.6, such g may be approximated by $g_{\tau,m}$ which lies in the convex hull of the set of functions $\{c + \gamma\mathrm{SP}_\tau(\epsilon z - b) : |c| \leqslant B, \epsilon \in \{\pm 1\}, |b| \leqslant 1, \gamma \leqslant 4B\}$. Moreover, $\|g - g_{\tau,m}\|_{W^{1,\infty}([-1,1])} \leqslant 4B(1+\tau)/\tau^2$. As a consequence, we have

$$\|g(w\cdot x) - g_{\tau,m}(w \cdot x)\|_{H^1(\Omega)} \leqslant \|g - g_{\tau,m}\|_{W^{1,\infty}([-1,1])} \leqslant 4B\tau^{-1}(1+\tau^{-1}).$$

By Lemma B.2, there exists a function v_τ in the convex hull of $\mathcal{F}_{\mathrm{SP}_\tau}(B)$ such that

$$\|u - \varphi v_\tau\|_{H^1(\Omega)} \leqslant 4\sqrt{21}B\tau^{-1}(1+\tau^{-1}).$$

Thanks to Lemma 6.4 and the bound (6.10), there exists $v_m \in \mathcal{F}_{\mathrm{SP}_\tau,m}(B)$, which is a convex combination of m functions in $\mathcal{F}_{\mathrm{SP}_\tau}(B)$ such that

$$\|\varphi v_\tau - \varphi v_m\|_{H^1(\Omega)} \leqslant \sqrt{21}\,\|v_\tau - v_m\|_{H^1(\Omega)} \leqslant \sqrt{\frac{21}{m}}B\left(13 + 4\tau^{-1}\right),$$

where the first inequality follows from Lemma B.2.

Combining the last two inequalities and setting $\tau = 9\sqrt{m}$, we obtain (6.9). $\square$

6.4 Bounding the approximation error

The following theorem bounds the approximation error in (5.3) when $\mathcal{F} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}$. We postpone the proof of Theorem 6.7 to Appendix B.5.

Theorem 6.7. *Under Assumptions 2.1 and 3.3, let $B_{u^*} = \|u^*\|_{\mathfrak{B}^s}$ and $v_m \in \mathcal{F}_{\mathrm{SP}_\tau,m}(B_{u^*})$ be defined in Theorem 3.2. Assume in addition that $\eta(B_{u^*}, m) := 64B_{u^*}/\sqrt{m} \leqslant 1/2$. Then,*

$$L_k(\varphi v_m) - L_k(u^*) \leqslant (3\max\{1, V_{\max}\} + 7\lambda_k + 5\beta)\,\eta(B_{u^*}, m).$$

7 Statistical error

By Theorem 5.4, in order to bound the statistical error, we need to control

$$\mathcal{K}_n\left(\frac{\mathcal{G}_1}{M_{\mathcal{F}}^2}, \frac{r}{M_{\mathcal{F}}}\right), \quad \mathcal{K}_n\left(\frac{\mathcal{G}_2}{M_{\mathcal{G}_2}}, r\sqrt{\frac{\lambda_1}{M_{\mathcal{G}_2}}}\right) \quad \text{and} \quad \mathcal{K}_n\left(\frac{\mathcal{F}_j}{2\mu_j M_{\mathcal{F}}} + \frac{1}{2}, \sqrt{\frac{r}{4\mu_j M_{\mathcal{F}}}}\right), \quad 1 \leqslant j \leqslant k-1. \tag{7.1}$$

To this end, we firstly bound the covering numbers of the properly rescaled function classes in §7.1. Secondly, we derive inequalities to bound the quantity $K_{n,q}(r,s]$ defined in (5.1) with respect to covering numbers in §7.2. Finally, we obtain the bounds for the quantities in (7.1).

7.1 Bounding the covering numbers

For fixed positive constants C, Γ, W and T, we consider the set of two-layer neural networks

$$\widetilde{\mathcal{F}}_m=\left\{v_\theta(x)=c+\sum_{i=1}^m\gamma_i\phi\left(w_i\cdot x+t_i\right)\ :\ x\in\Omega,|c|\leqslant C,\sum_{i=1}^m|\gamma_i|\leqslant\Gamma,|w_i|_1\leqslant W,|t_i|\leqslant T\right\},\tag{7.2}$$

where ϕ is the activation function, $\theta=(c,\{\gamma_i\}_{i=1}^m,\{w_i\}_{i=1}^m,\{t_i\}_{i=1}^m)$ denotes the parameters of the two-layer neural network. Denote the parameter space

$$\Theta=\Theta_c\times\Theta_\gamma\times\Theta_w\times\Theta_t=[-C,C]\times B_1^m(\Gamma)\times\left(B_1^d(W)\right)^m\times[-T,T]^m.$$

We consider the set Θ endowed with the metric ρ defined for $\theta=(c,\gamma,w,t),\ \theta'=(c',\gamma',w',t')$ in Θ by

$$\rho_\Theta\left(\theta,\theta'\right)=\max\left\{\left|c-c'\right|,\left|\gamma-\gamma'\right|_1,\max_i\left|w_i-w_i'\right|_1,\left\|t-t'\right\|_\infty\right\}.\tag{7.3}$$

Assume that ϕ satisfies the following assumption, which is valid for the Softplus activation function.

Assumption 7.1. $\phi\in C^2(\mathbb{R})$ and ϕ (resp. ϕ', the derivative of ϕ) is L-Lipschitz (resp. is L'-Lipschitz) for some L, $L'>0$. Moreover, there exist positive constants $\phi_{\max}$ and $\phi'_{\max}$ such that

$$\sup_{w\in\Theta_w,t\in\Theta_t,x\in\Omega}|\phi(w\cdot x+t)|\leqslant\phi_{\max}\quad\text{and}\quad\sup_{w\in\Theta_w,t\in\Theta_t,x\in\Omega}|\phi'(w\cdot x+t)|\leqslant\phi'_{\max}.$$

Example 7.2. Let $\Theta_w=\left(B_1^d(1)\right)^m$ and $\Theta_t=[-1,1]^m$. It is clear that $\left\|\mathrm{SP}_\tau'\right\|_{L^\infty(\mathbb{R})}\leqslant1$ and $\left\|\mathrm{SP}_\tau''\right\|_{L^\infty(\mathbb{R})}\leqslant\tau$, hence SP_τ satisfies Assumption 7.1 with

$$L=\phi'_{\max}=1,\quad L'=\tau,\quad\text{and}\quad\phi_{\max}\leqslant2+1/\tau.\tag{7.4}$$

Example 7.3. The activation function tanh satisfies Assumption 7.1 with

$$L=\phi'_{\max}=1,\ L'=4\sqrt{3}/9\quad\text{and}\quad\phi_{\max}=1.$$

Let (E,ρ) be a metric space with metric ρ. A δ-cover of a set $A\subset E$ with respect to ρ is a collection of points $\{x_1,\cdots,x_n\}\subset A$ such that for every $x\in A$, there exists $i\in\{1,\cdots,n\}$ such that $\rho\left(x,x_i\right)\leqslant\delta$. The δ-covering number $\mathcal{N}(\delta,A,\rho)$ is the cardinality of the smallest δ cover of the set A with respect to the metric ρ. Equivalently, the δ-covering number $\mathcal{N}(\delta,A,\rho)$ is the minimal number of balls $B_\rho(x,\delta)$ of radius δ required to cover A.

Let Q be any probability measure on Ω and $\|g\|_*=\sup_{x\in\Omega}|g(x)|$. Define

$$\begin{aligned}
\mathcal{G}_m^1&:=\left\{g:\Omega\to\mathbb{R}\ \middle|\ g=\varphi^2v_\theta^2\text{ where }v_\theta\in\widetilde{\mathcal{F}}_m\right\},\\
\mathcal{G}_m^2&:=\left\{g:\Omega\to\mathbb{R}\ \middle|\ g=\left|\nabla\left(\varphi v_\theta\right)\right|^2+V\varphi^2v_\theta^2\text{ where }v_\theta\in\widetilde{\mathcal{F}}_m\right\},\\
\mathcal{G}_m^3&:=\left\{g:\Omega\to\mathbb{R}\ \middle|\ g=\varphi\psi v_\theta\text{ where }v_\theta\in\widetilde{\mathcal{F}}_m\right\}.
\end{aligned}$$

Thanks to Assumption 7.1,

$$\begin{aligned}
&\max_{\theta\in\Theta}\|v_\theta\|_*\leqslant|c|+\sum_{i=1}^m|\gamma_i|\sup_{w_i\in\Theta_w,t_i\in\Theta_t,x\in\Omega}\left|\phi\left(w_i\cdot x+t_i\right)\right|\leqslant C+\Gamma\phi_{\max},\qquad(7.5)\\
&\max_{\theta\in\Theta}\||\nabla v_\theta|\|_*\leqslant\sum_{i=1}^m|\gamma_i|\,|w_i|\sup_{w_i\in\Theta_w,t_i\in\Theta_t,x\in\Omega}\left|\phi'\left(w_i\cdot x+t_i\right)\right|\leqslant\Gamma W\phi'_{\max}.
\end{aligned}$$

Since $|\nabla(\varphi v_\theta)| \leqslant |\varphi|\,|\nabla v_\theta| + |v_\theta|\,|\nabla\varphi|$ and $0 \leqslant V \leqslant V_{\max}$,

$$\sup_{g\in\mathcal{G}_m^2} \|g\|_* \leqslant \left[\Gamma W\phi'_{\max}/d + \pi\left(C+\Gamma\phi_{\max}\right)\right]^2 + V_{\max}\left(C+\Gamma\phi_{\max}\right)^2 d^{-2}. \tag{7.6}$$

The next proposition provides upper bounds for $\mathcal{N}\left(\delta, \mathcal{G}_m^i/M_i, \|\cdot\|_{L^2(Q)}\right)$, where $\{M_i\}_{i=1}^3$ are certain scaling parameters. Given C, Γ, W and T in (7.2), as in [47], we define

$$\mathcal{M}(\delta,\Lambda,m,d) := \frac{2C\Lambda}{\delta}\left(\frac{3\Gamma\Lambda}{\delta}\right)^m\left(\frac{3W\Lambda}{\delta}\right)^{dm}\left(\frac{3T\Lambda}{\delta}\right)^m.$$

Proposition 7.4. *If $0 \leqslant V \leqslant V_{\max}$ and ϕ satisfies Assumption 7.1, then for $i = 1,2,3$,*

$$\mathcal{N}\left(\delta, \mathcal{G}_m^i/M_i, \|\cdot\|_{L^2(Q)}\right) \leqslant \mathcal{M}\left(\delta, \Lambda_i/M_i, m, d\right),$$

where

$$\begin{aligned}
\Lambda_1 &:= 2\left(C+\Gamma\phi_{\max}\right)\left(1+\phi_{\max}+2L\Gamma\right)/d^2,\\
\Lambda_2 &:= 2\left[\Gamma W\phi'_{\max}/d + \pi\left(C+\Gamma\phi_{\max}\right)\right]\left[\Big((W+\Gamma)\phi'_{\max} + 2\Gamma W L'\Big)/d + \pi\left(1+\phi_{\max}+2L\Gamma\right)\right]\\
&\quad + 2V_{\max}\left(C+\Gamma\phi_{\max}\right)\left(1+\phi_{\max}+2L\Gamma\right),\\
\Lambda_3 &:= \|\psi\|_{L^2(Q)}\left(1+\phi_{\max}+2L\Gamma\right)/d.
\end{aligned}$$

The proof is postponed to Appendix C.1.

7.2 Estimates of the expectation of suprema of empirical processes

Let $\{\epsilon_i\}_{i=1}^\infty$ be independent Rademacher variables[5)] independent from $\{X_i\}_{i=1}^\infty$, and let $F \geqslant \sup_{f\in\mathcal{F}}|f|$ be a measurable envelope of the function class $\mathscr{F}$. Here, we call $\mathscr{F}$ a VC class if there exist some finite $A \geqslant 3\sqrt{e}$ and $v \geqslant 1$ such that for all probability measures Q and $0<\tau<1$, $\mathcal{N}\left(\tau\|F\|_{L^2(Q)}, \mathscr{F}, \|\cdot\|_{L^2(Q)}\right) \leqslant (A/\tau)^v$. We firstly recall the following fundamental lemma.

Lemma 7.5. *[25, Proposition 2.1] Let $\mathscr{F}$ be a measurable uniformly bounded VC class. Let $U \geqslant \sup_{f\in\mathscr{F}}\|f\|_{L^\infty}$ and $\sigma^2 \geqslant \sup_{f\in\mathscr{F}} E_{\mathcal{P}} f^2$ be such that $0<\sigma\leqslant U$. Then there exists a universal constant C such that for all $n\in\mathbb{N}$,*

$$\mathbf{E}\left\|\sum_{i=1}^n \epsilon_i f\left(X_i\right)\right\|_{\mathscr{F}} \leqslant C\left[vU\ln\frac{AU}{\sigma} + \sigma\sqrt{vn\ln\frac{AU}{\sigma}}\right].$$

To bound $\mathbf{E}\left\|\mathcal{P}_n - \mathcal{P}\right\|_{\mathscr{F}}$ with $\mathscr{F}$ a subset of a rescaled class $\mathcal{G}_m^i/M_i$ $(1\leqslant i\leqslant 3)$ in §7.1, we need the following lemma, whose proof is deferred to Appendix C.2.

Lemma 7.6. *Let $\mathscr{F}$ be a class of real valued measurable functions taking values in $[-1,1]$. Let $F\leqslant 1$ be a measurable envelope of $\mathscr{F}$ and $\sup_{f\in\mathscr{F}}\operatorname{Var}_{\mathcal{P}} f \leqslant \sigma^2 \leqslant 1$. Assume that for all $0<\tau<1$, there exists a universal net $\{f_i\}_{i=1}^M$ for all probability measures Q such that $M\leqslant (A/\tau)^v$ and for any $f\in\mathscr{F}$,*

$$\min_{1\leqslant i\leqslant M}\|f-f_i\|_{L^2(Q)} \leqslant \tau\|F\|_{L^2(Q)}.$$

Then, there exists a universal C such that for all $n\in\mathbb{N}$,

$$\mathbf{E}\left\|\mathcal{P}_n-\mathcal{P}\right\|_{\mathscr{F}} \leqslant C\left(\frac{v}{n}\ln\frac{A}{\sigma} + \sigma\sqrt{\frac{v}{n}\ln\frac{A}{\sigma}}\right).$$

Remark 7.7. If $\mathscr{F}\subset\mathcal{G}_m^i/M_i$ $(1\leqslant i\leqslant 3)$, then the universal nets exist due to Proposition C.1 and the procedure by which we control the covering numbers. The universal nets correspond to the nets for the parameter space Θ.

[5)] A Rademacher variable ϵ is one that satisfies $\mathbf{P}(\epsilon=1)=\mathbf{P}(\epsilon=-1)=1/2$.

Recall $K^{\mathscr{F}}_{n,q}(r,s]$ defined in (5.1) and $\mathcal{K}_n(\mathscr{F},r)=K^{\mathscr{F}}_{n,\sqrt{2}}(r,1]$. Corollary 7.8 is a direct consequence of Lemma 7.6. We postpone its proof to Appendix C.2.

Corollary 7.8. *Let $\mathscr{F}$ satisfy the assumptions in Lemma 7.6 and all functions in $\mathscr{F}$ take values in $[0,1]$. For all $n\in\mathbb{N}$, there holds*

$$]\mathcal{K}_n(\mathscr{F},r)\leqslant C\left(\frac{v}{nr^2}\ln\frac{A}{r}+\sqrt{\frac{v}{nr^2}\ln\frac{A}{r}}\right),$$

where $C\geqslant 1$ is an absolute constant. In particular, if $[v/(nr^2)]\ln(A/r)\leqslant 1$, then

$$\mathcal{K}_n(\mathscr{F},r)\leqslant 2C\sqrt{\frac{v}{nr^2}\ln\frac{A}{r}}.$$

7.3 Bounding $\mathcal{K}_n$ in the statistical error

We take $\mathcal{F}=\varphi\widetilde{\mathcal{F}}_m$ and $\widetilde{\mathcal{F}}_m=\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ with $\tau=9\sqrt{m}$, and consider

$$\begin{aligned}
&\mathcal{G}_1=\mathcal{G}_{\mathrm{SP}_\tau,m,1}(B):=\left\{g:g=\varphi^2v^2\text{ where }v\in\mathcal{F}_{\mathrm{SP}_\tau,m}(B)\right\},\\
&\mathcal{G}_2=\mathcal{G}_{\mathrm{SP}_\tau,m,2}(B):=\left\{g:g=|\nabla(\varphi v)|^2+V|\varphi v|^2\text{ where }v\in\mathcal{F}_{\mathrm{SP}_\tau,m}(B)\right\},\\
&\mathcal{F}_j=\mathcal{F}_{\mathrm{SP}_\tau,m,j}(B):=\{g:g=\varphi\psi_jv\text{ where }v\in\mathcal{F}_{\mathrm{SP}_\tau,m}(B)\}\quad\text{for }j=1,2,\ldots,k-1.
\end{aligned}$$

Note that $\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ coincides with the set $\widetilde{\mathcal{F}}_m$ defined in (7.2) with

$$C=B,\ \Gamma=4B,\ W=1,\ T=1.\tag{7.8}$$

By (7.5) and (7.6), using (7.4) and (7.8), we take $M_{\mathcal{F}}$, $M_{\mathcal{G}_2}$ as

$$\sup_{g\in\mathcal{F}}\|g\|_{L^\infty(\Omega)}\leqslant 9.5B/d=:M_{\mathcal{F}},\quad\sup_{g\in\mathcal{G}_2}\|g\|_{L^\infty(\Omega)}\leqslant 34^2B^2+V_{\max}(9.5B/d)^2=:M_{\mathcal{G}_2}.\tag{7.9}$$

Applying Lemma 7.6 or Corollary 7.8 to certain rescaled function classes, we estimate $\mathcal{K}_n$ as follows

Theorem 7.9. *Assume that $0\leqslant V\leqslant V_{\max}$ and $\|\psi_j\|_{L^\infty(\Omega)}\leqslant\mu_j$ for $1\leqslant j\leqslant k-1$. Consider the sets $\mathcal{F}=\varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$, $\mathcal{G}_1=\mathcal{G}_{\mathrm{SP}_\tau,m,1}(B)$, $\mathcal{G}_2=\mathcal{G}_{\mathrm{SP}_\tau,m,2}(B)$ and $\mathcal{F}_j=\mathcal{F}_{\mathrm{SP}_\tau,m,j}(B)$ with $\tau=9\sqrt{m}$ and $B\geqslant 1$. Assume that n is large enough such that*

$$C_0\frac{mB^2(1+V_{\max})}{nr^2}\ln\frac{B(1+\sqrt{m}/d)(1+V_{\max})}{rd}\leqslant 1,\tag{7.10}$$

where C_0 is an absolute constant. There exists an absolute constant C such that

$$\mathcal{K}_n\left(\mathcal{G}_1/M_{\mathcal{F}}^2,r/M_{\mathcal{F}}\right)\leqslant C\sqrt{\frac{mB^2}{ndr^2}\ln\frac{B}{rd}},\tag{7.11a}$$

$$\mathcal{K}_n\left(\mathcal{G}_2/M_{\mathcal{G}_2},r\sqrt{\lambda_1/M_{\mathcal{G}_2}}\right)\leqslant C\sqrt{\frac{mB^2(1+V_{\max})}{nr^2}\ln\frac{B(1+\sqrt{m}/d)(1+V_{\max})}{rd}},\tag{7.11b}$$

$$\mathcal{K}_n\left(\frac{\mathcal{F}_j}{2\mu_jM_{\mathcal{F}}}+\frac{1}{2},\sqrt{\frac{r}{4\mu_jM_{\mathcal{F}}}}\right)\leqslant C\left[\frac{m\mu_jB}{nr}\ln\left(\frac{\mu_jB}{rd}\right)+\sqrt{\frac{m\mu_jB}{nr}\ln\left(\frac{\mu_jB}{rd}\right)}\right].\tag{7.11c}$$

The proof is postponed to Appendix C.3.

8 Proof of the main generalization theorem

Combining Theorem 5.4, Theorem 6.7 and Theorem 7.9, we are ready to prove Theorem 3.4.

Proof of Theorem 3.4. Thanks to Theorem 5.4, taking $\mathcal{F}_{>r} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B) \cap \{\|u\|_{L^2(\Omega)} > r\}$ with $B = \|u^*\|_{\mathfrak{B}^s(\Omega)}$ and $\tau = 9\sqrt{m}$, if $2\xi_1 + \xi_2 < 1/2$, for any $u_{\mathcal{F}} \in \mathcal{F}_{>r}$,

$$L_k\left(u_n^m\right) - \lambda_k \leqslant 4\lambda_k\left(\xi_1 + \xi_2\right) + \beta\left(k\xi_3^2 + 4\sqrt{k}\xi_3\right) + 3\left(L_k\left(u_{\mathcal{F}}\right) - \lambda_k\right). \tag{8.1}$$

By Theorem 3.2, there exists $u_{\mathcal{F}} \in \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ such that $\|u^* - u_{\mathcal{F}}\|_{H^1(\Omega)} \leqslant 64B/\sqrt{m}$. Since $64B/\sqrt{m} \leqslant 1/2$ and $0 < r < 1/2$, $\|u_{\mathcal{F}}\|_{L^2(\Omega)} \geqslant \|u^*\|_{L^2(\Omega)} - 64B/\sqrt{m} > r$. Thus, $u_{\mathcal{F}} \in \mathcal{F}_{>r}$. By Theorem 6.7,

$$L_k\left(u_{\mathcal{F}}\right) - \lambda_k \leqslant 64\left(3\max\left\{1, V_{\max}\right\} + 7\lambda_k + 5\beta\right)B/\sqrt{m}. \tag{8.2}$$

Substituting the bounds in Theorem 7.9 into $\{\xi_i(n,r,\delta)\}_{i=1}^3$, we obtain, if (7.10) holds, then

$$\begin{aligned}
&\xi_1(n,r,\delta) \leqslant \frac{CB}{rd\sqrt{n}}\left(\sqrt{md\ln\frac{B}{rd}} + \sqrt{\ln(1/\delta)}\right),\\
&\xi_2(n,r,\delta) \leqslant C\Upsilon_1(n,m,B,r,\delta),\\
&\xi_3(n,r,\delta) \leqslant C\Upsilon_2(n,m,k,B,\bar{\mu}_k,r,\delta)/\sqrt{k},
\end{aligned} \tag{8.3}$$

where we have used (7.9), the fact $\lambda_1 \geqslant d\pi^2$ and (3.3b). Note that the bound for ξ_1 is smaller than the bound for ξ_2. Hence, there exists a constant C such that (3.3a) ensures both (7.10) and $2\xi_1 + \xi_2 \leqslant 1/2$. A combination of (8.1), (8.2) and (8.3) completes the proof. □

9 Solution theory in the spectral Barron Spaces

In this section we aim to prove the regularity of the eigenfunctions in the sine Barron space, as stated in Theorem 3.8. The regularity properties of PDEs within Barron spaces have been investigated only recently. In [20], the regularity results for the screened Poisson equation and various time-dependent equations in the Barron space have been proved through integral representations. The authors in [47] established a solution theory for the Poisson equation and the Schrödinger equations on the hypercube with homogeneous Neumann boundary condition. This work was further extended in [46] to include regularity estimates for the ground state of the Schrödinger operator. In addition, Chen et al. [14] proved the regularity theory for the static Schrödinger equations on $\mathbb{R}^d$ within the spectral Barron space.

