COMMUNICATIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 2, Pages 1–21 (January 31, 2022)
https://doi.org/10.1090/cams/5

A PRIORI GENERALIZATION ERROR ANALYSIS OF TWO-LAYER NEURAL NETWORKS FOR SOLVING HIGH DIMENSIONAL SCHRÖDINGER EIGENVALUE PROBLEMS

JIANFENG LU AND YULONG LU

ABSTRACT. This paper analyzes the generalization error of two-layer neural networks for computing the ground state of the Schrödinger operator on a d-dimensional hypercube with Neumann boundary condition. We prove that the convergence rate of the generalization error is independent of dimension d, under the a priori assumption that the ground state lies in a spectral Barron space. We verify such assumption by proving a new regularity estimate for the ground state in the spectral Barron space. The latter is achieved by a fixed point argument based on the Krein-Rutman theorem.

1. INTRODUCTION

High dimensional partial differential equations (PDEs) arise ubiquitously from scientific and engineering problems which involve many degrees of freedom, examples include many-body quantum mechanics, phase space description of chemical dynamics, learning and control of complex systems, spectral methods for high dimensional data, just to name a few. While numerical methods for partial differential equations in low-dimension are quite standard, the numerical solution to high dimensional PDEs has remained an outstanding challenge due to the well-known curse of dimensionality. Namely, the computational cost can grow exponentially as the dimension increases. Perhaps the most celebrated and important example of such challenge is to determine the ground state of many-body quantum systems, which amounts to solving eigenvalue problems for high dimensional PDEs.

In recent years, neural networks have shown great success in representing high-dimensional classifiers or probability distributions in a variety of machine learning tasks and have led to the tremendous success and popularity of deep learning [32, 38]. Motivated by those recent success, researchers have been actively exploring using deep learning techniques to solve high dimensional PDEs [10, 18, 