Without loss of generality, we assume that $V \geqslant 0$. For $f \in L^2(\Omega)$, consider the static Schrödinger equation with Dirichlet boundary condition

$$\mathcal{H}u = -\Delta u + Vu = f \quad \text{on } \Omega, \qquad\qquad u = 0 \quad \text{on } \partial\Omega. \tag{9.1}$$

To show the boundedness of $\mathcal{H}^{-1}$, we prove an estimate for the solution of (9.1).

Theorem 9.1. *Assume that $f \in \mathfrak{B}^s(\Omega)$ and $V \in \mathfrak{C}^s(\Omega)$ with $s \geqslant 0$ and $V(x) \geqslant 0$ for $x \in \Omega$. Then (9.1) has a unique solution $u \in \mathfrak{B}^{s+2}(\Omega)$. Moreover, there exists $C > 0$ depending on V and d such that*

$$\|u\|_{\mathfrak{B}^{s+2}(\Omega)} \leqslant C(V,d)\|f\|_{\mathfrak{B}^s(\Omega)}.$$

Corollary 9.2. *Assume that $V \in \mathfrak{C}^s(\Omega)$ with $V(x) \geqslant 0$ for $x \in \Omega$. Let $\mathcal{S} := \mathcal{H}^{-1}$. Then $\mathcal{S} : \mathfrak{B}^s(\Omega) \to \mathfrak{B}^{s+2}(\Omega)$ is bounded and $\mathcal{S}$ is a compact operator on $\mathfrak{B}^s(\Omega)$.*

We prove the above two results in Appendix D.2. To prove Theorem 3.8, we start with $s = 0$.

Proposition 9.3. *If $V \in \mathfrak{C}^0(\Omega)$, then any eigenfunction of Problem (2.1) lies in $\mathfrak{B}^2(\Omega)$.*

Proof. By the definition of $\mathfrak{C}^s(\Omega)$ and Lemma D.2, $V \in \mathfrak{C}^0(\Omega)$ if and only if $\tilde{V}_e \in \ell^1\left(\mathbb{Z}^d\right)$. Specifically,

$$\begin{aligned}
\|\tilde{V}_e\|_{\ell^1(\mathbb{Z}^d)} &= \sum_{k\in\mathbb{N}_0^d} 2^{\sum_{i=1}^d \mathbf{1}_{k_i\neq 0}}|\tilde{V}_e(k)| = \sum_{k\in\mathbb{N}_0^d} 2^{\mathbf{1}_{k\neq 0}}|\check{V}(k)| \leqslant \|V\|_{\mathfrak{C}^0(\Omega)},\\
\|V\|_{\mathfrak{C}^0(\Omega)} &= \sum_{k\in\mathbb{N}_0^d} 2\beta_k^{-1}|\tilde{V}_e(k)| \leqslant 2\|\tilde{V}_e\|_{\ell^1(\mathbb{Z}^d)},
\end{aligned}$$

where we have used $\beta_k = 2^{\mathbf{1}_{k\neq 0}-\sum_{i=1}^d \mathbf{1}_{k_i\neq 0}} \in [2^{1-d},1]$. Since $\tilde{V}_e \in \ell^1\left(\mathbb{Z}^d\right)$, Young's inequality implies that for $p\in[1,2]$, if $\{a_k\}_{k\in\mathbb{Z}^d} \in \ell^p\left(\mathbb{Z}^d\right)$, then $(\tilde{V}_e * a) \in \ell^p\left(\mathbb{Z}^d\right)$ and

$$\|\tilde{V}_e * a\|_{\ell^p} \leqslant \|\tilde{V}_e\|_{\ell^1}\|a\|_{\ell^p}. \tag{9.2}$$

For any eigenmode $\lambda,\psi\in H_0^1(\Omega)$, let ψ_o be the odd extension of ψ on $[-1,1]^d$, which may be regarded as an eigenfunction of the Schrödinger operator $\widetilde{\mathcal{H}}\colon = -\Delta + V_e$ with periodic boundary conditions: $\widetilde{\mathcal{H}}\psi_o = -\Delta\psi_o + V_e\psi_o = \lambda\psi_o$.

Next, we claim that if $\tilde{\psi}_o \in \ell^p$ with some $p\in[1,2]$, then $(\pi^2|k|_2^2+1)\tilde{\psi}_o \in \ell^p$. Notice that

$$\begin{aligned}\tilde{\psi}_o &= \mathscr{F}\left[(-\Delta+I)^{-1}(\widetilde{\mathcal{H}}+I-V_e)\psi_o\right]\\ &= \left(\pi^2|k|_2^2+1\right)^{-1}(\lambda+1)\,\tilde{\psi}_o - \left(\pi^2|k|_2^2+1\right)^{-1}(\tilde{\psi}_o * \tilde{V}_e),\end{aligned}$$

where $\mathscr{F}$ represents the Fourier transform. Since $\tilde{V}_e \in \ell^1\left(\mathbb{Z}^d\right), \tilde{\psi}_o \in \ell^p\left(\mathbb{Z}^d\right)$, we conclude $(\lambda+1)\tilde{\psi}_o \in \ell^p\left(\mathbb{Z}^d\right)$ and $(\tilde{V}_e * \tilde{\psi}_o) \in \ell^p\left(\mathbb{Z}^d\right)$ from (9.2). Hence

$$(\lambda+1)\,\tilde{\psi}_o - (\tilde{\psi}_o * \tilde{V}_e) = \left(\pi^2|k|_2^2+1\right)\tilde{\psi}_o \in \ell^p\left(\mathbb{Z}^d\right).$$

This proves the claim.

Now we complete the proof through a bootstrap argument. Since $\psi\in L^2(\Omega)$, we have $\psi_o \in L^2(\widetilde{\Omega})$ and its Fourier transform $\tilde{\psi}_o \in \ell^2\left(\mathbb{Z}^d\right)$. It follows from the above claim that $\left(\pi^2|k|_2^2+1\right)\tilde{\psi}_o \in \ell^2\left(\mathbb{Z}^d\right)$. For all $r>d/2$, $\left(\pi^2|k|_2^2+1\right)^{-1} \in \ell^r\left(\mathbb{Z}^d\right)$. By Hölder's inequality, as long as $q^{-1} < 2/d + 1/2$ and $q\geqslant 1$, there exists $r>d/2$ such that $q^{-1} = r^{-1} + 2^{-1}$ and

$$\|\tilde{\psi}_o\|_{\ell^q(Z^d)} \leqslant \left\|\left(\pi^2|k|_2^2+1\right)^{-1}\right\|_{\ell^r(Z^d)} \left\|\left(\pi^2|k|_2^2+1\right)\tilde{\psi}_o\right\|_{\ell^2(Z^d)} < \infty.$$

Thus $\tilde{\psi}_o \in \ell^q\left(\mathbb{Z}^d\right)$. Repeating this argument j times, we have $\tilde{\psi}_o \in \ell^q\left(\mathbb{Z}^d\right)$ as long as $q^{-1} < 2j/d + 1/2$ and $q\geqslant 1$. Choosing j properly, we conclude that $\tilde{\psi}_o \in \ell^1\left(\mathbb{Z}^d\right)$. Repeating the claim again, we have $\left(\pi^2|k|_2^2+1\right)\tilde{\psi}_o \in \ell^1\left(\mathbb{Z}^d\right)$. Using Lemma D.1, we get

$$\|\psi\|_{\mathfrak{B}^2(\Omega)} = \sum_{k\in\mathbb{N}_+^d}\left(1+\pi^2|k|_1^2\right)2^d|\tilde{\psi}_o(k)| = \sum_{k\in\mathbb{Z}^d}\left(1+\pi^2|k|_1^2\right)|\tilde{\psi}_o(k)|.$$

Hence, $\psi\in\mathfrak{B}^2(\Omega)$ because $\|\psi\|_{\mathfrak{B}^2(\Omega)} \leqslant d\|\left(1+\pi^2|k|_2^2\right)\tilde{\psi}_o\|_{\ell^1(Z^d)}$. □

With the aid of Proposition 9.3 and Corollary 9.2, we are ready to prove Theorem 3.8.

Proof of Theorem 3.8. Note that $\mathfrak{B}^r(\Omega) \hookrightarrow \mathfrak{B}^s(\Omega)$ and $\mathfrak{C}^r(\Omega) \hookrightarrow \mathfrak{C}^s(\Omega)$ for $0\leqslant r\leqslant s$. Take any eigenmode (λ,ψ) of Problem (2.1) such that $\mathcal{H}\psi = \lambda\psi$. Since $V\in\mathfrak{C}^s(\Omega)$ with $s\geqslant 0$, $V\in\mathfrak{C}^0(\Omega)$ and thus $\psi\in\mathfrak{B}^2(\Omega)$ according to Proposition 9.3. For any $0\leqslant r\leqslant s$, $V\in\mathfrak{C}^r(\Omega)$, it follows from Corollary 9.2 that $\mathcal{S}:\mathfrak{B}^r(\Omega)\to\mathfrak{B}^{r+2}(\Omega)$ is bounded. Notice that $\psi = \lambda\mathcal{S}\psi$. Hence $\psi\in\mathfrak{B}^2(\Omega)$ implies $\psi\in\mathfrak{B}^{\min(s+2,4)}(\Omega)$. Repeating this argument j times, we conclude that $\psi\in\mathfrak{B}^{\min(s+2,2j+2)}(\Omega)$. When j is large enough so that $2j+2>s+2$, we obtain $\psi\in\mathfrak{B}^{s+2}(\Omega)$, which completes the proof. □

Acknowledgements The AI-driven experiments, simulations and model training were performed on the robotic AI-Scientist platform of Chinese Academy of Sciences. The work of Ming was supported by the National Natural Science Foundation of China under the Grant No. 12371438.

References

1 Adams R A, Fournier J J F. Sobolev Spaces, Pure Appl Math, Academic Press, vol. 140. New York, NY: Academic Press, 2003

2 Alexander K S. The central limit theorem for weighted empirical processes indexed by sets. J Multivariate Anal, 1987, 22: 313–339

3 Alexander K S. Rates of growth and sample moduli for weighted empirical processes indexed by sets. Probab Theory Relat Fields, 1987, 75: 379–423

4 Artin E. The Gamma Function. Athena Series: Selected Topics in Mathematics. Holt, Rinehart and Winston, 1964

5 Ascher U M, Greif C. A First Course in Numerical Methods, Comput Sci Eng, vol. 7. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2011

6 Barron A R. Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans Inf Theory, 1993, 39(3): 930–945

7 Bartlett P, Harvey N, Liaw C, et al. Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. J Mach Learn Res, 2019, 20(63): 1–17

8 Bartlett P, Lugosi G. An inequality for uniform deviations of sample averages from their means. Stat Probab Lett, 1999, 44(1): 55–62

9 Berg J, Nyström K. A unified deep artificial neural network approach to partial differential equations in complex geometries. Neurocomputing, 2018, 317: 28–41

10 Berner J, Grohs P, Jentzen A. Analysis of the generalization error: empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. SIAM J Math Data Sci, 2020, 2(3): 631–657

11 Cai Z, Liu J. Approximating quantum many-body wave functions using artificial neural networks. Phys Rev B, 2018, 97: 035116

12 Cao H, Shi Y, Zhang Z. Quantitative Green's function estimates for lattice quasi-periodic Schrödinger operators. Sci China Math, 2024, 67(5): 1011

13 Carleo G, Troyer M. Solving the quantum many-body problem with artificial neural networks. Science, 2017, 355(6325): 602–606

14 Chen Z, Lu J, Lu Y, et al. A regularity theory for static Schrödinger equations on $\mathbb{R}^d$ in spectral Barron spaces. SIAM J Math Anal, 2023, 55(1): 557–570

15 Choo K, Mezzacapo A, Carleo G. Fermionic neural-network states for ab-initio electronic structure. Nat Commun, 2020, 11(1): 2368

16 Davies E B. Heat Kernels and Spectral Theory, Camb Tracts Math, vol. 92. Cambridge : Cambridge University Press, 1989

17 Dudley R M. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. J Funct Anal, 1967, 1: 290–330

18 Dwivedi V, Srinivasan B. Physics informed extreme learning machine (pielm)–a rapid method for the numerical solution of partial differential equations. Neurocomputing, 2020, 391: 96–118,

19 E W, Han J, Jentzen A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. Commun Math Stat, 2017, 5(4): 349–380,

20 E W, Wojtowytsch S. Some observations on high-dimensional partial differential equations with Barron data. In: Bruna J, Hesthaven J, Zdeborova L (eds.) Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference, Proceedings of Machine Learning Research, vol. 145. 2022, 253–269

21 E W, Yu B. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. Commun Math Stat, 2018, 6(1): 1–12

22 Evans L C. Partial Differential Equations, Grad Stud Math, vol. 19. Providence, RI: American Mathematical Society (AMS), 2010

23 Gao X, Duan L M. Efficient representation of quantum many-body states with deep neural networks. Nat Commun, 2017, 8(1): 662

24 Van de Geer S A. Applications of Empirical Process Theory, Camb Ser Stat Probab Math, vol. 6. Cambridge: Cambridge University Press, 2000

25 Giné E, Guillou A. On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. Ann Inst Henri Poincaré, Probab Stat, 2001, 37(4): 503–522

26 Giné E, Koltchinskii V. Concentration inequalities and asymptotic results for ratio type empirical processes. Ann Probab, 2006, 34(3): 1143–1216

27 Giné E, Koltchinskii V, Wellner J A. Ratio limit theorems for empirical processes. In: Giné E, Houdré C, Nualart D (eds.) Stochastic Inequalities and Applications. Birkhäuser Basel, Basel, 2003, 249–278

28 Gonon L. Random feature neural networks learn Black-Scholes type PDEs without curse of dimensionality. J Mach Learn Res, 2023, 24(189): 1–51

29 Gonon L, Schwab C. Deep ReLU neural networks overcome the curse of dimensionality for partial integrodifferential equations. Anal Appl, 2023, 21(1): 1–47

30 Grohs P, Hornung F, Jentzen A, et al. A Proof that Artificial Neural Networks Overcome the Curse of Dimensionality in the Numerical Approximation of Black-Scholes Partial Differential Equations, Mem Am Math Soc, vol. 1410. Providence, RI: American Mathematical Society (AMS), 2023

31 Guo Y, Ming P. A deep learning method for computing eigenvalues of the fractional Schrödinger operator. J Syst Sci Complex, 2024, 37(2): 391–412

32 Han J, Jentzen A, E W. Solving high-dimensional partial differential equations using deep learning. Proc Natl Acad Sci USA, 2018, 115(34): 8505–8510

33 Han J, Lu J, Zhou M. Solving high-dimensional eigenvalue problems using deep neural networks: a diffusion Monte Carlo like approach. J Comput Phys, 2020, 423: 13

34 Han J, Zhang L, E W. Solving many-electron Schrödinger equation using deep neural networks. J Comput Phys, 2019, 399: 8

35 Hermann J, Schätzle Z, Noé F. Deep-neural-network solution of the electronic Schrödinger equation. Nat Chem, 2020, 12(10): 891–897

36 Holliday E G, Lindner J F, Ditto W L. Solving quantum billiard eigenvalue problems with physics-informed machine learning. AIP Adv, 2023, 13: 085013

37 Hutzenthaler M, Jentzen A, Kruse T. Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. Found Comput Math, 2022, 22(4): 905–966

38 Ji X, Jiao Y, Lu X, et al. Deep Ritz method for elliptical multiple eigenvalue problems. J Sci Comput, 2024, 98(2): 34

39 Khoo Y, Lu J, Ying L. Solving for high-dimensional committor functions using artificial neural networks. Res Math Sci, 2019, 6(1): 13

40 Kreĭn M G, Rutman M A. Linear operators leaving invariant a cone in a Banach space. Usp Mat Nauk, 1948, 3(1): 3–95

41 Krishnapriyan A, Gholami A, Zhe S, et al. Characterizing possible failure modes in physics-informed neural networks. In: Ranzato M, Beygelzimer A, Dauphin Y, et al (eds.) Advances in Neural Information Processing Systems, vol. 34. Curran Associates, Inc. 2021, 26548–26560

42 Lagaris I, Likas A, Fotiadis D. Artificial neural networks for solving ordinary and partial differential equations. IEEE Trans Neural Netw, 1998, 9(5): 987–1000

43 Lagaris I, Likas A, Papageorgiou D. Neural-network methods for boundary value problems with irregular boundaries. IEEE Trans Neural Netw, 2000, 11(5): 1041–1049

44 Li H, Ying L. A semigroup method for high dimensional elliptic PDEs and eigenvalue problems based on neural networks. J Comput Phys, 2022, 453: 15

45 Liao Y, Ming P. Spectral Barron space for deep neural network approximation. SIAM J Math Data Sci, 2025, In press

46 Lu J, Lu Y. A priori generalization error analysis of two-layer neural networks for solving high dimensional Schrödinger eigenvalue problems. Commun Am Math Soc, 2022, 2: 1–21

47 Lu Y, Lu J, Wang M. A priori generalization analysis of the deep Ritz method for solving high dimensional elliptic partial differential equations. In: Belkin M, Kpotufe S (eds.) Proceedings of Thirty Fourth Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 134. 2021, 3196–3241

48 Lubich C. From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis. Zur Lect Adv Math. Zürich: European Mathematical Society (EMS), 2008

49 Luo T, Yang H. Two-layer neural networks for partial differential equations: optimization and generalization theory. In: Mishra S, Townsend A (eds.) Numerical Analysis Meets Machine Learning, Handbook of Numerical Analysis, vol. 25. Elsevier, 2024, 515–554

50 Lyu L, Wu K, Rui D, et al. Enforcing exact boundary and initial conditions in the deep mixed residual method. CSIAM Trans Appl Math, 2021, 2(4): 748–775

51 Mathias D, Ehrlacher V. Two-layers neural networks for Schrödinger eigenvalue problems. arXiv:2409.01640, 2024

52 Massart P. Some applications of concentration inequalities to statistics. Ann Fac Sci Toulouse Math, 2000, 9(2): 245–303

53 McFall K S, Mahan J R. Artificial neural network method for solution of boundary value problems with exact satisfaction of arbitrary boundary conditions. IEEE Trans Neural Netw, 2009, 20(8): 1221–1233

54 Nielsen M A, Chuang I L. Quantum Computation and Quantum Information. Cambridge: Cambridge University Press, 2010

55 Pisier G. Remarques sur un résultat non publié de b. maurey. Séminaire Analyse fonctionnelle (dit "Maurey-Schwartz"), 1980-1981, 1–12

56 Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys, 2019, 378: 686–707

57 Sheng H, Yang C. PFNN: a penalty-free neural network method for solving a class of second-order boundary-value problems on complex geometries. J Comput Phys, 2021, 428: 13

58 Siegel J W, Hong Q, Jin X, et al. Greedy training algorithms for neural networks and applications to PDEs. J

Comput Phys, 2023, 484: 27
59 Siegel J W, Xu J. High-order approximation rates for shallow neural networks with cosine and ReLU^k activation functions. Appl Comput Harmon Anal, 2022, 58: 1–26
60 Sirignano J, Spiliopoulos K. DGM: a deep learning algorithm for solving partial differential equations. J Comput Phys, 2018, 375: 1339–1364
61 Sui Y, Zhang D, Cao J, et al. An efficient finite element method and error analysis for eigenvalue problem of Schrödinger equation with an inverse square potential on spherical domain. Adv Difference Equ, 2020, 2020:14
62 Sukumar N, Srivastava A. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. Comput Methods Appl Mech Eng, 2022, 389: 50
63 Vershynin R. High-Dimensional Probability. An Introduction with Applications in Data Science, Camb Ser Stat Probab Math, vol. 47. Cambridge: Cambridge University Press, 2018
64 Wainwright M J. High-dimensional Statistics. A Non-Asymptotic Viewpoint, Camb Ser Stat Probab Math, vol. 48. Cambridge: Cambridge University Press, 2019
65 Wang S, Teng Y, Perdikaris P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. SIAM J Sci Comput, 2021, 43(5): a3055–a3081
66 Wang Y, Xie H. Computing multi-eigenpairs of high-dimensional eigenvalue problems using tensor neural networks. J Comput Phys, 2024, 506: 112928
67 Wellner J A. Limit theorems for the ratio of the empirical distribution function to the true distribution function. Z Wahrscheinlichkeitstheor Verw Geb, 1978, 45: 73–88
68 Weyl H. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). Math Ann, 1912, 71: 441–479
69 Weyl H. Über die Abhängigkeit der Eigenschwingungen einer Membran von deren Begrenzung. J Reine Angew Math, 1912, 141: 1–11
70 Xu J. Finite neuron method and convergence analysis. Commun Comput Phys, 2020, 28(5): 1707–1745
71 Yarotsky D. Error bounds for approximations with deep ReLU networks. Neural Netw, 2017, 94: 103–114
72 Yarotsky D. Optimal approximation of continuous functions by very deep ReLU networks. In: Bubeck S, Perchet V, Rigollet P (eds.) Proceedings of the 31st Conference On Learning Theory, Proceedings of Machine Learning Research, vol. 75. 2018, 639–649
73 Zang Y, Bao G, Ye X, et al. Weak adversarial networks for high-dimensional partial differential equations. J Comput Phys, 2020, 411: 13

Appendix A Stability estimate of the k-th eigenfunction

In this appendix, we prove Proposition 2.2 and Proposition 2.3. Define P as the orthogonal projection operator from $L^2(\Omega)$ to U_k and $P^\perp$ as the orthogonal projection from $L^2(\Omega)$ to $U_k^\perp$. Recall that $\{\psi_j\}_{j=1}^{k-1}$ are the first $k-1$ normalized orthogonal eigenfunctions. For $u \in H^1(\Omega)$, we write

$$u = Pu + P^\perp u = Pu + w + z \quad \text{with} \quad w\colon= \sum_{j=1}^{k-1} \langle u, \psi_j\rangle\, \psi_j, \quad z\colon= P^\perp u - w.$$

Note that w is the orthogonal projection of u onto the subspace $W_k = \operatorname{span}\{\psi_1, \psi_2, \dots, \psi_{k-1}\}$ and z is the orthogonal projection of u onto the subspace $Z_k = W_k^\perp \cap \ker(\mathcal{H} - \lambda_k I)^\perp$. Recall that $\lambda_{k'}$ is the first eigenvalue of $\mathcal{H}$ that is strictly larger than λ_k. For any $z \in Z_k$,

$$\langle z, \mathcal{H} z\rangle \geqslant \lambda_{k'} \|z\|^2_{L^2(\Omega)} \geqslant \lambda_k \|z\|^2_{L^2(\Omega)}. \tag{A.1}$$

Notice that $\mathcal{H}$ leaves the three orthogonal subspaces W_k, U_k, Z_k invariant. Therefore,

$$\begin{aligned}\langle u, \mathcal{H} u\rangle &= \langle w, \mathcal{H} w\rangle + \langle Pu, \mathcal{H} Pu\rangle + \langle z, \mathcal{H} z\rangle \\ &= \sum_{j=1}^{k-1} \lambda_j \langle u, \psi_j\rangle^2 + \lambda_k \langle Pu, Pu\rangle + \langle z, \mathcal{H} z\rangle.\end{aligned} \tag{A.2}$$

Using the definition (2.3), the decomposition (A.2) and

$$\|u\|^2_{L^2(\Omega)} = \langle w, w\rangle + \langle Pu, Pu\rangle + \langle z, z\rangle, \tag{A.3}$$

we obtain

$$\begin{aligned}(L_k(u)-\lambda_k)\,\|u\|_{L^2(\Omega)}^2 &= \langle u,\mathcal{H}u\rangle+\beta\sum_{j=1}^{k-1}\langle u,\psi_j\rangle^2-\lambda_k\langle u,u\rangle\\ &=\sum_{j=1}^{k-1}(\beta+\lambda_j-\lambda_k)\,\langle u,\psi_j\rangle^2+\langle z,(\mathcal{H}-\lambda_k)z\rangle.\end{aligned}\tag{A.4}$$

This identity is the key to proving Proposition 2.2 and Proposition 2.3.

Proof of Proposition 2.2. Note that $0\leqslant\lambda_k-\lambda_j\leqslant\lambda_k-\lambda_1$ for each $1\leqslant j\leqslant k-1$. It follows from (A.2), (A.3) and (A.1) that

$$\begin{aligned}\left|\langle u,\mathcal{H}u\rangle-\lambda_k\|u\|_{L^2(\Omega)}^2\right| &= |\langle w,(\mathcal{H}-\lambda_k)w\rangle+\langle Pu,(\mathcal{H}-\lambda_k)Pu\rangle+\langle z,(\mathcal{H}-\lambda_k)z\rangle|\\ &\leqslant\sum_{j=1}^{k-1}(\lambda_k-\lambda_j)\,\langle u,\psi_j\rangle^2+\langle z,(\mathcal{H}-\lambda_k)z\rangle\\ &\leqslant\max\left\{\frac{\lambda_k-\lambda_1}{\beta+\lambda_1-\lambda_k},1\right\}(L_k(u)-\lambda_k)\,\|u\|_{L^2(\Omega)}^2,\end{aligned}$$

where we have used (A.4) and $\beta>\lambda_k-\lambda_1$ in the last line. This gives Proposition 2.2. □

Proof of Proposition 2.3. It follows from (A.1) that $\langle z,(\mathcal{H}-\lambda_k)\,z\rangle\geqslant(\lambda_{k'}-\lambda_k)\,\|z\|_{L^2(\Omega)}^2$. Using (A.4) and the facts $0\leqslant\lambda_k-\lambda_j\leqslant\lambda_k-\lambda_1$ for all $1\leqslant j\leqslant k-1$, we obtain

$$\begin{aligned}(L_k(u)-\lambda_k)\,\|u\|_{L^2(\Omega)}^2 &\geqslant(\beta+\lambda_1-\lambda_k)\,\|w\|_{L^2(\Omega)}^2+(\lambda_{k'}-\lambda_k)\,\|z\|_{L^2(\Omega)}^2\\ &\geqslant\min\{\beta+\lambda_1-\lambda_k,\lambda_{k'}-\lambda_k\}\left\|P^\perp u\right\|_{L^2(\Omega)}^2,\end{aligned}\tag{A.5}$$

where we have used $\left\|P^\perp u\right\|_{L^2(\Omega)}^2=\|w\|_{L^2(\Omega)}^2+\|z\|_{L^2(\Omega)}^2$. Since $\beta+\lambda_1-\lambda_k$ and $\lambda_{k'}-\lambda_k$ are both strictly greater than zero, the inequality (A.5) implies the estimate (2.7a).

To bound $\left\|\nabla P^\perp u\right\|^2$, we note that

$$\begin{aligned}(L_k(u)-\lambda_k)\,\|u\|_{L^2(\Omega)}^2 &= \langle Pu,(\mathcal{H}-\lambda_k)Pu\rangle+\left\langle P^\perp u,(\mathcal{H}-\lambda_k)P^\perp u\right\rangle+\beta\|w\|_{L^2(\Omega)}^2\\ &\geqslant\left\langle P^\perp u,(\mathcal{H}-\lambda_k)P^\perp u\right\rangle\\ &=\int_\Omega\left|\nabla P^\perp u\right|^2\,\mathrm{d}\,x+\int_\Omega(V-\lambda_k)\left|P^\perp u\right|^2\,\mathrm{d}\,x,\end{aligned}$$

where we have used $\langle Pu,(\mathcal{H}-\lambda_k)Pu\rangle=0$. Rearranging the terms, we arrive at

$$\begin{aligned}\left\|\nabla P^\perp u\right\|_{L^2(\Omega)}^2 &\leqslant(L_k(u)-\lambda_k)\,\|u\|_{L^2(\Omega)}^2-\int_\Omega(V-\lambda_k)\left|P^\perp u\right|^2\,\mathrm{d}\,x\\ &\leqslant(L_k(u)-\lambda_k)\,\|u\|_{L^2(\Omega)}^2+(\lambda_k-V_{\min})\left\|P^\perp u\right\|_{L^2(\Omega)}^2.\end{aligned}$$

Substituting (2.7a) into the above inequality, we obtain (2.7b). □

Appendix B Missing proof in section 6

Appendix B.1 Preliminaries

Proof of Lemma 6.1. (1) For $u\in\mathfrak{B}^0(\Omega)$, it follows from the fact $\|\Phi_k\|_{L^\infty(\Omega)}\leqslant 1$ that

$$\|u\|_{L^\infty(\Omega)}\leqslant\sum_{k\in\mathbb{N}_+^d}|\hat{u}(k)|=\frac{1}{2}\|u\|_{\mathfrak{B}^0(\Omega)}.$$

Moreover, since $u\in\mathfrak{B}^s(\Omega)$ with $s\geqslant 0$ have summable sine coefficients, the sum of sine expansion converges uniformly, which implies that $u\in C(\overline{\Omega})$ and u vanishes on the boundary of Ω.

(2) If $u \in \mathfrak{B}^2(\Omega)$, then $|\hat{u}(k)| \leqslant \|u\|_{\mathfrak{B}^2(\Omega)}$ for each $k \in \mathbb{N}_+^d$ by definition. It follows from the Cauchy-Schwarz inequality that

$$\|u\|_{H^1(\Omega)}^2 \leqslant 2^{-d}\|u\|_{\mathfrak{B}^2(\Omega)} \sum_{k\in\mathbb{N}_+^d} \left(1+\pi^2|k|_1^2\right)|\hat{u}(k)| \leqslant 2^{-d}\|u\|_{\mathfrak{B}^2(\Omega)}^2.$$

Hence $u \in H_0^1(\Omega)$. □

Appendix B.2 Upper bounds for the cut-off function φ

The following lemma gives upper bounds for the cut-off function φ and its gradient.

Lemma B.1. *For all $x \in \Omega$, it holds that $0 < \varphi(x) < 1/d$ and $|\nabla\varphi(x)| < \pi$.*

Proof. For any $x = (x_1, x_2, \cdots, x_d)^T \in \Omega$ and $1 \leqslant i \leqslant d$, $0 < \sin(\pi x_i) < 1$. It is evident that $0 < \varphi(x) < 1/d$. A straightforward calculation yields

$$|\nabla\varphi(x)|^2 = \pi^2 \frac{\sum_{l=1}^d \cos^2(\pi x_l)\left[\prod_{j\neq l}\sin(\pi x_j)\right]^4}{\left[\sum_{l=1}^d \prod_{j\neq l}\sin(\pi x_j)\right]^4}.$$

Since for every $1 \leqslant l \leqslant d$ and $x \in (0,1)^d$, $\prod_{j\neq l}\sin(\pi x_j) > 0$ and $\cos^2(\pi x_l) \in [0,1)$, we have

$$\sum_{l=1}^d \cos^2(\pi x_l)\left[\prod_{j\neq l}\sin(\pi x_j)\right]^4 < \sum_{l=1}^d \left[\prod_{j\neq l}\sin(\pi x_j)\right]^4 \leqslant \left[\sum_{l=1}^d\prod_{j\neq l}\sin(\pi x_j)\right]^4,$$

which implies $|\nabla\varphi(x)|^2 < \pi^2$ for all $x \in (0,1)^d$. □

Lemma B.2. *For any $h \in H^1(\Omega)$, $\|\varphi h\|_{H^1(\Omega)} \leqslant \sqrt{21}\,\|h\|_{H^1(\Omega)}$. Particularly, if $\{h_j\}_{j=1}^\infty$ converges to h in $H^1(\Omega)$, then $\{\varphi h_j\}_{j=1}^\infty$ converges to φh in $H^1(\Omega)$.*

Proof. By Lemma B.1, for $h \in H^1(\Omega)$, a direct calculation yields $\|\varphi h\|_{L^2(\Omega)} \leqslant \|h\|_{L^2(\Omega)}/d$ and

$$\int_\Omega |\nabla(\varphi(x)h(x))|^2 \,\mathrm{d}\,x \leqslant 2\int_\Omega \left(h^2|\nabla\varphi|^2 + \varphi^2|\nabla h|^2\right)\mathrm{d}\,x \leqslant 2\int_\Omega \left(\pi^2 h^2 + \frac{1}{d^2}|\nabla h|^2\right)\mathrm{d}\,x.$$

Notice that $2\pi^2 + 1 < 21$ and $d \geqslant 1$. Therefore, we obtain

$$\|\varphi h\|_{H^1(\Omega)}^2 \leqslant \left(2\pi^2 + \frac{1}{d^2}\right)\|h\|_{L^2(\Omega)}^2 + \frac{2}{d^2}\|\nabla h\|_{L^2(\Omega)}^2 \leqslant 21\,\|h\|_{H^1(\Omega)}^2.$$

In particular, if $\{h_j\}_{j=1}^\infty$ converges to h in $H^1(\Omega)$, then $\|\varphi h_j - \varphi h\|_{H^1(\Omega)} \leqslant \sqrt{21}\,\|h_j - h\|_{H^1(\Omega)} \to 0$. □

Appendix B.3 Sine Spectral Barron Space and Neural Network Approximation

To prove Proposition 6.2, we need the following elementary facts.

Lemma B.3. *The following expansion holds for $m \in \mathbb{N}_+$ and $x \in (0,1)$*

$$\frac{\sin(m\pi x)}{\sin(\pi x)} = \begin{cases} 1 + \displaystyle\sum_{l=1}^{(m-1)/2} 2\cos(2l\pi x), & m \text{ is odd}, \\ \displaystyle\sum_{l=1}^{m/2} 2\cos((2l-1)\pi x), & m \text{ is even}. \end{cases}$$

The proof is straightforward, and we omit the proof.

Proof of Proposition 6.2. For $u \in \mathfrak{B}^{s+1}$, dividing both sides of $u = \sum_{k\in\mathbb{N}_+^d} \hat{u}(k)\Phi_k$ by φ, we get

$$\frac{u(x)}{\varphi(x)} = \sum_{k\in\mathbb{N}_+^d} \hat{u}(k) \left(\sum_{i=1}^d \frac{\sin(k_i\pi x_i)}{\sin(\pi x_i)} \prod_{j\neq i} \sin(k_j \pi x_j) \right). \tag{B.1}$$

The sum in (B.1) is absolutely convergent for all $x\in\Omega$ because $\left|\frac{\sin(k_i\pi x)}{\sin(\pi x)}\right| \leqslant k_i$ and

$$\sum_{k\in\mathbb{N}_+^d} |\hat{u}(k)| \left(\sum_{i=1}^d \frac{|\sin(k_i\pi x_i)|}{\sin(\pi x_i)} \prod_{j\neq i} |\sin(k_j\pi x_j)| \right) \leqslant \sum_{k\in\mathbb{N}_+^d} |\hat{u}(k)| \left(\sum_{i=1}^d k_i\right) \leqslant \|u\|_{\mathfrak{B}^1(\Omega)}.$$

Expanding $\frac{\sin(k_i\pi x_i)}{\sin(\pi x_i)}$ with $k_i\in\mathbb{N}_+$,using by Lemma B.3 and substituting this expansion into (B.1), we obtain that u/φ has an expansion of the form (6.1), where $(k,i)\in\Gamma$ if and only if $k\in\mathbb{N}_0^d$ has at most one zero component at position k_i. Furthermore, the coefficients $\hat{v}(k,i)$ can be expressed in terms of $\hat{u}(k)$ as

$$\hat{v}(k,i) = \left(1 + \mathbf{1}_{\{k_i\geqslant 1\}}\right) \sum_{l=0}^{\infty} \hat{u}\left(k + (2l+1)e_i\right) \qquad \text{for } (k,i)\in\Gamma, \tag{B.2}$$

where e_i is the i-th cannonical basis.

By (B.2), we obtain

$$\sum_{(k,i)\in\Gamma} (1+\pi^s|k|_1^s)\,|\hat{v}(k,i)| \leqslant \sum_{i=1}^d \left(\sum_{k\in\mathbb{N}_+^d, k_i \text{ is odd}} A_1|\hat{u}(k)| + \sum_{k\in\mathbb{N}_+^d, k_i \text{ is even}} A_2|\hat{u}(k)| \right), \tag{B.3}$$

where

$$A_1 = 1 + \pi^s\,|k - k_ie_i|_1^s + \mathbf{1}_{\{k_i\geqslant 3\}}\cdot 2 \sum_{l=0}^{(k_i-3)/2} \left[1 + \pi^s\,|k-(2l+1)e_i|_1^s\right],$$

$$A_2 = 2\sum_{l=1}^{k_i/2} \left[1+\pi^s\,|k-(2l-1)e_i|_1^s\right].$$

Notice that $s\geqslant 0$, t^s is a nondecreasing function for $t\geqslant 0$. When $k\in\mathbb{N}_+^d$ and k_i is odd,

$$A_1 \leqslant k_i + 2\pi^s \sum_{l=0}^{(k_i-1)/2} (|k|_1 - 2l - 1)^s \leqslant k_i + 2\pi^s \int_{|k|_1-k_i}^{|k|_1} t^s\,\mathrm{d}\,t.$$

Similarly, when $k\in\mathbb{N}_+^d$ and k_i is even,

$$A_2 = k_i + 2\pi^s \sum_{l=1}^{k_i/2} (|k|_1 - 2l + 1)^s \leqslant k_i + 2\pi^s \int_{|k|_1-k_i}^{|k|_1} t^s\,\mathrm{d}\,t.$$

Hence,

$$|A_1|\,, |A_2| \leqslant \frac{2\pi^s}{s+1}\left(|k|_1^{s+1} - (|k|_1-k_i)^{s+1}\right).$$

Substituting the above bound into (B.3) and exchanging the order of summation, we bound (B.3) by

$$\begin{aligned}\sum_{(k,i)\in\Gamma} (1+\pi^s|k|_1^s)\,|\hat{v}(k,i)| &\leqslant \sum_{i=1}^d \left\{ \sum_{k\in\mathbb{N}_+^d} |\hat{u}(k)| \left[k_i + \frac{2\pi^s}{s+1}\left(|k|_1^{s+1} - (|k|_1-k_i)^{s+1}\right)\right]\right\} \\ &= \sum_{k\in\mathbb{N}_+^d} |\hat{u}(k)| \left\{ |k|_1 + \frac{2\pi^s}{s+1}\left[d|k|_1^{s+1} - \sum_{i=1}^d (|k|_1-k_i)^{s+1}\right]\right\}\end{aligned} \tag{B.4}$$

Using Jensen's inequality and the fact that t^{s+1} is convex, we get

$$\frac{1}{d}\sum_{i=1}^{d}(|k|_1-k_i)^{s+1}\geqslant\left[\frac{1}{d}\sum_{i=1}^{d}(|k|_1-k_i)\right]^{s+1}=\left(\frac{d-1}{d}|k|_1\right)^{s+1}.$$

Combining the above two inequalities and using the Bernoulli inequality

$$(1-1/d)^{s+1}>1-\frac{s+1}{d},$$

we obtain

$$\begin{aligned}\sum_{(k,i)\in\Gamma}(1+\pi^s|k|_1^s)\,|\hat{v}(k,i)|&\leqslant\leqslant\sum_{k\in\mathbb{N}_+^d}|\hat{u}(k)|\left\{|k|_1+\frac{2\pi^s}{s+1}\left[d|k|_1^{s+1}-d\left(\frac{d-1}{d}|k|_1\right)^{s+1}\right]\right\}\\&=\sum_{k\in\mathbb{N}_+^d}|\hat{u}(k)|\left\{|k|_1+\frac{2d}{\pi(s+1)}\left[1-\left(\frac{d-1}{d}\right)^{s+1}\right]\pi^{s+1}|k|_1^{s+1}\right\}\\&\leqslant\sum_{k\in\mathbb{N}_+^d}|\hat{u}(k)|\left\{|k|_1+\frac{2}{\pi}\pi^{s+1}|k|_1^{s+1}\right\}\\&\leqslant(3/\pi)\|u\|_{\mathfrak{B}^{s+1}(\Omega)}\leqslant\|u\|_{\mathfrak{B}^{s+1}(\Omega)}.\end{aligned}$$

The estimate (6.2) follows from (B.3), (B.4) and the above bound. □

The following lemma is useful to show that u/φ lies in the convex hull of $\mathcal{F}_s(B)$.

Lemma B.4. *[47, Lemma 4.2] For any $\theta=(\theta_1,\theta_2,\cdots,\theta_d)^T\in\mathbb{R}^d$,*

$$\prod_{i=1}^{d}\cos\theta_i=\frac{1}{2^d}\sum_{\xi\in\{1,-1\}^d}\cos(\xi\cdot\theta).$$

Proof of Proposition 6.3. **Step 1:** Show that u lies in $\overline{\operatorname{conv}(\varphi\mathcal{F}_s(B_u))}$ with $B_u=\|u\|_{\mathfrak{B}^{s+1}(\Omega)}$. By Lemma B.4 and $\sin\theta_j=\cos(\theta_j-\pi/2)$, let $\theta=(k_1\pi x_1,k_2\pi x_2,\dots,k_d\pi x_d)^\top$, hence

$$\cos(k_i\pi x_i)\prod_{j\neq i}\sin(k_j\pi x_j)=\frac{1}{2^d}\sum_{\xi\in\{1,-1\}^d}\cos\left(\pi k_\xi\cdot x-\frac{\pi}{2}\sum_{j\neq i}\xi_j\right),\tag{B.5}$$

where $k_\xi=(k_1\xi_1,k_2\xi_2,\dots,k_d\xi_d)$. Since $u\in\mathfrak{B}^{s+1}(\Omega)$ with $s\geqslant1$, plugging (B.5) into the expansion of u/φ in Proposition 6.2 yields

$$\begin{aligned}\frac{u(x)}{\varphi(x)}&=\sum_{(k,i)\in\Gamma}\hat{v}(k,i)\cdot\frac{1}{2^d}\sum_{\xi\in\{1,-1\}^d}\cos\left(\pi k_\xi\cdot x-\frac{\pi}{2}\sum_{j\neq i}\xi_j\right)\\&=\sum_{(k,i)\in\Gamma}\frac{|\hat{v}(k,i)|\,(1+\pi^s|k|_1^s)}{Z_v}\frac{Z_v}{1+\pi^s|k|_1^s}\frac{1}{2^d}\sum_{\xi\in\{1,-1\}^d}\operatorname{sign}(\hat{v}(k,i))\cos\left(\pi k_\xi\cdot x-\frac{\pi}{2}\sum_{j\neq i}\xi_j\right),\end{aligned}$$

where Z_v is a constant to be specified later on. We define a probability measure on Γ by

$$\mu(d(k,i)){:}=\frac{|\hat{v}(k,i)|\,(1+\pi^s|k|_1^s)}{Z_v}\delta(d(k,i))$$

with $Z_v=\sum_{(k,i)\in\Gamma}|\hat{v}(k,i)|\,(1+\pi^s|k|_1^s)\leqslant B_u$. Let $\operatorname{sign}(\hat{v}(k,i))=(-1)^{\theta_{k,i}}$ with $\theta_{k,i}\in\{0,1\}$, and

$$g(x,k,i)=\frac{Z_v}{1+\pi^s|k|^s}\cdot\frac{1}{2^d}\sum_{\xi\in\{1,-1\}^d}\cos\left(\pi\left(k_\xi\cdot x-\frac{1}{2}\sum_{j\neq i}\xi_j+\theta_{k,i}\right)\right)\varphi(x).$$

Then, for any $x \in \Omega$,

$$u(x) = \sum_{(k,i)\in\Gamma} g(x,k,i)\mu(d(k,i)) = \mathbf{E}_\mu g(x,k,i). \tag{B.6}$$

For $\xi \in \{1,-1\}^d$, $\sum_{j\neq i}\xi_j$ is even when d is odd, and $\sum_{j\neq i}\xi_j$ is odd when d is even. It is clear that $g(x,k,i)$ is a convex combination of 2^d elements in $\varphi\mathcal{F}_s(B_u)$. Moreover, thanks to (B.6) and the uniform boundness of $\|g(\cdot,k,i)\|_{H^1(\Omega)}$ derived from the following step, u lies in $\overline{\operatorname{conv}(\varphi\mathcal{F}_s(B_u))}$.

Step 2: Check that $\varphi\mathcal{F}_s(B)$ with $s \geqslant 1$ is a bounded set in $H^1(\Omega)$. By Lemma B.1, whether f is a sine function or a cosine function,

$$\int_\Omega \varphi^2(x) f^2(\pi(k\cdot x+b))\,\mathrm{d}\,x \leqslant \int_\Omega \varphi^2(x)\,\mathrm{d}\,x \leqslant d^{-2},$$

and

$$\begin{aligned}\int_\Omega |\nabla(\varphi(x)f(\pi(k\cdot x+b)))|^2\,\mathrm{d}\,x &\leqslant 2\int_\Omega \left[f^2(\pi(k\cdot x+b))|\nabla\varphi|^2 + \varphi^2|\nabla f(\pi(k\cdot x+b))|^2\right]\,\mathrm{d}\,x\\ &\leqslant 2\int_\Omega \left(|\nabla\varphi|^2 + \pi^2|k|^2\varphi^2\right)\,\mathrm{d}\,x\\ &\leqslant 2\pi^2(1+d^{-2}).\end{aligned}$$

Hence, for any $w \in \varphi\mathcal{F}_s(B)$, $w(x) = \gamma(1+\pi^s|k|^s)^{-1}\varphi(x)f(\pi(k\cdot x+b))$,

$$\|w\|_{H^1(\Omega)}^2 \leqslant \frac{B^2}{(1+\pi^s|k|_1^s)^2}\left(\frac{1}{d^2} + 2\pi^2 + \frac{2\pi^2|k|^2}{d^2}\right).$$

When $|k|_1 \geqslant 1$, since $s \geqslant 1$ and

$$1/d^2 + 2\pi^2 + 2\pi^2|k|^2/d^2 \leqslant 4 + 8\pi|k|_1 + 4\pi^2|k|_1^2 \leqslant 4\left(1+\pi^s|k|_1^s\right)^2,$$

we have $\|w\|_{H^1(\Omega)}^2 \leqslant 4B^2$. Since $k \in \Gamma_1$, the only case where $k=0$ can occur is when $d=1$. In this case, $w(x) = \gamma\varphi(x) = \gamma\sin(\pi x)$ and

$$\|w\|_{H^1(\Omega)}^2 \leqslant \gamma^2\left(1+\pi^2\right)/2 \leqslant 6B^2.$$

Thus, we conclude that for any $w \in \varphi\mathcal{F}_s(B)$ with $s \geqslant 1$,

$$\|w\|_{H^1(\Omega)} \leqslant \sqrt{6}B.$$

Step 3: Using Lemma 6.4 along with the facts that u lies in $\overline{\operatorname{conv}(\varphi\mathcal{F}_s(B_u))}$ and $\|w\|_{H^1(\Omega)} \leqslant \sqrt{6}B$ for any $w \in \varphi\mathcal{F}_s(B)$, we obtain Proposition 6.3. $\square$

Appendix B.4 Reduction to ReLU and Softplus Activation Functions

Proof of Lemma 6.5. Let g_m be the piecewise linear interpolant of g with respect to grid $\{z_j\}_{j=0}^{2m}$. Let $h = \max(h_1,h_2)$. Then, $h \leqslant (1+\rho)/m \leqslant 3/(2m)$, According to [5],

$$\|g-g_m\|_{L^\infty([-1,1])} \leqslant \frac{h^2}{8}\|g''\|_{L^\infty([-1,1])}.$$

Consider $z \in [z_j, z_{j+1}]$ for some $0 \leqslant j \leqslant 2m-1$. By the mean value theorem, there exist $\xi,\eta \in (z_j,z_{j+1})$ such that $\left(g(z_{j+1})-g(z_j)\right)/(z_{j+1}-z_j) = g'(\xi)$ and

$$\left|g'(z) - \frac{g\left(z_{j+1}\right)-g\left(z_i\right)}{z_{j+1}-z_j}\right| = |g'(z)-g'(\xi)| = |g''(\eta)|\,|z-\xi|,$$

which implies that $\|g'-g_m'\|_{L^\infty([-1,1])} \leqslant h\|g''\|_{L^\infty([-1,1])}$. Thus,

$$\|g-g_m\|_{W^{1,\infty}([-1,1])} \leqslant \frac{h^2}{8}\|g''\|_{L^\infty([-1,1])} + h\|g''\|_{L^\infty([-1,1])} \leqslant \frac{19}{16}Bh \leqslant \frac{2B}{m}.$$

This proves (6.6).

Next, it is straightforward to verify that g_m can be rewritten as a two-layer ReLU neural network

$$g_m(z) = c + \sum_{i=1}^{m} a_i \,\mathrm{ReLU}\,(z_i - z) + \sum_{i=m+1}^{2m} a_i \,\mathrm{ReLU}\,(z - z_{i-1}), \quad z \in [-1,1], \tag{B.7}$$

where $c = g(z_m) = g(\rho)$ and the parameters a_i defined by

$$a_i = \begin{cases} (g(z_{m+1}) - g(z_m))/h_2, & \text{if } i = m+1,\\ (g(z_{m-1}) - g(z_m))/h_1, & \text{if } i = m,\\ (g(z_i) - 2g(z_{i-1}) + g(z_{i-2}))/h_2, & \text{if } i > m+1,\\ (g(z_{i-1}) - 2g(z_i) + g(z_{i+1}))/h_1, & \text{if } i < m. \end{cases}$$

Furthermore, by the mean value theorem, there exists $\xi_1, \xi_2 \in (z_m, z_{m+1})$ such that

$$|a_{m+1}| = |g'(\xi_1)| = |g'(\xi_1) - g'(\rho)| = |g''(\xi_2)\xi_1| \leqslant Bh_2.$$

In a similar manner we obtain $|a_m| \leqslant Bh_1$, $|a_i| \leqslant 2Bh_2$ if $i > m+1$ and $|a_i| \leqslant 2Bh_1$ if $i < m$. Therefore,

$$\sum_{i=1}^{2m} |a_i| \leqslant 2Bmh_1 + 2Bmh_2 = 4B.$$

Finally, setting $\epsilon_i = -1$, $b_i = -z_i$ for $i = 1, \cdots, m$ and $\epsilon_i = 1$, $b_i = z_{i-1}$ for $i = m+1, \cdots, 2m$, we obtain that g_m satisfies (6.5). This completes the proof. □

Next, we prove the approximation results with the Softplus activation. To this end, we recall a lemma from [47] which shows that ReLU may be well approximated by SP_τ for $\tau \gg 1$.

Lemma B.5. *[47, Lemma 4.6] The following inequalities hold:*

$$|\mathrm{ReLU}(z) - \mathrm{SP}_\tau(z)| \leqslant \tau^{-1}\mathrm{e}^{-\tau|\mathrm{z}|}, \quad \forall \mathrm{z} \in [-2,2], \tag{B.8a}$$

$$\left|\mathrm{ReLU}'(z) - \mathrm{SP}'_\tau(z)\right| \leqslant \mathrm{e}^{-\tau|\mathrm{z}|}, \quad \forall \mathrm{z} \in [-2,0) \cup (0,2], \tag{B.8b}$$

$$\|\mathrm{SP}_\tau\|_{W^{1,\infty}([-2,2])} \leqslant 3 + \tau^{-1}. \tag{B.8c}$$

Proof of Lemma 6.6. Thanks to Lemma 6.5, there exists g_m of the form (B.7) such that

$$\|g - g_m\|_{W^{1,\infty}([-1,1])} \leqslant \frac{2B}{m}.$$

Let $g_{\tau,m}$ be the function obtained by replacing ReLU in g_m by SP_τ, i.e.,

$$g_{\tau,m}(z) = c + \sum_{i=1}^{m} a_i \mathrm{SP}_\tau(z_i - z) + \sum_{i=m+1}^{2m} a_i \mathrm{SP}_\tau(z - z_{i-1}), \quad z \in [-1,1]. \tag{B.9}$$

By (B.8a) and the bounds of $|a_i|$ in Lemma 6.5, the difference $|g_m(z) - g_{\tau,m}(z)|$ may be bounded by

$$\begin{aligned}
&\sum_{i=1}^{m} |a_i|\,|\mathrm{ReLU}(z_i - z) - \mathrm{SP}_\tau(z_i - z)| + \sum_{i=m+1}^{2m} |a_i|\,|\mathrm{ReLU}(z - z_{i-1}) - \mathrm{SP}_\tau(z - z_{i-1})|\\
&\leqslant \sum_{i=1}^{m-1} \frac{2Bh_1}{\tau}\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z_i}|} + \frac{\mathrm{Bh_1}}{\tau}\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z_m}|} + \frac{\mathrm{Bh_2}}{\tau}\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z_m}|} + \sum_{\mathrm{i=m+2}}^{\mathrm{2m}} \frac{2\mathrm{Bh_2}}{\tau}\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z_{i-1}}|}\\
&\leqslant \frac{2B}{m\tau} + \frac{2B}{\tau}\left(\sum_{i=1}^{m-1} h_1\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z_i}|} + \sum_{\mathrm{i=m+1}}^{\mathrm{2m-1}} \mathrm{h_2}\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z_i}|}\right) =: \frac{2B}{m\tau} + \frac{2B}{\tau}I,
\end{aligned}$$

where we have used $\mathrm{e}^{-\tau|\mathrm{z}-\mathrm{z}_\mathrm{m}|} \leqslant 1$ and $h_1 + h_2 = 2/m$ in the last inequality. Suppose that $z \in (z_j, z_{j+1})$ for some fixed $0 \leqslant j \leqslant 2m-1$. When $j = 0$, the term $i = j+1 = 1$ in the sum I may be bounded by $h = \max(h_1, h_2) \leqslant 3/(2m)$. The other terms can be bounded by

$$\int_{z_{j+1}}^{z_{2m-1}} \mathrm{e}^{-\tau(\mathrm{x}-\mathrm{z})}\mathrm{dx} \leqslant \frac{1}{\tau}\left(1-\mathrm{e}^{-\tau(\mathrm{z}_{2\mathrm{m}-1}-\mathrm{z}_{\mathrm{j}+1})}\right) \leqslant \frac{1}{\tau}\left(1-\mathrm{e}^{-2\tau}\right).$$

Thus, $I \leqslant h + (1-\mathrm{e}^{-2\tau})/\tau$. Similar bound holds when $j = 2m-1$. When $1 \leqslant j \leqslant 2m-2$, the term $i = j$ and $i = j+1$ in the sum I is bounded by h, respectively. The other terms may be bounded by

$$\begin{aligned}\int_{z_1}^{z_j} \mathrm{e}^{-\tau(\mathrm{z}-\mathrm{x})}\mathrm{dx} + \int_{\mathrm{z}_{\mathrm{j}+1}}^{\mathrm{z}_{2\mathrm{m}-1}} \mathrm{e}^{-\tau(\mathrm{x}-\mathrm{z})}\mathrm{dx} &\leqslant \frac{1}{\tau}\left(2-\mathrm{e}^{-\tau(\mathrm{z}_\mathrm{j}-\mathrm{z}_1)}-\mathrm{e}^{-\tau(\mathrm{z}_{2\mathrm{m}-1}-\mathrm{z}_{\mathrm{j}+1})}\right)\\ &\leqslant \frac{2}{\tau}\left(1-\mathrm{e}^{-\tau(\mathrm{z}_{2\mathrm{m}-1}-\mathrm{z}_1)/2}\right) \leqslant \frac{2}{\tau}\left(1-\mathrm{e}^{-\tau}\right),\end{aligned}$$

where we have used Jensen's inequality in the second inequality. Since $2(1-\mathrm{e}^{-\tau}) > 1-\mathrm{e}^{-2\tau}$ for all $\tau > 0$, we summarize that $I \leqslant 2h + 2(1-\mathrm{e}^{-\tau})/\tau$ for all $0 \leqslant j \leqslant 2m-1$. Hence,

$$\|g_m - g_{\tau,m}\|_{L^\infty([-1,1])} \leqslant \frac{2B}{\tau}\left(\frac{1}{m} + 2h + \frac{2(1-\mathrm{e}^{-\tau})}{\tau}\right) \leqslant \frac{4B}{\tau}\left(\frac{2}{m} + \frac{1-\mathrm{e}^{-\tau}}{\tau}\right).$$

Using (B.8b), proceeding along the same line that leads to the above estimate, we obtain

$$\left\|g_m' - g_{\tau,m}'\right\|_{L^\infty([-1,1])} \leqslant 4B\left(\frac{2}{m} + \frac{1-\mathrm{e}^{-\tau}}{\tau}\right).$$

A combination of the above estimates and $\|g - g_m\|_{W^{1,\infty}([-1,1])} \leqslant 2B/m$ yields

$$\begin{aligned}\|g - g_{\tau,m}\|_{W^{1,\infty}([-1,1])} &\leqslant \|g - g_m\|_{W^{1,\infty}([-1,1])} + \|g_m - g_{\tau,m}\|_{W^{1,\infty}([-1,1])}\\ &\leqslant \frac{2B}{m} + 4B\left(1+\frac{1}{\tau}\right)\left(\frac{2}{m} + \frac{1-\mathrm{e}^{-\tau}}{\tau}\right)\\ &\leqslant 4B\left(1+\frac{1}{\tau}\right)\left(\frac{3}{m} + \frac{1-\mathrm{e}^{-\tau}}{\tau}\right).\end{aligned}$$

Since $\tau > 0$, choose $m \in \mathbb{N}_+$ such that $m \geqslant \max\{3\tau e^\tau, 2\}$. Then, we obtain (6.9). Finally, we rewrite (B.9) in the form (6.8), which completes the proof. □

Appendix B.5 Bounding the approximation error

Proof of Theorem 6.7. We denote φv_m by u_m. Since $u^* \in U_k$, $L_k(u^*) = \lambda_k$. Note that

$$L_k(u_m) - \lambda_k = \frac{\mathcal{E}_V(u_m) - \mathcal{E}_V(u^*)}{\mathcal{E}_2(u_m)} + \frac{\mathcal{E}_P(u_m) - \mathcal{E}_P(u^*)}{\mathcal{E}_2(u_m)} + \frac{\mathcal{E}_2(u^*) - \mathcal{E}_2(u_m)}{\mathcal{E}_2(u_m)} L_k(u^*). \tag{B.10}$$

Since $u^* \in \mathfrak{B}^s(\Omega)$ for some $s \geqslant 3$, by Theorem 3.2, we have

$$\|u^* - u_m\|_{H^1(\Omega)} \leqslant \eta(B_{u^*}, m) \leqslant 1/2 \quad \text{with} \quad B_{u^*} = (1+2/\pi)\|u^*\|_{\mathfrak{B}^s(\Omega)}.$$

Combining this inequality with $\|u^*\|_{L^2(\Omega)}^2 = 1$ yields

$$1/2 \leqslant 1 - \|u^* - u_m\|_{H^1(\Omega)} \leqslant \|u_m\|_{L^2(\Omega)} \leqslant 1 + \|u^* - u_m\|_{H^1(\Omega)} \leqslant 3/2, \tag{B.11}$$

and

$$\begin{aligned}|\mathcal{E}_2(u_m) - \mathcal{E}_2(u^*)| &= \left(\|u^*\|_{L^2(\Omega)} + \|u_m\|_{L^2(\Omega)}\right)\left|\|u^*\|_{L^2(\Omega)} - \|u_m\|_{L^2(\Omega)}\right|\\ &\leqslant (2+\eta(B_{u^*}, m))\,\eta(B_{u^*}, m)\\ &\leqslant 5\eta(B_{u^*}, m)/2.\end{aligned} \tag{B.12}$$

Since $a^2 - b^2 = (a-b)^2 + 2b(a-b)$, it follows from $0 \leqslant V \leqslant V_{\max}$ that

$$
\begin{aligned}
|\mathcal{E}_V(u) - \mathcal{E}_V(u^*)| &\leqslant \int_\Omega \Big(|\nabla u - \nabla u^*|^2 + 2\,|\nabla u^*|\,|\nabla u - \nabla u^*|\Big)\,\mathrm{d}\,x \\
&\quad + \int_\Omega \Big(V\,|u-u^*|^2 + 2V\,|u^*|\,|u-u^*|\Big)\,\mathrm{d}\,x \\
&\leqslant \max\{1, V_{\max}\}\,\|u-u^*\|_{H^1(\Omega)}^2 + 2\,\|\nabla u^*\|_{L^2(\Omega)}\,\|\nabla u - \nabla u^*\|_{L^2(\Omega)} \\
&\quad + 2\left\|\sqrt{V}u^*\right\|_{L^2(\Omega)}\left\|\sqrt{V}(u-u^*)\right\|_{L^2(\Omega)} \\
&\leqslant \max\{1, V_{\max}\}\,\|u-u^*\|_{H^1(\Omega)}^2 + 2\left(\mathcal{E}_V(u^*)\,\mathcal{E}_V(u-u^*)\right)^{1/2} \\
&\leqslant \max\{1, V_{\max}\}\,\|u-u^*\|_{H^1(\Omega)}^2 + 2\sqrt{\lambda_k \max\{1, V_{\max}\}}\,\|u-u^*\|_{H^1(\Omega)},
\end{aligned}
\tag{B.13}
$$

where we have used $\mathcal{E}_V(u) = \|\nabla u\|_{L^2(\Omega)}^2 + \left\|\sqrt{V}u\right\|_{L^2(\Omega)}^2$ and $\mathcal{E}_V(u^*) = \lambda_k$. Hence,

$$
|\mathcal{E}_V(u_m) - \mathcal{E}_V(u^*)| \leqslant (3\max\{1, V_{\max}\}/2 + \lambda_k)\,\eta(B_{u^*}, m).
$$

Moreover, since $\{\psi_j\}_{j=1}^{k-1}$ are normalized orthogonal eigenfunctions,

$$
\begin{aligned}
|\mathcal{E}_P(u_m) - \mathcal{E}_P(u^*)| &\leqslant \beta\left(\sum_{j=1}^{k-1}\langle u_m + u^*, \psi_j\rangle^2\right)^{1/2}\left(\sum_{j=1}^{k-1}\langle u_m - u^*, \psi_j\rangle^2\right)^{1/2} \\
&\leqslant \beta\left(\|u^*\|_{L^2(\Omega)} + \|u_m\|_{L^2(\Omega)}\right)\|u^* - u_m\|_{L^2(\Omega)} \\
&\leqslant \frac{5}{2}\beta\eta(B_{u^*}, m).
\end{aligned}
$$

Substituting all the above estimates into (B.10), we obtain Theorem 6.7. □

Appendix C Missing proof in Section 7

Appendix C.1 Bounding the covering numbers

The following proposition gives an upper bound for the covering number $\mathcal{N}(\delta, \Theta, \rho_\Theta)$.

Proposition C.1. *[47, Proposition 5.1] Consider the metric space (Θ, ρ_Θ) with ρ_Θ defined in (7.3). Then for any $\delta > 0$, the covering number $\mathcal{N}(\delta, \Theta, \rho_\Theta)$ satisfies that*

$$
\mathcal{N}(\delta, \Theta, \rho_\Theta) \leqslant \frac{2C}{\delta}\left(\frac{3\Gamma}{\delta}\right)^m\left(\frac{3W}{\delta}\right)^{dm}\left(\frac{3T}{\delta}\right)^m.
$$

Proof of Proposition 7.4. **Bounding $\mathcal{N}\left(\delta, \mathcal{G}_m^1/M_1, \|\cdot\|_{L^2(Q)}\right)$.** For $\theta, \theta' \in \Theta$, by adding and subtracting terms, we may bound $|v_\theta(x) - v_{\theta'}(x)|$ by

$$
\begin{aligned}
&|c - c'| + \left|\sum_{i=1}^m \gamma_i\phi(w_i\cdot x - t_i) - \sum_{i=1}^m \gamma_i'\phi(w_i'\cdot x - t_i')\right| \\
&\leqslant |c-c'| + \sum_{i=1}^m |\gamma_i - \gamma_i'|\,|\phi(w_i\cdot x - t_i)| + \sum_{i=1}^m |\gamma_i'|\,|\phi(w_i\cdot x - t_i) - \phi(w_i'\cdot x - t_i')|.
\end{aligned}
\tag{C.1}
$$

Since ϕ satisfies Assumption 7.1, $|\phi(w_i\cdot x - t_i) - \phi(w_i'\cdot x - t_i')| \leqslant L\left(|w_i - w_i'|_1 + |t_i - t_i'|\right)$. Therefore, it follows from (C.1) that

$$
\begin{aligned}
|v_\theta(x) - v_{\theta'}(x)| &\leqslant |c-c'| + \phi_{\max}|\gamma - \gamma'|_1 + L\Gamma\left(\max_i |w_i - w_i'|_1 + |t - t'|_\infty\right) \\
&\leqslant (1 + \phi_{\max} + 2L\Gamma)\,\rho_\Theta(\theta, \theta').
\end{aligned}
\tag{C.2}
$$

Consequently, using 7.5 and C.2, we obtain

$$\begin{aligned}\left\|\varphi^2 v_\theta^2-\varphi^2 v_{\theta'}^2\right\|_{L^2(Q)} &\leqslant \left\|\varphi^2\right\|_{L^2(Q)}\left(\|v_\theta\|_*+\|v_\theta'\|_*\right)\|v_\theta-v_\theta'\|_*\\ &\leqslant 2\left(C+\Gamma\phi_{\max}\right)\left(1+\phi_{\max}+2L\Gamma\right)\rho_\Theta\left(\theta,\theta'\right)/d^2.\end{aligned}$$

Thus, for any $g_\theta, g_{\theta'}\in\mathcal{G}_m^1$, $\|g_\theta/M_1-g_{\theta'}/M_1\|_{L^2(Q)}\leqslant\Lambda_1\rho_\Theta\left(\theta,\theta'\right)/M_1$ and then

$$\mathcal{N}\left(\delta,\mathcal{G}_m^1/M_1,\|\cdot\|_{L^2(Q)}\right)\leqslant\mathcal{N}\left(M_1\delta/\Lambda_1,\Theta,\rho_\Theta\right)\leqslant\mathcal{M}\left(\delta,\Lambda_1/M_1,m,d\right),$$

where the second inequality follows from Proposition C.1 with δ replaced by $M_1\delta/\Lambda_1$.

Bounding $\mathcal{N}\left(\delta,\mathcal{G}_m^2/M_2,\|\cdot\|_{L^2(Q)}\right)$. As in (C.1), by adding and subtracting terms,

$$\begin{aligned}|\nabla v_\theta(x)-\nabla v_{\theta'}(x)| &\leqslant\sum_{i=1}^m|\gamma_i-\gamma_i'|\,|w_i|_1\,|\phi'\left(w_i\cdot x+t_i\right)|+\sum_{i=1}^m|\gamma_i'|\,|w_i-w_i'|_1\,|\phi'\left(w_i\cdot x+t_i\right)|\\ &\quad+\sum_{i=1}^m|\gamma_i'|\,|w_i'|_1\,|\phi'\left(w_i\cdot x+t_i\right)-\phi'\left(w_i'\cdot x+t_i'\right)|\\ &\leqslant W\phi_{\max}'|\gamma-\gamma'|_1+\Gamma\phi_{\max}'\max_i|w_i-w_i'|_1\\ &\quad+\Gamma WL'\left(\max_i|w_i-w_i'|_1+|t-t'|_\infty\right)\\ &\leqslant\left((W+\Gamma)\phi_{\max}'+2\Gamma WL'\right)\rho_\Theta\left(\theta,\theta'\right).\end{aligned}$$

Combining (C.2) and the above estimate, we obtain

$$\begin{aligned}\||\nabla(\varphi v_\theta)-\nabla(\varphi v_{\theta'})|\|_{L^2(Q)} &\leqslant\|\varphi\|_{L^2(Q)}\|\nabla v_\theta-\nabla v_{\theta'}\|_*+\|\nabla\varphi\|_{L^2(Q)}\|v_\theta-v_{\theta'}\|_*\\ &\leqslant\left[\left((W+\Gamma)\phi_{\max}'+2\Gamma WL'\right)/d+\pi\left(1+\phi_{\max}+2L\Gamma\right)\right]\rho_\Theta\left(\theta,\theta'\right).\end{aligned}$$

Using the fact $\max_{\theta\in\Theta}\||\nabla\left(\varphi v_\theta\right)|\|_*\leqslant\Gamma W\phi_{\max}'/d+\pi\left(C+\Gamma\phi_{\max}\right)$, we have

$$\begin{aligned}\left\||\nabla\left(\varphi v_\theta\right)|^2-|\nabla\left(\varphi v_{\theta'}\right)|^2\right\|_{L^2(Q)} &\leqslant\||\nabla\left(\varphi v_\theta\right)|+|\nabla\left(\varphi v_{\theta'}\right)|\|_*\|\nabla\left(\varphi v_\theta\right)-\nabla\left(\varphi v_{\theta'}\right)\|_{L^2(Q)}\\ &\leqslant\Lambda_{21}\rho_\Theta\left(\theta,\theta'\right),\end{aligned}$$

where $\Lambda_{21}\colon=2\left[\Gamma W\phi_{\max}'/d+\pi\left(C+\Gamma\phi_{\max}\right)\right]\left[\left((W+\Gamma)\phi_{\max}'+2\Gamma WL'\right)/d+\pi\left(1+\phi_{\max}+2L\Gamma\right)\right]$. It follows from $0\leqslant V\leqslant V_{\max}$ and (C.2) that

$$\begin{aligned}\left\|V\varphi^2v_\theta^2-V\varphi^2v_{\theta'}^2\right\|_{L^2(Q)} &\leqslant V_{\max}\|\varphi\|_*^2\left(\|v_\theta\|_*+\|v_{\theta'}\|_*\right)\|v_\theta-v_{\theta'}\|_*\\ &\leqslant 2V_{\max}\left(C+\Gamma\phi_{\max}\right)\left(1+\phi_{\max}+2L\Gamma\right)\rho_\Theta\left(\theta,\theta'\right)/d^2\\ &=:\Lambda_{22}\rho_\Theta\left(\theta,\theta'\right).\end{aligned}$$

Combining the last two estimates, for any $g_\theta, g_{\theta'}\in\mathcal{G}_m^2$, we get

$$\begin{aligned}\|g_\theta-g_{\theta'}\|_{L^2(Q)} &\leqslant\left\||\nabla\left(\varphi v_\theta\right)|^2-|\nabla\left(\varphi v_{\theta'}\right)|^2\right\|_{L^2(Q)}+\left\|V\varphi^2v_\theta^2-V\varphi^2v_{\theta'}^2\right\|_{L^2(Q)}\\ &\leqslant\left(\Lambda_{21}+\Lambda_{22}\right)\rho_\Theta\left(\theta,\theta'\right)=\Lambda_2\rho_\Theta\left(\theta,\theta'\right).\end{aligned}$$

Dividing both sides of the above inequality by M_2, we obtain

$$\mathcal{N}\left(\delta,\mathcal{G}_m^2/M_2,\|\cdot\|_{L^2(Q)}\right)\leqslant\mathcal{N}\left(M_2\delta/\Lambda_2,\Theta,\rho_\Theta\right)\leqslant\mathcal{M}\left(\delta,\Lambda_1/M_1,m,d\right),$$

where the second inequality follows from Proposition C.1 with δ replaced by $M_2\delta/\Lambda_2$.

Bounding $\mathcal{N}\left(\delta,\mathcal{G}_m^3/M_3,\|\cdot\|_{L^2(Q)}\right)$. It follows from (C.2) that

$$\begin{aligned}\|\varphi\psi_jv_\theta-\varphi\psi_jv_{\theta'}\|_{L^2(Q)} &\leqslant\|\varphi\|_*\|\psi_j\|_{L^2(Q)}\|v_\theta-v_{\theta'}\|_*\\ &\leqslant\|\psi_j\|_{L^2(Q)}\left(1+\phi_{\max}+2L\Gamma\right)\rho_\Theta\left(\theta,\theta'\right)/d.\end{aligned}$$

The bound for $\mathcal{N}\left(\delta,\mathcal{G}_m^3/M_3,\|\cdot\|_{L^2(Q)}\right)$ follows from a similar argument that leads to the bound of $\mathcal{N}\left(\delta,\mathcal{G}_m^2/M_2,\|\cdot\|_{L^2(Q)}\right)$. $\square$

Appendix C.2 Estimates of the expectation of suprema of empirical processes

Proof of Lemma 7.6. By standard symmetrization,

$$\mathbf{E}\left\|\sum_{i=1}^{n}\left(f\left(X_i\right)-\mathcal{P}f\right)\right\|_{\mathscr{F}} \leqslant 2\mathbf{E}\left\|\sum_{i=1}^{n}\epsilon_i\left(f\left(X_i\right)-\mathcal{P}f\right)\right\|_{\mathscr{F}}.$$

Next, we shall apply Lemma 7.5 to the centered class $\bar{\mathscr{F}} := \{f-\mathcal{P}f : f\in\mathscr{F}\}$. To this end, we verify the assumptions in Lemma 7.5 for $\bar{\mathscr{F}}$. If functions in $\mathscr{F}$ take values in $[-1,1]$, then functions in $\bar{\mathscr{F}}$ take values in $[-2,2]$. If F is a measurable envelope of $\mathscr{F}$, then $F+\mathcal{P}F$ is a measurable envelope of $\bar{\mathscr{F}}$. Since for any $f\in\mathscr{F}$ and $0<\tau<1$, there exists f_i with $\|f-f_i\|_{L^2(Q)}\leqslant\tau\|F\|_{L^2(Q)}$ for all probability measures Q, then $|\mathcal{P}\left(f-f_i\right)|\leqslant\mathcal{P}\left|f-f_i\right|\leqslant\tau\mathcal{P}F$. Hence,

$$\begin{aligned}\|(f-\mathcal{P}f)-(f_i-\mathcal{P}f_i)\|_{L^2(Q)} &\leqslant \|f-f_i\|_{L^2(Q)}+|\mathcal{P}\left(f-f_i\right)|\\ &\leqslant \tau\left(\|F\|_{L^2(Q)}+\mathcal{P}F\right)\leqslant\sqrt{2}\tau\|F+\mathcal{P}F\|_{L^2(Q)},\end{aligned}$$

which means that $\{f_i-\mathcal{P}f_i\}_{i=1}^{M}$ is a $\sqrt{2}\tau\|F+\mathcal{P}F\|_{L^2(Q)}$-net for $\bar{\mathscr{F}}$. Therefore, $\bar{\mathscr{F}}$ is a VC class and

$$\mathcal{N}\left(\tau\|F+\mathcal{P}F\|_{L^2(Q)},\bar{\mathscr{F}},\|\cdot\|_{L^2(Q)}\right)\leqslant\left(\sqrt{2}A/\tau\right)^{v}$$

for all probability measures Q and $0<\tau<1$. The proof is completed by applying Lemma 7.5 to $\bar{\mathscr{F}}$. □

Proof of Corollary 7.8. Applying Lemma 7.6 to $\mathscr{F}\left(\rho_j\right)$ with $\rho_j=2^{j/2}r$, we have

$$\mathcal{K}_n(\mathscr{F},r)\leqslant\max_{1\leqslant j\leqslant l}\frac{\mathbf{E}\left\|\mathcal{P}_n-\mathcal{P}\right\|_{\mathscr{F}(\rho_j)}}{\rho_{j-1}^2}\leqslant C\max_{1\leqslant j\leqslant l}\left(\frac{v}{n\rho_{j-1}^2}\ln\frac{A}{\rho_j}+\sqrt{\frac{2v}{n\rho_{j-1}^2}\ln\frac{A}{\rho_j}}\right).$$

Notice that the quantity in parenthesis decreases as j increases and the maximum value reaches at $j=1$, which completes the proof. □

Appendix C.3 Bounding $\mathcal{K}_n$ in the statistical error

Proof of Theorem 7.9. With (7.4) and (7.8), a direct calculation yields

$$\begin{aligned}&\Lambda_1\leqslant 19B\left(4+8B\right)/d^2,\\ &\Lambda_2\leqslant 68B\left(11+30B+17B\sqrt{m}/d\right)+19V_{\max}B\left(4+8B\right)/d^2,\\ &\Lambda_3\leqslant\|\psi\|_{L^2(Q)}\left(4+8B\right)/d,\end{aligned}\tag{C.3}$$

and for $B\geqslant 1$,

$$\begin{aligned}\mathcal{M}(\delta,\Lambda,m,d)&=2^{2m+1}3^{(d+2)m}B^{m+1}\left(\Lambda/\delta\right)^{(d+2)m+1}\\ &\leqslant\left(2^{\frac{3}{d+3}}B^{\frac{2}{d+3}}3\Lambda/\delta\right)^{(d+2)m+1}.\end{aligned}\tag{C.4}$$

For $\mathcal{G}_1/M_{\mathcal{F}}^2$, it has a measurable envelope $F\equiv 1$. By Lemma 7.4, (7.9), (C.3) and (C.4),

$$\begin{aligned}\mathcal{N}\left(\tau\|F\|_{L^2(Q)},\mathcal{G}_1/M_{\mathcal{F}}^2,\|\cdot\|_{L^2(Q)}\right)&\leqslant\mathcal{M}\left(\tau,\Lambda_1/M_{\mathcal{F}}^2,m,d\right)\\ &\leqslant\left(2^{\frac{3}{d+3}}B^{\frac{2}{d+3}}6B\left(4+8B\right)/\left(9.5B^2\tau\right)\right)^{(d+2)m+1}\\ &\leqslant\left(2^{\frac{3}{d+3}+3}B^{\frac{2}{d+3}}/\tau\right)^{(d+2)m+1}.\end{aligned}$$

Thus, we may take $A=2^{\frac{3}{d+3}+3}B^{\frac{2}{d+3}}$, $v=(d+2)m+1$, $\hat{r}=r/M_{\mathcal{F}}$ and

$$\frac{v}{n\hat{r}^2}\ln\frac{A}{\hat{r}}\leqslant C_1\frac{mB^2}{ndr^2}\ln\frac{B}{rd}=:I_1,$$

where C_1 is an absolute constant. If $I_1 \leqslant 1$, then we apply Corollary 7.8 to $\mathcal{G}_1/M_{\mathcal{F}}^2$, and obtain (7.11a).

For $\mathcal{G}_2/M_{\mathcal{G}_2}$, it has a measurable envelope $F \equiv 1$. By Lemma 7.4, (7.9), (C.3) and (C.4),

$$\begin{aligned}\mathcal{N}\left(\tau\|F\|_{L^2(Q)}, \mathcal{G}_2/M_{\mathcal{G}_2}, \|\cdot\|_{L^2(Q)}\right) &\leqslant \mathcal{M}\left(\tau, \Lambda_2/M_{\mathcal{G}_2}, m, d\right)\\ &\leqslant \left(2^{\frac{3}{d+3}} B^{\frac{2}{d+3}}\left(8+3\sqrt{m}/d\right)/\tau\right)^{(d+2)m+1}.\end{aligned}$$

Thus, we may take $A = 2^{\frac{3}{d+3}} B^{\frac{2}{d+3}}(8+3\sqrt{m}/d)$, $v=(d+2)m+1$, $\hat{r} = r\sqrt{\lambda_1/M_{\mathcal{G}_2}}$ and thus

$$\begin{aligned}\frac{v}{n\hat{r}^2}\ln\frac{A}{\hat{r}} &\leqslant \tilde{C}_2\frac{dm\left(B^2+V_{\max}\left(B/d\right)^2\right)}{n\lambda_1 r^2}\ln\frac{B^{\frac{2}{d+3}}\left(1+\sqrt{m}/d\right)\left(B^2+V_{\max}\left(B/d\right)^2\right)}{r\sqrt{\lambda_1}}\\ &\leqslant C_2\frac{mB^2\left(1+V_{\max}\right)}{nr^2}\ln\frac{B\left(1+\sqrt{m}/d\right)\left(1+V_{\max}\right)}{rd} =: I_2,\end{aligned}$$

where we have used $\lambda_1 \geqslant d\pi^2$ in the second inequality. If $I_2 \leqslant 1$, then (7.11b) follows from applying Corollary 7.8 to $\mathcal{G}_2/M_{\mathcal{G}_2}$. Let $C_0 = \max\{C_1, C_2\}$, and then (7.10) ensures both $I_1 \leqslant 1$ and $I_2 \leqslant 1$.

To bound $\mathcal{K}_n(\mathcal{F}_j/(2\mu_j M_{\mathcal{F}})+1/2, \sqrt{\hat{r}/4\mu_j M_{\mathcal{F}}})$, recall (5.11) for the choice of $\sigma_{\mathcal{P}}(f)$, and note that

$$\mathcal{K}_n\left(\frac{\mathcal{F}_j}{2\mu_j M_{\mathcal{F}}}+\frac{1}{2}, \sqrt{\frac{r}{4\mu_j M_{\mathcal{F}}}}\right) = \frac{1}{2}\max_{1\leqslant i\leqslant l}\frac{1}{\rho_{i-1}^2}\mathbf{E}\sup_{\|u\|_{L^2}/(\mu_j M_{\mathcal{F}})\in\left(4\rho_{i-1}^2, 4\rho_i^2\right]}\frac{\left|\mathcal{P}_n\left(u\psi_j\right)-\mathcal{P}\left(u\psi_j\right)\right|}{\mu_j M_{\mathcal{F}}}. \tag{C.5}$$

To estimate the expectation, we may apply Lemma 7.6 to the set

$$\left\{\frac{u\psi_j}{M_{\mathcal{F}}\mu_j}\,\middle|\, u\in\mathcal{F}, (\mu_j M_{\mathcal{F}})^{-1}\|u\|_{L^2}\in\left(4\rho_{i-1}^2, 4\rho_i^2\right]\right\},$$

which has a measurable envelope $F_j = \psi_j/\mu_j$. Taking $\psi = \psi_j$, by Lemma 7.4, (7.9), (C.3) and (C.4), we get

$$\begin{aligned}\mathcal{N}\left(\tau\|F_j\|_{L^2(Q)}, \mathcal{F}_j/(\mu_j M_{\mathcal{F}}), \|\cdot\|_{L^2(Q)}\right) &\leqslant \mathcal{M}\left(\tau\|\psi_j\|_{L^2(Q)}/\mu_j, \Lambda_{3,j}/(\mu_j M_{\mathcal{F}}), m, d\right)\\ &\leqslant \left(8B^{\frac{2}{d+3}}/\tau\right)^{(d+2)m+1}.\end{aligned}$$

Taking $A = 8B^{\frac{2}{d+3}}$, $v=(d+2)m+1$, $\sigma = 2\rho_i$ in Lemma 7.6, we obtain that for all $n\in\mathbb{N}$,

$$\mathbf{E}\sup_{\|u\|_{L^2}/(\mu_j M_{\mathcal{F}})\in\left(4\rho_{i-1}^2, 4\rho_i^2\right]}\frac{\left|\mathcal{P}_n\left(u\psi\right)-\mathcal{P}\left(u\psi\right)\right|}{\mu M_{\mathcal{F}}} \leqslant C\left(\frac{dm}{n}\ln\frac{B}{\rho_j}+\rho_j\sqrt{\frac{dm}{n}\ln\frac{B}{\rho_j}}\right),$$

where C is an absolute constant. Therefore, substituting the above bound into (C.5) and noting that $\rho_i = 2^{i/2}\sqrt{r/(4\mu_j M_{\mathcal{F}})} = 2^{i/2}\sqrt{rd/(38\mu_j B)}$, we obtain (7.11c). □

Appendix D Missing proof in section 9

Appendix D.1 Some useful facts on sine and cosine series

Assume that $u\in L^1(\Omega)$ admits the sine series expansion $u(x) = \sum_{k\in\mathbb{N}_+^d}\hat{u}(k)\Phi_k(x)$, where $\hat{u}(k)$ are the sine expansion coefficients, i.e.,

$$\hat{u}(k) = \frac{\int_\Omega u(x)\Phi_k(x)\,\mathrm{d}\,x}{\int_\Omega \Phi_k^2(x)\,\mathrm{d}\,x} = 2^d\int_\Omega u(x)\Phi_k(x)\,\mathrm{d}\,x. \tag{D.1}$$

Let $\widetilde{\Omega} := [-1,1]^d$ and define the odd extension u_o of a function u by

$$u_o(x) = u_o\left(x_1, x_2, \ldots, x_d\right) = \operatorname{sign}\left(x_1 x_2\cdots x_d\right) u\left(\left|x_1\right|, \left|x_2\right|, \ldots, \left|x_d\right|\right), \quad x\in\widetilde{\Omega},$$

where $\operatorname{sign}(y) = \mathbf{1}_{\{y>0\}} - \mathbf{1}_{\{y<0\}}$. Let $\tilde{u}_o(k)$ be the Fourier coefficients of u_o, i.e., $u_o(x) = \sum_{k\in\mathbb{Z}^d}\tilde{u}_o(k)e^{i\pi k\cdot x}$, where $\tilde{u}_o(k) = 2^{-d}\int_{\widetilde{\Omega}} u_o(x)e^{-i\pi k\cdot x}\,\mathrm{d}\,x$. By abuse of notation, we use $|k|$ to stand for the vector $(|k_1|, |k_2, |, \dots, |k_d|)$ in this section and denote by $|k|_2$ the Euclid norm of the vector k. Since u_o is real and odd, $\tilde{u}_o(k) = \operatorname{sign}(k_1 k_2 \cdots k_d)\, \tilde{u}_o(|k|)$. Particularly, when d is odd, $u_o(x) = -u_o(-x)$, $\tilde{u}_o(k) = -\tilde{u}_o(-k)$ and

$$u_o(x) = \sum_{k \in \mathbb{Z}^d} \tilde{u}_o(k) i \sin(\pi k \cdot x) \quad \text{with} \quad i\tilde{u}_o(k) = 2^{-d} \int_{\widetilde{\Omega}} u_o(x) \sin(\pi k \cdot x) \,\mathrm{d}\, x. \tag{D.2}$$

When d is even, $u_o(x) = u_o(-x)$, $\tilde{u}_o(k) = \tilde{u}_o(-k)$ and

$$u_o(x) = \sum_{k \in \mathbb{Z}^d} \tilde{u}_o(k) \cos(\pi k \cdot x) \quad \text{with} \quad \tilde{u}_o(k) = 2^{-d} \int_{\widetilde{\Omega}} u_o(x) \cos(\pi k \cdot x) \,\mathrm{d}\, x. \tag{D.3}$$

We may extend $\{\hat{u}(k)\}$ from a sequence on $\mathbb{N}_+^d$ to a sequence on $\mathbb{N}_0^d$ by letting $\hat{u}(k) = 0$ if $k \in \mathbb{N}_0^d \setminus \mathbb{N}_+^d$. The relation between $\tilde{u}_o(k)$ and $\hat{u}(k)$ is established in the following lemma.

Lemma D.1. *For every $k \in \mathbb{Z}^d$, there holds*

$$\begin{aligned} i\tilde{u}_o(k) &= 2^{-d}(-1)^{\frac{d-1}{2}} \operatorname{sign}(k_1 k_2 \cdots k_d)\, \hat{u}(|k|), && \text{if } d \text{ is odd;} \\ \tilde{u}_o(k) &= 2^{-d}(-1)^{\frac{d}{2}} \operatorname{sign}(k_1 k_2 \cdots k_d)\, \hat{u}(|k|), && \text{if } d \text{ is even.} \end{aligned}$$

Proof. When d is odd, by the oddness of $u_o(x)$, we obtain

$$\begin{aligned} \int_{\widetilde{\Omega}} u_o(x) \sin(\pi k \cdot x) \mathrm{d}\, x &= \underbrace{\int_{\widetilde{\Omega}} u_o(x) \sin\left(\pi \left(\sum_{i=1}^{d-1} k_i x_i\right)\right) \cos(\pi k_d x_d) \,\mathrm{d}\, x}_{=0} \\ &\quad + \int_{\widetilde{\Omega}} u_o(x) \cos\left(\pi \left(\sum_{i=1}^{d-1} k_i x_i\right)\right) \sin(\pi k_d x_d) \,\mathrm{d}\, x \\ &= \underbrace{\int_{\widetilde{\Omega}} u_o(x) \cos\left(\pi \left(\sum_{i=1}^{d-2} k_i x_i\right)\right) \cos(\pi k_{d-1} x_{d-1}) \sin(\pi k_d x_d) \,\mathrm{d}\, x}_{=0} \\ &\quad - \int_{\widetilde{\Omega}} u_o(x) \sin\left(\pi \left(\sum_{i=1}^{d-2} k_i x_i\right)\right) \sin(\pi k_{d-1} x_{d-1}) \sin(\pi k_d x_d) \,\mathrm{d}\, x \\ &= (-1)^{\frac{d-1}{2}} \int_{\widetilde{\Omega}} u_o(x) \prod_{i=1}^{d} \sin(\pi k_i x_i) \,\mathrm{d}\, x \\ &= (-1)^{\frac{d-1}{2}} 2^d \operatorname{sign}(k_1 k_2 \cdots k_d) \int_{\Omega} u_o(x) \Phi_{|k|}(x) \mathrm{d}\, x. \end{aligned}$$

According to (D.2) and (D.1), for every $k \in \mathbb{Z}^d$, we have

$$i\tilde{u}_o(k) = 2^{-d} \int_{\widetilde{\Omega}} u_o(x) \sin(\pi k \cdot x) \,\mathrm{d}\, x = 2^{-d}(-1)^{\frac{d-1}{2}} \operatorname{sign}(k_1 k_2 \cdots k_d)\, \hat{u}(|k|).$$

When d is even, similarly, by oddness of $u_o(x)$,

$$\begin{aligned} \int_{\widetilde{\Omega}} u_o(x) \cos(\pi k \cdot x) \,\mathrm{d}\, x &= \underbrace{\int_{\widetilde{\Omega}} u_o(x) \cos\left(\pi \left(\sum_{i=1}^{d-1} k_i x_i\right)\right) \cos(\pi k_d x_d) \,\mathrm{d}\, x}_{=0} \\ &\quad - \int_{\widetilde{\Omega}} u_o(x) \sin\left(\pi \left(\sum_{i=1}^{d-1} k_i x_i\right)\right) \sin(\pi k_d x_d) \,\mathrm{d}\, x. \end{aligned}$$

Then, we have

$$
\begin{aligned}
\int_{\widetilde{\Omega}} u_o(x)\cos(\pi k\cdot x)\,\mathrm{d}\,x &= \underbrace{-\int_{\widetilde{\Omega}} u_o(x)\sin\left(\pi\left(\sum_{i=1}^{d-2}k_i x_i\right)\right)\cos\left(\pi k_{d-1}x_{d-1}\right)\sin\left(\pi k_d x_d\right)\mathrm{d}\,x}_{=0}\\
&\quad -\int_{\widetilde{\Omega}} u_o(x)\sin\left(\pi\left(\sum_{i=1}^{d-2}k_i x_i\right)\right)\sin\left(\pi k_{d-1}x_{d-1}\right)\sin\left(\pi k_d x_d\right)\,\mathrm{d}\,x\\
&= (-1)^{\frac{d}{2}}\int_{\widetilde{\Omega}} u_o(x)\prod_{i=1}^{d}\sin\left(\pi k_i x_i\right)\,\mathrm{d}\,x\\
&= (-1)^{\frac{d}{2}}2^d\,\mathrm{sign}\left(k_1k_2\cdots k_d\right)\int_{\Omega} u_o(x)\Phi_{|k|}(x)\,\mathrm{d}\,x.
\end{aligned}
$$

According to (D.3) and (D.1), for every $k\in\mathbb{Z}^d$, we have

$$
\tilde{u}_o(k) = 2^{-d}\int_{\widetilde{\Omega}} u_o(x)\cos(\pi k\cdot x)\mathrm{d}\,x = 2^{-d}(-1)^{\frac{d}{2}}\,\mathrm{sign}\left(k_1k_2\cdots k_d\right)\hat{u}(|k|).
$$

The lemma is proved. □

Assume that $V(x)\in L^1(\Omega)$ admits the cosine series expansion $V(x)=\sum_{k\in\mathbb{N}_0^d}\check{V}(k)\Psi_k$, where

$$
\check{V}(k) = 2^{\sum_{i=1}^{d}\mathbf{1}_{k_i\neq 0}}\int_{\Omega} V(x)\left(\prod_{i=1}^{d}\cos\left(k_i\pi x_i\right)\right)\mathrm{d}\,x.
$$

Define the even extension V_e of the function V by

$$
V_e(x) = V_e\left(x_1,\cdots,x_d\right) = V\left(\left|x_1\right|,\cdots,\left|x_d\right|\right),\quad x\in\widetilde{\Omega}.
$$

Let $\tilde{V}_e(k)$ be the Fourier coefficients of V_e. Since V_e is real and even, $V_e(x)=\sum_{k\in\mathbb{Z}^d}\tilde{V}_e(k)\cdot\cos(\pi k\cdot x)$, where

$$
\tilde{V}_e(k) = \frac{\int_{\widetilde{\Omega}} V_e(x)\cos(\pi k\cdot x)\mathrm{d}\,x}{\int_{\widetilde{\Omega}}\cos^2(\pi k\cdot x)\,\mathrm{d}\,x} = 2^{-d+\mathbf{1}_{k\neq 0}}\int_{\widetilde{\Omega}} V_e(x)\cos(\pi k\cdot x)\,\mathrm{d}\,x.
$$

It follows from [47, Lemma B.1] that the relation between $\tilde{V}_e(k)$ and $\check{V}(k)$ reads as

Lemma D.2. *[47, Lemma B.1] For every $k\in\mathbb{Z}^d$, $\tilde{V}_e(k)=\beta_k\check{V}(|k|)$ where $\beta_k=2^{\mathbf{1}_{k\neq 0}-\sum_{i=1}^{d}\mathbf{1}_{k_i\neq 0}}$.*

Let $w(x)=u(x)V(x)$ in Ω and its odd extension $w_o(x)=u_o(x)V_e(x)$ in $\widetilde{\Omega}$ with Fourier coefficients $\tilde{w}_o(k)$. By the oddness of $w_o(x)$, we have $w_o(x)=\sum_{k\in\mathbb{Z}^d}i\tilde{w}_o(k)\sin(\pi k\cdot x)$, where $\tilde{w}_o(k)=\mathrm{sign}\left(k_1k_2\cdots k_d\right)\tilde{w}_o(|k|)$. By the properties of Fourier transform,

$$
\tilde{w}_o(k) = \sum_{m\in\mathbb{Z}^d}\tilde{u}_o(m)\tilde{V}_e(k-m). \tag{D.4}
$$

Similar to $u_o(x)$, $w_o(x)$ admits the sine series expansion $w_o(x)=\sum_{k\in\mathbb{N}_+^d}\hat{w}(k)\Phi_k(x)$ on Ω.

The following proposition gives a representation of $\hat{w}(k)$ in terms of $\hat{u}(k)$ and $\check{V}(k)$.

Proposition D.3. *Let $\beta_k=2^{\mathbf{1}_{k\neq 0}-\sum_{i=1}^{d}\mathbf{1}_{k_i\neq 0}}$. For any $k\in\mathbb{N}_+^d$, there holds*

$$
\hat{w}(k) = \widehat{(uV)}(k) = \sum_{m\in\mathbb{Z}^d}\mathrm{sign}\left(m_1m_2\cdots m_d\right)\beta_{|m-k|}\hat{u}(|m|)\check{V}(|k-m|).
$$

Proof. Thanks to Lemma D.1, Lemma D.2 and relation (D.4), when d is odd, for each $k\in\mathbb{N}_+^d$,

$$
\begin{aligned}
\hat{w}(k) &= 2^d(-1)^{\frac{d-1}{2}}i\tilde{w}_o(k) = 2^d(-1)^{\frac{d-1}{2}}\sum_{m\in\mathbb{Z}^d}i\tilde{u}_o(m)\tilde{V}_e(k-m)\\
&= 2^d(-1)^{\frac{d-1}{2}}\sum_{m\in\mathbb{Z}^d}2^{-d}(-1)^{\frac{d-1}{2}}\,\mathrm{sign}\left(m_1m_2\cdots m_d\right)\hat{u}(|m|)\cdot\beta_{|m-k|}\check{V}(|k-m|)\\
&= \sum_{m\in\mathbb{Z}^d}\mathrm{sign}\left(m_1m_2\cdots m_d\right)\beta_{|m-k|}\hat{u}(|m|)\check{V}(|k-m|),
\end{aligned}
$$

where $\beta_{|m-k|} = 2^{\mathbf{1}_{k\neq m}-\sum_{i=1}^{d}\mathbf{1}_{k_i\neq m_i}}$.

Similarly, when d is even, for each $k\in\mathbb{N}_+^d$,

$$\begin{aligned}\hat{w}(k) &= 2^d(-1)^{\frac{d}{2}}\tilde{w}_o(k) = 2^d(-1)^{\frac{d}{2}}\sum_{m\in\mathbb{Z}^d}\tilde{u}_o(m)\tilde{V}_e(k-m)\\ &= 2^d(-1)^{\frac{d}{2}}\sum_{m\in\mathbb{Z}^d}2^{-d}(-1)^{\frac{d}{2}}\,\mathrm{sign}\,(m_1m_2\cdots m_d)\,\hat{u}(|m|)\cdot\beta_{|m-k|}\check{V}(|k-m|)\\ &= \sum_{m\in\mathbb{Z}^d}\mathrm{sign}\,(m_1m_2\cdots m_d)\,\beta_{|m-k|}\hat{u}(|m|)\check{V}(|k-m|),\end{aligned}$$

where $\beta_{|m-k|} = 2^{\mathbf{1}_{k\neq m}-\sum_{i=1}^{d}\mathbf{1}_{k_i\neq m_i}}$. □

Appendix D.2 Boundedness of $\mathcal{H}^{-1}:\mathfrak{B}^s(\Omega)\to\mathfrak{B}^{s+2}(\Omega)$

Proof of Theorem 9.1. It is clear that there exists a unique solution $u\in H_0^1(\Omega)$ such that

$$\|\nabla u\|_{L^2(\Omega)}^2 \leqslant \|f\|_{L^2(\Omega)}\|u\|_{L^2(\Omega)}. \tag{D.5}$$

To show $u\in\mathfrak{B}^{s+2}(\Omega)$, we firstly derive an operator equation that is equivalent to (9.1). Multiplying Φ_k on both sides of (9.1) and then integrating the resulting equation, we get

$$\pi^2|k|_2^2\hat{u}(k)+\widehat{(Vu)}(k) = \hat{f}(k),\quad k\in\mathbb{N}_+^d. \tag{D.6}$$

Using Proposition D.3, we rewrite (D.6) as

$$\pi^2|k|_2^2\hat{u}(k)+\sum_{m\in\mathbb{Z}^d}\mathrm{sign}\,(m_1m_2\cdots m_d)\,\beta_{|m-k|}\hat{u}(|m|)\check{V}(|k-m|) = \hat{f}(k),\quad k\in\mathbb{N}_+^d,$$

where $\beta_k = 2^{\mathbf{1}_{k\neq 0}-\sum_{i=1}^{d}\mathbf{1}_{k_i\neq 0}}$. Define the operator $\mathbb{M}:\hat{u}\mapsto\mathbb{M}\hat{u}$ by

$$(\mathbb{M}\hat{u})(k) = \pi^2|k|_2^2\hat{u}(k),\quad k\in\mathbb{N}_+^d.$$

We may extend $\hat{u}(k)$ as 0 when $k\in\mathbb{N}_0^d\setminus\mathbb{N}_+^d$. Define the operator $\mathbb{V}:\hat{u}\mapsto\mathbb{V}\hat{u}$ by

$$(\mathbb{V}\hat{u})(k) = \sum_{m\in\mathbb{Z}^d}\mathrm{sign}\,(m_1m_2\cdots m_d)\,\beta_{|m-k|}\hat{u}(|m|)\check{V}(|k-m|),\quad k\in\mathbb{N}_+^d.$$

We rewrite (D.6) as

$$(\mathbb{M}+\mathbb{V})\hat{u} = \hat{f}. \tag{D.7}$$

Since the diagonal operator $\mathbb{M}$ is invertible, this operator equation is equivalent to

$$\left(\mathbb{I}+\mathbb{M}^{-1}\mathbb{V}\right)\hat{u} = \mathbb{M}^{-1}\hat{f}. \tag{D.8}$$

Next, we claim that equation (D.8) has a unique solution $\hat{u}\in\ell_{W_s}^1\left(\mathbb{N}_+^d\right)$ and there exists C_1 depending on V and d such that

$$\|\hat{u}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)} \leqslant C_1(V,d)\|\hat{f}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)}. \tag{D.9}$$

It follows from the compactness of $\mathbb{M}^{-1}\mathbb{V}$ as shown in Lemma D.5 that $\mathbb{I}+\mathbb{M}^{-1}\mathbb{V}$ is a Fredholm operator on $\ell_{W_s}^1\left(\mathbb{N}_+^d\right)$. By the celebrated Fredholm alternative theorem, the operator $\mathbb{I}+\mathbb{M}^{-1}\mathbb{V}$ has a bounded inverse $\left(\mathbb{I}+\mathbb{M}^{-1}\mathbb{V}\right)^{-1}$ if and only if $\left(\mathbb{I}+\mathbb{M}^{-1}\mathbb{V}\right)\hat{u}=0$ has a trivial solution. By the equivalence between equation (9.1) and (D.8), we only need to show that the only solution of (9.1) is zero when $f=0$, which is a direct consequence of (D.5) and the Poincaré's inequality.

It follows from (D.7) and the boundedness of $\mathbb{V}$ on $\ell_{W_s}^1\left(\mathbb{N}_+^d\right)$ proved in Lemma D.5 that

$$\begin{aligned}\|\mathbb{M}\hat{u}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)} &\leqslant \|\mathbb{V}\hat{u}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)}+\|\hat{f}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)}\\ &\leqslant C_2(V,d)\|\hat{u}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)}+\|\hat{f}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)}\\ &\leqslant C_3(V,d)\|\hat{f}\|_{\ell_{W_s}^1(\mathbb{N}_+^d)},\end{aligned} \tag{D.10}$$

where we have used (D.9) in the last step. Therefore, the estimate (D.10) implies

$$
\begin{aligned}
\|u\|_{\mathcal{B}^{s+2}(\Omega)} &= \sum_{k\in\mathbb{N}_+^d} \left(1+\pi^{s+2}|k|_1^{s+2}\right)|\hat{u}(k)| = \sum_{k\in\mathbb{N}_+^d} \frac{1+\pi^{s+2}|k|_1^{s+2}}{\pi^2|k|_2^2}\cdot\pi^2|k|_2^2\,|\hat{u}(k)| \\
&\leqslant \left(\pi^{-2}+d\right)\|\mathbb{M}\hat{u}\|_{\ell^1_{W_s}\left(\mathbb{N}_+^d\right)} \leqslant C(d,V)\|\hat{f}\|_{\ell^1_{W_s}\left(\mathbb{N}_+^d\right)},
\end{aligned}
$$

which completes the proof. □

Proceeding along the same line that leads to [47, Lemma 7.2], we obtain

Lemma D.4. *Suppose that $\mathbb{T}$ is a multiplication operator on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$ defined for each $a=(a(k))_{k\in\mathbb{N}_+^d}$ that $(\mathbb{T}a)_k=\lambda_k a_k$ with $\lambda_k\to 0$ as $|k|_2\to\infty$. Then $\mathbb{T}:\ell^1_{W_s}\left(\mathbb{N}_+^d\right)\to\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$ is compact.*

The following lemma shows that the operator $\mathbb{V}$ is bounded on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$.

Lemma D.5. *Assume that $V\in\mathfrak{C}^s(\Omega)$. Then the operator $\mathbb{V}$ is bounded on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$ and the operator $\mathbb{M}^{-1}\mathbb{V}$ is compact on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$.*

Proof. Since $\mathbb{M}^{-1}$ is a multiplication operator on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$ with the diagonal entries converging to zero, it follows from Lemma D.4 that $\mathbb{M}^{-1}$ is compact on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$. To show the compactness of $\mathbb{M}^{-1}\mathbb{V}$, it suffices to show that the operator $\mathbb{V}$ is bounded on $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$. Since $\beta_k=2^{\mathbf{1}_{k\neq 0}-\sum_{i=1}^d\mathbf{1}_{k_i\neq 0}}\in[2^{1-d},1]$ and $V\in\mathfrak{C}^s(\Omega)$, using Proposition D.3, one has that for any $\hat{u}\in\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$ with $\hat{u}(k)=0$ when $k\in\mathbb{N}_0^d\setminus\mathbb{N}_+^d$,

$$
\begin{aligned}
\|\mathbb{V}\hat{u}\|_{\ell^1_{W_s}\left(\mathbb{N}_+^d\right)} &= \sum_{k\in\mathbb{N}_+^d}\left(1+\pi^s|k|_1^s\right)\left|\sum_{m\in\mathbb{Z}^d}\operatorname{sign}\left(m_1m_2\cdots m_d\right)\beta_{|m-k|}\hat{u}(|m|)\check{V}(|k-m|)\right| \\
&\leqslant \sum_{m\in\mathbb{Z}^d}\sum_{k\in\mathbb{N}_+^d}\left(1+\pi^s\max\left(2^{s-1},1\right)\left(|m-k|_1^s+|m|_1^s\right)\right)|\hat{u}(|m|)|\left|\check{V}(|k-m|)\right| \\
&\leqslant 2^d\max\left(2^{s-1},1\right)\left(\|\hat{u}\|_{\ell^1\left(\mathbb{N}_+^d\right)}\|\check{V}\|_{\ell^1_{W_s}\left(\mathbb{N}_0^d\right)}+\|\hat{u}\|_{\ell^1_{W_s}\left(\mathbb{N}_+^d\right)}\|\check{V}\|_{\ell^1\left(\mathbb{N}_0^d\right)}\right) \\
&\leqslant 2^{d+1}\max\left(2^{s-1},1\right)\|V\|_{\mathfrak{C}^s(\Omega)}\|\hat{u}\|_{\ell^1_{W_s}\left(\mathbb{N}_+^d\right)},
\end{aligned}
$$

where we have used the elementary inequality $|a+b|^s\leqslant\max\left(2^{s-1},1\right)(|a|^s+|b|^s)$ in the first inequality and the fact $\sum_{m\in\mathbb{Z}^d}|\hat{u}(|m|)|\leqslant 2^d\|\hat{u}\|_{\ell^1\left(\mathbb{N}_+^d\right)}\leqslant 2^d\|\hat{u}\|_{\ell^1_{W_s}\left(\mathbb{N}_+^d\right)}$ in the second inequality. □

Proof of Corollary 9.2. Note that the operator $\mathcal{S}:\mathfrak{B}^s(\Omega)\to\mathfrak{B}^{s+2}(\Omega)$ is bounded, as proved in Theorem 9.1. It remains to show the inclusion $\mathcal{J}:\mathfrak{B}^{s+2}(\Omega)\hookrightarrow\mathfrak{B}^s(\Omega)$ is compact. Indeed, by definition, the space $\mathfrak{B}^s(\Omega)$ may be viewed as a weighted ℓ^1 space $\ell^1_{W_s}\left(\mathbb{N}_0^d\right)$ of the sine coefficients defined on the lattice $\mathbb{N}_+^d$ with the weight $W_s(k)=(1+\pi^s|k|_1^s)$. Therefore, the inclusion satisfies

$$
\|\mathcal{J}u\|_{\mathfrak{B}^s(\Omega)}=\sum_{k\in\mathbb{N}_+^d}W_s(k)|\hat{u}(k)|=\sum_{k\in\mathbb{N}_+^d}\frac{W_s(k)}{W_{s+2}(k)}W_{s+2}(k)|\hat{u}(k)|.
$$

Since $\frac{W_s(k)}{W_{s+2}(k)}\to 0$ as $|k|_2\to\infty$, by a similar argument as the proof of Lemma D.4, one can conclude that $\mathcal{J}$ is compact from $\ell^1_{W_{s+2}}\left(\mathbb{N}_+^d\right)$ to $\ell^1_{W_s}\left(\mathbb{N}_+^d\right)$ and hence from $\mathcal{B}^{s+2}(\Omega)$ to $\mathcal{B}^s(\Omega)$.

Consequently, Corollary 9.2 is a direct consequence of the boundness of $\mathcal{S}:\mathfrak{B}^s(\Omega)\to\mathfrak{B}^{s+2}(\Omega)$ and the compactness of the inclusion $\mathcal{J}$ from $\mathfrak{B}^{s+2}(\Omega)$ to $\mathfrak{B}^s(\Omega)$. □

Appendix E About the penalty method

Firstly, we prove that $\|u_n\|_{L^2(\Omega)}\geqslant 1/2$ with high probability when γ is sufficiently large. To this end, we decompose $\mathcal{E}_2(u_n)-1$ as

$$
\mathcal{E}_2(u_n)-1=\mathcal{E}_{n,2}(u_n)-1-\left(\mathcal{E}_{n,2}(u_n)-\mathcal{E}_2(u_n)\right)=:\mathcal{E}_{n,2}(u_n)-1-R_1,
$$

and for any $u_{\mathcal{F}} \in \mathcal{F}$,

$$\begin{aligned}\gamma\left(\mathcal{E}_{n,2}\left(u_n\right)-1\right)^2 &\leqslant \mathscr{L}_{k,n}\left(u_n\right) \leqslant \mathscr{L}_{k,n}\left(u_{\mathcal{F}}\right)\\ &=\Big(L_{k,n}(u_{\mathcal{F}})-L_k(u_{\mathcal{F}})\Big)+\Big(L_k(u_{\mathcal{F}})-\lambda_k\Big)+\lambda_k\\ &\quad+\gamma\left[\Big(\mathcal{E}_{n,2}(u_{\mathcal{F}})-\mathcal{E}_2(u_{\mathcal{F}})\Big)+\Big(\mathcal{E}_2(u_{\mathcal{F}})-1\Big)\right]^2\\ &=: R_2+R_3+\lambda_k+\gamma\left(R_4+R_5\right)^2,\end{aligned} \tag{E.1}$$

where the second inequality follows from the fact that u_n is a minimizer of $\mathscr{L}_{k,n}(u)$. Therefore,

$$|\mathcal{E}_2(u_n)-1| \leqslant |R_1|+\left[\frac{\lambda_k}{\gamma}+\frac{R_2+R_3}{\gamma}+\left(R_4+R_5\right)^2\right]^{1/2}. \tag{E.2}$$

Note that R_1 is the statistical error, R_2, R_4 are the Monte Carlo error, and R_3, R_5 are the approximation error.

Bounding R_1. To control R_1, we employ the well-known tool of Rademacher complexity. We recall the definition firstly.

Definition E.1. For a set of random variables $\{Z_j\}_{j=1}^n$ independently distributed according to P and a function class $\mathcal{G}$, the empirical Rademacher complexity is defined by

$$\widehat{\mathscr{R}}_n(\mathcal{G}) := \mathbf{E}_\sigma\left[\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^n \sigma_j g\left(Z_j\right)\right| \mid Z_1,\cdots,Z_n\right],$$

where the expectation $\mathbf{E}_\sigma$ is taken with respect to the independent uniform Bernoulli sequence $\{\sigma_j\}_{j=1}^n$ with $\sigma_j \in \{\pm 1\}$. The Rademacher complexity of $\mathcal{G}$ is defined by

$$\mathscr{R}_n(\mathcal{G}) = \mathbf{E}_{P^n}\left[\widehat{\mathscr{R}}_n(\mathcal{G})\right].$$

We introduce the following generalization bound via the Rademacher complexity.

Lemma E.2. *[64, Theorem 4.10] Let $\mathcal{G}$ be a class of integrable real valued functions such that $\sup_{g\in\mathcal{G}}\|g\|_{L^\infty(\Omega)} \leqslant M_{\mathcal{G}}$. Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. random samples from some distribution P over Ω. Then for any positive integer $n \geqslant 1$ and any scalar $\delta \geqslant 0$, with probability at least $1-\delta$,*

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n g\left(Z_i\right)-\mathbf{E}g(Z)\right| \leqslant 2\mathscr{R}_n(\mathcal{G})+M_{\mathcal{G}}\sqrt{\frac{2\ln(1/\delta)}{n}}.$$

Recall the function class $\mathcal{G}_1$ defined in (5.4) and $\sup_{u\in\mathcal{F}}\|u\|_{L^\infty(\Omega)} \leqslant M_{\mathcal{F}}$, hence $\sup_{g\in\mathcal{G}_1}\|g\|_{L^\infty(\Omega)} \leqslant M_{\mathcal{F}}^2$. Define for $n \in \mathbb{N}$ and $\delta > 0$ the constant

$$\xi_4(n,\delta) := 2\mathscr{R}_n\left(\mathcal{G}_1\right)+M_{\mathcal{F}}^2\sqrt{\frac{2\ln(1/\delta)}{n}},$$

and the event $A_4(n,\delta) := \{|\mathcal{E}_{n,2}\left(u_n\right)-\mathcal{E}_2\left(u_n\right)| \leqslant \xi_4(n,\delta)\}$. Applying Lemma E.2 to $\mathcal{G}_1$, we have

$$\mathbf{P}\left[A_4(n,\delta)\right] \geqslant 1-\delta. \tag{E.3}$$

Next, the celebrated Dudley's theorem will be used to bound the Rademacher complexity in terms of the metric entropy.

Theorem E.3. *[17] Let $\mathscr{F}$ be a class of real functions, $\{Z_i\}_{i=1}^n$ be random i.i.d. samples and the empirical measure $P_n = n^{-1}\sum_{i=1}^n \delta_{Z_i}$. If*

$$\sup_{f\in\mathscr{F}}\|f\|_{L^2(P_n)} := \sup_{f\in\mathscr{F}}\left(\frac{1}{n}\sum_{i=1}^n f^2\left(Z_i\right)\right)^{1/2} \leqslant c,$$

then

$$\widehat{\mathscr{R}}_n(\mathscr{F}) \leqslant \inf_{\epsilon\in[0,c/2]}\left(4\epsilon+\frac{12}{\sqrt{n}}\int_\epsilon^{c/2}\sqrt{\ln\mathcal{N}\left(\delta,\mathscr{F},\|\cdot\|_{L^2(P_n)}\right)}\,\mathrm{d}\delta\right).$$

By Lemma 7.4, (7.9), (C.3) and (C.4), the covering number

$$\mathcal{N}\left(\delta, \mathcal{G}_1, \|\cdot\|_{L^2(P_n)}\right) \leqslant \left(2^{\frac{3}{d+3}} B^{\frac{2}{d+3}} 3 \cdot 19B(4+8B)/(d^2\delta)\right)^{(d+2)m+1}$$
$$\leqslant \left(1368 B^{5/2}/(d^2\delta)\right)^{(d+2)m+1}.$$

Using Theorem E.3 and (7.9), we obtain, there exists an absolute constant C such that

$$\begin{aligned}
\widehat{\mathfrak{R}}_n(\mathcal{G}_1) &\leqslant 12\sqrt{\frac{(d+2)m+1}{n}} \int_0^{M_{\mathcal{F}}^2/2} \sqrt{\ln\left(1368B^{2.5}/(d^2\delta)\right)}\,\mathrm{d}\delta \\
&\leqslant C\sqrt{\frac{dm}{n}}\left(M_{\mathcal{F}}^2\sqrt{1+\ln(B/d)} + \int_0^{M_{\mathcal{F}}^2/2}\sqrt{(\ln(1/\delta))_+}\,\mathrm{d}\delta\right) \\
&\leqslant \frac{CB}{d}\left(\frac{B}{d}+1\right)\sqrt{\frac{d\,(1+\ln B)\,m}{n}},
\end{aligned} \tag{E.4}$$

where in the last inequality we have used

$$\int_0^{M_{\mathcal{F}}^2/2}\sqrt{(\ln(1/\delta))_+}\,\mathrm{d}\delta \leqslant \int_0^{\min(1, M_{\mathcal{F}}^2/2)} \sqrt{1/\delta}\,\mathrm{d}\delta \leqslant \min(2, \sqrt{2}M_{\mathcal{F}}).$$

By (E.4) and (7.9), on the event $A_4(n,\delta)$,

$$R_1 \leqslant \xi_4(n,\delta) \leqslant \frac{CB}{d}\left(\frac{B}{d}+1\right)\sqrt{\frac{d\,(1+\ln B)\,m + \ln(1/\delta)}{n}}. \tag{E.5}$$

Bounding R_3 and R_5. Let $\mathcal{F} = \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ with $B = \|u^*\|_{\mathfrak{B}^s(\Omega)}$ and $\tau = 9\sqrt{m}$. By Theorem 3.2, there exists $u_{\mathcal{F}} \in \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ such that $\|u^* - u_{\mathcal{F}}\|_{H^1(\Omega)} \leqslant 64B/\sqrt{m} \leqslant 1/2$. By Theorem 6.7,

$$R_3 = L_k\left(u_{\mathcal{F}}\right) - \lambda_k \leqslant 64\left(3\max\{1, V_{\max}\} + 7\lambda_k + 5\beta\right)B/\sqrt{m}. \tag{E.6}$$

Similar as in (B.11) and (B.12), we have $1/2 \leqslant \|u_{\mathcal{F}}\|_{L^2(\Omega)} \leqslant 3/2$ and

$$R_5 \leqslant \left(\|u_{\mathcal{F}}\|_{L^2(\Omega)} + \|u^*\|_{L^2(\Omega)}\right)\left|\|u_{\mathcal{F}}\|_{L^2(\Omega)} - \|u^*\|_{L^2(\Omega)}\right| \leqslant 160B/\sqrt{m}. \tag{E.7}$$

Bounding R_2 and R_4. As a preparation, we introduce Hoeffding's inequality to control the Monte Carlo error.

Lemma E.4 (Hoeffding's inequality). *[63, Theorem 2.2.6] Let Z_1, Z_2, $\cdots$, Z_n be independent random variables. Assume that $Z_i \in [m_i, M_i]$ for every i. Then, for any $t > 0$,*

$$\mathbf{P}\left(\sum_{i=1}^n (Z_i - \mathbf{E}Z_i) \geqslant t\right) \leqslant \exp\left(-\frac{2t^2}{\sum_{i=1}^n (M_i - m_i)^2}\right).$$

In particular, if $Z_1, Z_2, \cdots, Z_n$ are identically distributed with $|Z_i| \leqslant M$, then for any $t > 0$,

$$\mathbf{P}\left(\left|\frac{\sum_{i=1}^n Z_i}{n} - \mathbf{E}Z_1\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{nt^2}{2M^2}\right).$$

Consider $u_{\mathcal{F}} \in \varphi\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ given by Theorem 3.2. To bound R_4, we define the constant $\xi_5(n,\delta) := M_{\mathcal{F}}^2\sqrt{\ln(2/\delta)/(2n)}$ and the event $A_5(n,\delta) := \{|\mathcal{E}_{n,2}(u_{\mathcal{F}}) - \mathcal{E}_2(u_{\mathcal{F}})| \leqslant \xi_5(n,\delta)\}$. On the event $A_5(n,\delta)$,

$$R_4 \leqslant \xi_5(n,\delta). \tag{E.8}$$

Since $\|u_{\mathcal{F}}^2\|_{L^\infty(\Omega)} \leqslant M_{\mathcal{F}}^2$, applying Lemma E.4 yields that $\mathbf{P}\left[A_5(n,\delta)\right] \geqslant 1-\delta$. To bound R_2, we decompose

$$\begin{aligned}
R_2 &\leqslant \frac{|\mathcal{E}_{n,V}(u_{\mathcal{F}}) - \mathcal{E}_V(u_{\mathcal{F}})|}{\mathcal{E}_{n,2}(u_{\mathcal{F}})} + \frac{\mathcal{E}_V(u_{\mathcal{F}}) + \mathcal{E}_P(u_{\mathcal{F}})}{\mathcal{E}_2(u_{\mathcal{F}})\,\mathcal{E}_{n,2}(u_{\mathcal{F}})}\left|\mathcal{E}_2(u_{\mathcal{F}}) - \mathcal{E}_{n,2}(u_{\mathcal{F}})\right| + \frac{|\mathcal{E}_{n,P}(u_{\mathcal{F}}) - \mathcal{E}_P(u_{\mathcal{F}})|}{\mathcal{E}_{n,2}(u_{\mathcal{F}})} \\
&=: R_{21} + R_{22} + R_{23}.
\end{aligned}$$

Recall the function classes $\mathcal{G}_2$, $\mathcal{F}_j$ defined in (5.4) and that we assume $\sup_{g\in\mathcal{G}_2}\|g\|_{L^\infty(\Omega)}\leqslant M_{\mathcal{G}_2}$. Since $\{\psi_j\}_{j=1}^{k-1}$ are normalized orthogonal eigenfunctions and $\|u_{\mathcal{F}}\|_{L^2(\Omega)}\leqslant 3/2$,

$$\mathcal{E}_P\left(u_{\mathcal{F}}\right)=\beta\sum_{j=1}^{k-1}\langle u_{\mathcal{F}},\psi_j\rangle^2\leqslant\beta\|u_{\mathcal{F}}\|_{L^2(\Omega)}^2\leqslant\frac{9}{4}\beta.$$

By (B.13) and $\mathcal{E}_V\left(u^*\right)=\lambda_k$,

$$\begin{aligned}\mathcal{E}_V(u_{\mathcal{F}})&\leqslant\max\left\{1,V_{\max}\right\}\|u-u^*\|_{H^1(\Omega)}^2+2\sqrt{\lambda_k\max\left\{1,V_{\max}\right\}}\|u-u^*\|_{H^1(\Omega)}+\mathcal{E}_V\left(u^*\right)\\&\leqslant\left(\max\left\{1,\sqrt{V_{\max}}\right\}/2+\sqrt{\lambda_k}\right)^2.\end{aligned}$$

Therefore, if $\xi_5(n,\delta)<1/2$, on the event $A_5(n,\delta)$, then

$$R_{22}\leqslant\frac{\left(\left(\max\left\{1,\sqrt{V_{\max}}\right\}+2\sqrt{\lambda_k}\right)^2+9\beta\right)\xi_5(n,\delta)}{1-2\xi_5(n,\delta)}. \tag{E.9}$$

To bound R_{21}, we define the constant $\xi_6(n,\delta):=M_{\mathcal{G}_2}\sqrt{\ln(2/\delta)/(2n)}$ and the event

$$A_6(n,\delta):=\left\{\left|\mathcal{E}_{n,V}(u_{\mathcal{F}})-\mathcal{E}_V(u_{\mathcal{F}})\right|\leqslant\xi_6(n,\delta)\right\}.$$

Using Lemma E.4, we get $\mathbf{P}\left[A_6(n,\delta)\right]\geqslant 1-\delta$. Hence, if $\xi_5(n,\delta)<1/2$, within event $A_5(n,\delta)\cap A_6(n,\delta)$, then

$$R_{21}\leqslant\frac{2\xi_6(n,\delta)}{1-2\xi_5(n,\delta)}. \tag{E.10}$$

To bound R_{23}, recall that $\bar{\mu}_k=\max_{1\leqslant j\leqslant k-1}\|\psi_j\|_{L^\infty(\Omega)}$. We define the constant

$$\xi_7(n,\delta):=\bar{\mu}_kM_{\mathcal{F}}\sqrt{\frac{\ln(2k/\delta)}{2n}},$$

and the events

$$A_{7,j}(n,\delta):=\left\{\left|\frac{1}{n}\sum_{i=1}^n u_{\mathcal{F}}\left(X_i\right)\psi_j\left(X_i\right)-\langle u_{\mathcal{F}},\psi_j\rangle\right|\leqslant\xi_7(n,\delta)\right\}\quad\text{for each }1\leqslant j\leqslant k-1.$$

Let $A_7(n,\delta):=\bigcap_{j=1}^{k-1}A_{7,j}(n,\delta)$. By Lemma E.4, $\mathbf{P}\left[A_{7,j}(n,\delta)\right]\geqslant 1-\delta/k$ and hence $\mathbf{P}\left[A_7(n,\delta)\right]\geqslant 1-\delta$.

On event $A_7(n,\delta)$, it follows from the fact $a^2-b^2=(a-b)^2+2b(a-b)$ that

$$\begin{aligned}\beta^{-1}\left|\mathcal{E}_{n,P}\left(u_{\mathcal{F}}\right)-\mathcal{E}_P\left(u_{\mathcal{F}}\right)\right|&\leqslant\sum_{j=1}^{k-1}\left[\xi_7^2(n,\delta)+2\left|\langle u_{\mathcal{F}},\psi_j\rangle\right|\xi_7(n,\delta)\right]\\&\leqslant 2\left(\sum_{j=1}^{k-1}\langle u_{\mathcal{F}},\psi_j\rangle^2\right)^{1/2}\left(\sum_{j=1}^{k-1}\xi_7^2(n,\delta)\right)^{1/2}+(k-1)\xi_7^2(n,\delta)\\&\leqslant 3\sqrt{k}\xi_7(n,\delta)+k\xi_7^2(n,\delta),\end{aligned}$$

where we have used $\|u_{\mathcal{F}}\|_{L^2(\Omega)}\leqslant 3/2$ in the last inequality. Hence, on the event $A_5(n,\delta)\cap A_7(n,\delta)$,

$$R_{23}\leqslant\frac{2\beta\sqrt{k}\xi_7(n,\delta)\left(3+\sqrt{k}\xi_7(n,\delta)\right)}{1-2\xi_5(n,\delta)}. \tag{E.11}$$

Thus, it may be concluded from (E.9), (E.10) and (E.11) that if $\xi_5(n,\delta)<1/2$, within event $\bigcap_{i=5}^7A_i(n,\delta)$, then

$$R_2\leqslant\frac{\left(\left(\max\left\{1,\sqrt{V_{\max}}\right\}+2\sqrt{\lambda_k}\right)^2+9\beta\right)\xi_5(n,\delta)+2\xi_6(n,\delta)+2\beta\sqrt{k}\xi_7(n,\delta)\left(3+\sqrt{k}\xi_7(n,\delta)\right)}{1-2\xi_5(n,\delta)}.$$

By (7.9), there exist certain absolute constants C such that

$$\xi_5(n,\delta) \leqslant C\left(\frac{B}{d}\right)^2\sqrt{\frac{\ln(1/\delta)}{n}}, \quad \xi_6(n,\delta) \leqslant CB^2\left(1+V_{\max}\right)\sqrt{\frac{\ln(1/\delta)}{n}},$$
$$\xi_7(n,\delta) \leqslant \frac{C\bar{\mu}_k B}{d}\sqrt{\frac{\ln(k/\delta)}{n}},$$

and if $\xi_5(n,\delta) < 1/4$, then

$$R_2 \leqslant C\left(V_{\max}+\lambda_k+\beta\right)B^2\sqrt{\frac{\ln(1/\delta)}{n}} + C\beta\left(\bar{\mu}_k B/d\right)\sqrt{\frac{k\ln(k/\delta)}{n}}. \tag{E.12}$$

Note that the bound for ξ_5 is smaller than that for R_1 and R_2 up to an absolute constant. The bound for R_5 is smaller than that for R_3 up to an absolute constant. By the estimates (E.5), (E.6), (E.7), (E.8) and (E.12) with the choice $\gamma \geqslant 4\lambda_k$, (3.6) ensures that $\xi_5(n,\delta) < 1/4$ and $|R_1|$, R_2/γ, R_3/γ, $(R_4+R_5)^2$ are all bounded by $1/16$ on event $\bigcap_{i=4}^7 A_i(n,\delta)$. Then, for $0<\delta<1/4$, it follows from (E.2), (E.3) and $\mathbf{P}\left[A_i(n,\delta)\right] \geqslant 1-\delta$ for $i=5,6,7$ that

$$\mathbf{P}\left(\mathcal{E}_2(u_n) \geqslant 1/2\right) \geqslant \mathbf{P}\left(\bigcap_{i=4}^7 A_i(n,\delta)\right) \geqslant 1-4\delta, \tag{E.13}$$

which completes the proof of Theorem 3.9.

Next, we analyze the generalization error of the penalty method. To this end, we decompose

$$\begin{aligned} L_k(u_n)-\lambda_k &\leqslant \left[\mathscr{L}_k(u_n)-\mathscr{L}_{k,n}(u_n)\right]+\mathscr{L}_{k,n}(u_n)-\lambda_k \\ &=: R_6+R_2+R_3+\gamma\left(R_4+R_5\right)^2, \end{aligned} \tag{E.14}$$

where we have used (E.1). Further, we decompose R_6 as follows

$$\begin{aligned} R_6 &= L_k(u_n)-L_{k,n}(u_n)+\gamma\left(\mathcal{E}_2(u_n)-1\right)^2-\gamma\left(\mathcal{E}_{n,2}(u_n)-1\right)^2 \\ &\leqslant L_k(u_n)-L_{k,n}(u_n)+\gamma\left[\left(\mathcal{E}_2(u_n)-\mathcal{E}_{n,2}(u_n)\right)^2+2\left|\mathcal{E}_2(u_n)-1\right|\left|\mathcal{E}_2(u_n)-\mathcal{E}_{n,2}(u_n)\right|\right] \\ &=: R_{61}+\gamma\left[R_1^2+2\left|\mathcal{E}_2(u_n)-1\right|\left|R_1\right|\right]. \end{aligned} \tag{E.15}$$

Notice that under the assumptions of Theorem 3.9, within event $\bigcap_{i=4}^7 A_i(n,\delta)$, $\mathcal{E}_2(u_n) \geqslant 1/2$. Thus, the analysis in §5 is applicable to u_n and $u_{\mathcal{F}}$. Proceeding along the same line in §5 that leads to (5.15), we have, if $\xi_1(n,r,\delta)<1$, within event $\bigcap_{i=1}^3 A_i(n,r,\delta)$, then

$$R_{61} \leqslant \frac{\xi_1+\xi_2}{1-\xi_1}\left(L_k(u_n)-\lambda_k\right)+\lambda_k\frac{\xi_1+\xi_2}{1-\xi_1}+\frac{\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right). \tag{E.16}$$

Notice that R_2 coincides with T_2. Thus, by (5.17),

$$R_2 \leqslant \frac{\xi_1+\xi_2}{1-\xi_1}\left(L_k(u_{\mathcal{F}})-\lambda_k\right)+\lambda_k\frac{\xi_1+\xi_2}{1-\xi_1}+\frac{\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right).$$

Combining the above inequality with (E.14), (E.15) and (E.16), we obtain, if $2\xi_1+\xi_2 \leqslant 1/2$, then

$$\begin{aligned} L_k(u_n)-\lambda_k \leqslant{}& \frac{\xi_1+\xi_2}{1-\xi_1}\left(L_k(u_n)-\lambda_k\right)+2\lambda_k\frac{\xi_1+\xi_2}{1-\xi_1}+\frac{2\beta}{1-\xi_1}\left(\frac{k}{4}\xi_3^2+\sqrt{k}\xi_3\right) \\ &+\frac{\xi_1+\xi_2}{1-\xi_1}\left(L_k(u_{\mathcal{F}})-\lambda_k\right)+\gamma\left[\left(R_4+R_5\right)^2+R_1^2+|R_1|\right]+R_3, \end{aligned}$$

and so

$$\begin{aligned} L_k(u_n)-\lambda_k \leqslant{}& 4\lambda_k\left(\xi_1+\xi_2\right)+\beta\left(k\xi_3^2+4\sqrt{k}\xi_3\right)+2\left(L_k(u_{\mathcal{F}})-\lambda_k\right) \\ &+2\gamma\left(2R_4^2+2R_5^2+R_1^2+|R_1|\right)+2R_3. \end{aligned} \tag{E.17}$$

Note that under the assumptions of Theorem 3.9, $\xi_5(n,\delta)$, $|R_1|$, R_4+R_5 are all bounded by 1/4. When (3.3a) and (3.3b) with $r=0.49$ hold, $2\xi_1+\xi_2\leqslant 1/2$. Here, without loss of generality, we may take $r=0.49$ and $\delta\in(0,1/7)$. Substituting the bounds (E.5), (E.6), (E.8), (E.7) derived for R_1, R_3, R_4, R_5 earlier in this section, and the bounds (8.3) for $\{\xi_i\}_{i=1}^3$ into (E.17), we obtain (3.7) on the event $A_{in}=\left(\bigcap_{i=1}^3 A_i(n,r,\delta)\right)\bigcap\left(\bigcap_{i=4}^7 A_i(n,\delta)\right)$ with $\mathbf{P}\left(A_{in}\right)\geqslant 1-7\delta$ followed from (E.13) and (5.16), which completes the proof of Corollary 3.11.

Appendix F Analysis of the accumulative error

In this section, we prove Theorem 3.13 and Proposition 3.14. Recall the loss $L_k(u)$ defined in (2.4) and we take ψ_j to be the normalization of the orthogonal projection of $\mathfrak{u}_{\theta j}$ to subspace U_j, i.e., $\psi_j=P_j\mathfrak{u}_{\theta j}/\|P_j\mathfrak{u}_{\theta j}\|_{L^2(\Omega)}$. Since $\mathfrak{u}_{\theta k}$ is a minimizer of $\widetilde{L}_{k,n}(u)$ over $\mathcal{F}_{>r}$, $\widetilde{L}_{k,n}(\mathfrak{u}_n)-\widetilde{L}_{k,n}(u_{\mathcal{F}})\leqslant 0$ for any $u_{\mathcal{F}}\in\mathcal{F}_{>r}$. Similar to (5.2), for any $u_{\mathcal{F}}\in\mathcal{F}_{>r}$, we decompose the generalization error as

$$\begin{aligned}L_k(\mathfrak{u}_{\theta k})-\lambda_k&\leqslant\left(L_k(\mathfrak{u}_{\theta k})-\widetilde{L}_k(\mathfrak{u}_{\theta k})\right)+\left(\widetilde{L}_k(\mathfrak{u}_{\theta k})-\widetilde{L}_{k,n}(\mathfrak{u}_{\theta k})\right)+\left(\widetilde{L}_{k,n}(u_{\mathcal{F}})-\widetilde{L}_k(u_{\mathcal{F}})\right)\\&\quad+\left(\widetilde{L}_k(u_{\mathcal{F}})-L_k(u_{\mathcal{F}})\right)+\left(L_k(u_{\mathcal{F}})-\lambda_k\right)\\&=:S_1+S_2+S_3+S_4+S_5,\end{aligned}\tag{F.1}$$

Note that S_1, S_4 are the accumulative errors, S_2 is the statistical error, S_3 is the Monte Carlo error and S_5 is the approximation error.

Firstly, we give a uniform bound on the accumulative error $\widetilde{L}_k(u)-L_k(u)$.

Proposition F.1. *Assume that for each $j\in\mathbb{N}_+$, ψ_j is the normalization of the orthogonal projection of $u_{\theta j}$ to subspace U_j. Set $\beta=\beta_k$ in $L_k(u)$. Then, for any $u\in L^2(\Omega)$, there holds*

$$\left|\widetilde{L}_k(u)-L_k(u)\right|\leqslant 2\beta_k\sum_{j=1}^{k-1}\sqrt{\frac{L_j(u_{\theta j})-\lambda_j}{\min\left\{\beta_j+\lambda_1-\lambda_j,\lambda_{j'}-\lambda_j\right\}}}.$$

Proof. For any $u\in L^2(\Omega)$, we let $\bar{u}=u/\|u\|_{L^2(\Omega)}$ and $\bar{u}_{\theta j}=u_{\theta j}/\|u_{\theta j}\|_{L^2(\Omega)}$. Using the triangle inequality and the Cauchy's inequality, we get

$$\begin{aligned}\frac{1}{\beta_k}\left|\widetilde{L}_k(u)-L_k(u)\right|&\leqslant\sum_{j=1}^{k-1}|\langle\bar{u},\bar{u}_{\theta j}+\psi_j\rangle|\,|\langle\bar{u},\bar{u}_{\theta j}-\psi_j\rangle|\\&\leqslant\sum_{j=1}^{k-1}\|\bar{u}\|_{L^2}^2\|\bar{u}_{\theta j}+\psi_j\|_{L^2}\|\bar{u}_{\theta j}-\psi_j\|_{L^2}.\end{aligned}\tag{F.2}$$

Since $\bar{u}_{\theta j}$ and ψ_j are both normalized,

$$\|\bar{u}_{\theta j}+\psi_j\|_{L^2}\|\bar{u}_{\theta j}-\psi_j\|_{L^2}=2\sqrt{1-\langle\bar{u}_{\theta j},\psi_j\rangle^2}.\tag{F.3}$$

By the choice of ψ_j and the stability estimates in Proposition 2.3,

$$\sqrt{1-\langle\bar{u}_{\theta j},\psi_j\rangle^2}=\frac{\|P_j^{\perp}u_{\theta j}\|_{L^2}}{\|u_{\theta j}\|_{L^2}}\leqslant\sqrt{\frac{L_j(u_{\theta j})-\lambda_j}{\min\left\{\beta_j+\lambda_1-\lambda_j,\lambda_{j'}-\lambda_j\right\}}}.\tag{F.4}$$

The proof is completed by combining (F.2), (F.3) and (F.4). □

Recall the constants $\{\xi_i(n,r,\delta)\}_{i=1}^3$ and the events $\{A_i(n,r,\delta)\}_{i=1}^3$ defined in §5. We bound S_2 as

$$\begin{aligned}S_2&\leqslant\left|\frac{\mathcal{E}_{n,V}(\mathfrak{u}_{\theta k})}{\mathcal{E}_{n,2}(\mathfrak{u}_{\theta k})}-\frac{\mathcal{E}_V(\mathfrak{u}_{\theta k})}{\mathcal{E}_2(\mathfrak{u}_{\theta k})}\right|+\beta_k\sum_{j=1}^{k-1}\left|\frac{\mathscr{P}_{n,j}(\mathfrak{u}_{\theta k})}{\mathcal{E}_{n,2}(\mathfrak{u}_{\theta k})\mathcal{E}_{n,2}(u_{\theta j})}-\frac{\mathscr{P}_j(\mathfrak{u}_{\theta k})}{\mathcal{E}_2(\mathfrak{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}\right|\\&=:S_{21}+S_{22}.\end{aligned}$$

Proceeding along the same line that leads to (5.10), we obtain, if $\xi_1(n,r,\delta)<1$, then on the event $\bigcap_{i=1}^2 A_i(n,r,\delta)$, there holds

$$S_{21}\leqslant \frac{\xi_1(n,r,\delta)+\xi_2(n,r,\delta)}{1-\xi_1(n,r,\delta)}\cdot\frac{\mathcal{E}_V\left(\mathrm{u}_{\theta k}\right)}{\mathcal{E}_2\left(\mathrm{u}_{\theta k}\right)}. \tag{F.5}$$

For S_{22}, we get

$$\frac{S_{22}}{\beta_k}\leqslant \frac{|\mathscr{P}_{n,j}(\mathrm{u}_{\theta k})-\mathscr{P}_j(\mathrm{u}_{\theta k})|}{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}\frac{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}{\mathcal{E}_{n,2}(\mathrm{u}_{\theta k})\mathcal{E}_{n,2}(u_{\theta j})}+\frac{\mathscr{P}_j(\mathrm{u}_{\theta k})}{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}\left|\frac{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}{\mathcal{E}_{n,2}(\mathrm{u}_{\theta k})\mathcal{E}_{n,2}(u_{\theta j})}-1\right|.$$

To bound the statistical error, we use the normalization property and $L^\infty(\Omega)$ boundedness of ψ_j. We may define $\xi_{\theta 3}(n,r,\delta)$, $A_{\theta 3}(n,r,\delta)$ as $\xi_3(n,r,\delta)$, $A_3(n,r,\delta)$ by replacing ψ_j with $\bar{u}_{\theta j}$, and get bounds for them by replacing $\|\psi_j\|_{L^\infty(\Omega)}$ with $\|\bar{u}_{\theta j}\|_{L^\infty(\Omega)}$. Therefore, similar to (5.13), we obtain

$$\sum_{j=1}^{k-1}\frac{|\mathscr{P}_{n,j}(\mathrm{u}_{\theta k})-\mathscr{P}_j(\mathrm{u}_{\theta k})|}{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}\leqslant \frac{k}{4}\xi_{\theta 3}(n,r,\delta)^2+\sqrt{k}\xi_{\theta 3}(n,r,\delta).$$

If $\xi_1(n,r,\delta)<1$, on the event $A_1(n,r,\delta)$, then

$$(1+\xi_1)^{-2}<\frac{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}{\mathcal{E}_{n,2}(\mathrm{u}_{\theta k})\mathcal{E}_{n,2}(u_{\theta j})}<(1-\xi_1)^{-2}.$$

Hence,

$$\left|\frac{\mathcal{E}_2(u)\mathcal{E}_2(u_{\theta j})}{\mathcal{E}_{n,2}(u)\mathcal{E}_{n,2}(u_{\theta j})}-1\right|<\frac{\xi_1\left(2-\xi_1\right)}{\left(1-\xi_1\right)^2}.$$

Thus, on event $A_1(n,r,\delta)\cap A_{\theta 3}(n,r,\delta)$, if $\xi_1(n,r,\delta)<1$, then

$$S_{22}\leqslant \frac{\beta_k}{\left(1-\xi_1\right)^2}\left(\frac{k}{4}\xi_{\theta 3}^2+\sqrt{k}\xi_{\theta 3}\right)+\frac{\xi_1\left(2-\xi_1\right)}{\left(1-\xi_1\right)^2}\sum_{j=1}^{k-1}\frac{\beta_k\mathscr{P}_j(\mathrm{u}_{\theta k})}{\mathcal{E}_2(\mathrm{u}_{\theta k})\mathcal{E}_2(u_{\theta j})}. \tag{F.6}$$

We conclude from (F.5) and (F.6) that on event $A_1(n,r,\delta)\cap A_2(n,r,\delta)\cap A_{\theta 3}(n,r,\delta)$, if $\xi_1(n,r,\delta)<1$, then

$$S_2\leqslant \frac{2\xi_1+\xi_2}{\left(1-\xi_1\right)^2}L_k(\mathrm{u}_{\theta k})+\frac{\beta_k}{\left(1-\xi_1\right)^2}\left(\frac{k}{4}\xi_{\theta 3}^2+\sqrt{k}\xi_{\theta 3}\right).$$

Similarly, since $\|u_{\mathcal{F}}\|_{L^2(\Omega)}>r$, we obtain

$$S_3\leqslant \frac{2\xi_1+\xi_2}{\left(1-\xi_1\right)^2}L_k(u_{\mathcal{F}})+\frac{\beta_k}{\left(1-\xi_1\right)^2}\left(\frac{k}{4}\xi_{\theta 3}^2+\sqrt{k}\xi_{\theta 3}\right).$$

Then, it follows from the above two estimates and (F.1) that

$$\begin{aligned}L_k(\mathrm{u}_{\theta k})-\lambda_k\leqslant{}& \frac{2\xi_1+\xi_2}{\left(1-\xi_1\right)^2}\left(L_k(\mathrm{u}_{\theta k})-\lambda_k\right)+\frac{2\xi_1+\xi_2}{\left(1-\xi_1\right)^2}\left(L_k(u_{\mathcal{F}})-\lambda_k\right)+2\lambda_k\frac{2\xi_1+\xi_2}{\left(1-\xi_1\right)^2}\\&+\frac{2\beta_k}{\left(1-\xi_1\right)^2}\left(\frac{k}{4}\xi_{\theta 3}^2+\sqrt{k}\xi_{\theta 3}\right)+S_1+S_4+S_5,\end{aligned}$$

and if $4\xi_1+\xi_2\leqslant 1/2$, then

$$L_k(\mathrm{u}_{\theta k})-\lambda_k\leqslant 4\lambda_k\left(2\xi_1+\xi_2\right)+\beta_k\left(k\xi_{\theta 3}^2+4\sqrt{k}\xi_{\theta 3}\right)+2S_1+2S_4+3S_5. \tag{F.7}$$

Let $u_{\mathcal{F}}$ be given by Theorem 3.2. Substituting the bounds (8.3) for $\{\xi_i\}_{i=1}^3$, the bound for S_1, S_4 given by Proposition F.1 and the bound for S_5 given by Theorem 6.7 into (F.7), we prove Theorem 3.13.

Proof of Proposition 3.14. Let $d_0 = 0$ and $d_k = \sum_{j=1}^{k} \sqrt{\left(L_j\left(u_{\theta j}\right) - \lambda_j\right)/\beta_j}$. Since

$$\frac{L_k\left(u_{\theta k}\right) - \lambda_k}{\beta_k} \leqslant \frac{\Delta_k}{\beta_k} + 8\sum_{j=1}^{k-1} \sqrt{\frac{\beta_j}{\min\left(\beta_j + \lambda_1 - \lambda_j, \lambda_{j'} - \lambda_j\right)} \cdot \frac{L_j\left(u_{\theta j}\right) - \lambda_j}{\beta_j}},$$

we have

$$\left(d_k - d_{k-1}\right)^2 = \frac{L_k\left(u_{\theta k}\right) - \lambda_k}{\beta_k} \leqslant \tau_k + 2\rho_{k-1} d_{k-1} \tag{F.8}$$

and $d_k \leqslant d_{k-1} + \sqrt{\tau_k + 2\rho_{k-1}d_{k-1}}$. We claim that $d_k \leqslant \rho_{k-1}k^2/2 + \sqrt{\tau_k}k$ for all $k \in \mathbb{N}$, which may be proved by induction. When $k = 0$, the bound is trivial. Assume that the claim holds for k. For $k+1$, since $\rho_{k-1} \leqslant \rho_k$ and $\tau_k \leqslant \tau_{k+1}$,

$$\begin{aligned} d_{k+1} &\leqslant \frac{\rho_{k-1}}{2}k^2 + \sqrt{\tau_k}k + \sqrt{\tau_{k+1} + 2\rho_k\left(\rho_{k-1}k^2/2 + \sqrt{\tau_k}k\right)} \\ &= \frac{\rho_k}{2}k^2 + \sqrt{\tau_k}k + \left(\rho_k k + \sqrt{\tau_{k+1}}\right) \\ &\leqslant \frac{\rho_k}{2}(k+1)^2 + \sqrt{\tau_{k+1}}(k+1), \end{aligned}$$

which completes the proof of the claim. The estimate (3.9) follows from the claim and (F.8). □

Appendix G Properties of eigenvalues and eigenfunctions

In this part, we characterize the asymptotic distribution of the eigenvalues and estimate the maximum norm of the eigenfunctions.

Lemma G.1. *If $\mathcal{H}$ satisfies Assumption 2.1, then, there exists an absolute constant C such that*

$$\frac{1}{C}dk^{2/d} + V_{\min} \leqslant \lambda_k \leqslant Cdk^{2/d} + V_{\max}.$$

Proof. First, when $V \equiv 0$, we consider the eigenvalue problem of the Laplacian. By Weyl's formula[68, 69], the k-th eigenvalue ν_k satisfies

$$c_1 C(d)^{2/d} k^{2/d} \leqslant \nu_k \leqslant c_2 C(d)^{2/d} k^{2/d},$$

where c_1, c_2 are absolute constants and $C(d) = d2^{d-1}\pi^{d/2}\Gamma\left(d/2\right)$. A straightforward calculation gives $C(d)^{2/d} = 4\pi\Gamma(d/2+1)^{2/d}$. Using the Stirling's approximation formula [4, eq. 3.9],

$$\sqrt{2\pi x}\left(\frac{x}{\mathrm{e}}\right)^x \leqslant \Gamma(x+1) \leqslant \sqrt{2\pi x}\left(\frac{x}{\mathrm{e}}\right)^x \mathrm{e}^{1/(12x)}, \qquad x > 0,$$

we obtain

$$\frac{2\pi}{\mathrm{e}}(\pi d)^{1/d} \leqslant \frac{C(d)^{2/d}}{d} \leqslant \frac{2\pi}{\mathrm{e}}(\pi d)^{1/d}\mathrm{e}^{1/(3d^2)}.$$

Using the elementary facts

$$(\pi d)^{1/d} \geqslant 1 \qquad \text{and} \qquad (\pi d)^{1/d}\mathrm{e}^{1/(3d^2)} \leqslant \pi \mathrm{e}^{1/3},$$

we get

$$\frac{2\pi}{\mathrm{e}}d \leqslant C(d)^{2/d} \leqslant 4\pi d.$$

Therefore, there exists absolute constant C such that

$$\frac{1}{C}dk^{2/d} \leqslant \nu_k \leqslant Cdk^{2/d}.$$

Second, by the minimax principle, we get

$$\lambda_k = \min_{\dim E = k} \max_{u \in E\setminus\{0\}} \frac{\int_\Omega \left(|\nabla u|^2 + Vu^2\right) \mathrm{d}\,x}{\int_\Omega u^2 \,\mathrm{d}\,x},$$

where the minimum is taken over all k-dimensional subspace $E \subset H_0^1(\Omega)$. Using $V_{\min} \leqslant V(x) \leqslant V_{\max}$, we conclude

$$\nu_k + V_{\min} \leqslant \lambda_k \leqslant \nu_k + V_{\max},$$

which together with the upper and the lower bounds for ν_k completes the proof. □

Lemma G.2. *If $\mathcal{H}$ satisfies Assumption 2.1, then*

$$\|\psi_k\|_{L^\infty(\Omega)} \leqslant \left(Ck^{2/d} + \frac{\mathrm{e}\,(\mathrm{V_{max}} - \mathrm{V_{min}})}{\pi d}\right)^{d/4}, \tag{G.1}$$

where C is an absolute constant.

Proof. Combining Example 2.1.9 and Lemma 2.1.2 in [16], for any Schrödinger operator $\mathcal{H}_1 = -\Delta + V_1$ where $0 \leqslant V_1 \in L^1_{\mathrm{loc}}(\Omega)$, the kernel $K(t,x,y)$ of $\mathrm{e}^{-\mathcal{H}_1 t}$ satisfies

$$0 \leqslant K(t,x,y) \leqslant (4\pi t)^{-d/2} \mathrm{e}^{-(x-y)^2/4t} \leqslant (4\pi t)^{-d/2},$$

which implies that $\mathrm{e}^{-\mathcal{H}_1 t}$ is a symmetric Markov semigroup on $L^2(\Omega)$ and $\mathrm{e}^{-\mathcal{H}_1 t} : L^2(\Omega) \to L^\infty(\Omega)$ is a bounded operator with norm $\|\mathrm{e}^{-\mathcal{H}_1 t}\|_{\infty,2} \leqslant (4\pi t)^{-d/4}$ for all $0 < t < \infty$. Suppose that ϕ is a normalized eigenfunction of $\mathcal{H}_1$ associated with the eigenvalue λ. Hence,

$$\|\phi\|_{L^\infty(\Omega)} = \mathrm{e}^{\lambda t} \|\mathrm{e}^{-\mathcal{H}_1 t}\phi\|_{L^\infty(\Omega)} \leqslant (4\pi t)^{-d/4} \mathrm{e}^{\lambda t}. \tag{G.2}$$

Let $V_1 = V - V_{\min}$. By Assumption 2.1, $0 \leqslant V_1 \leqslant V_{\max} - V_{\min}$ and ψ_k is the k-th normalized eigenfunction of $\mathcal{H}_1$ associated with the eigenvalue $\lambda_k - V_{\min}$. Taking $t = d/[4\lambda], \phi = \psi_k$ and $\lambda = \lambda_k - V_{\min}$ in (G.2), we have

$$\|\psi_k\|_{L^\infty(\Omega)} \leqslant \left(\frac{\mathrm{e}\,(\mathrm{\lambda_k} - \mathrm{V_{min}})}{\pi d}\right)^{d/4} \quad \text{for all } k \geqslant 1.$$

Substituting the upper bound for λ_k in Lemma G.1 completes the proof. □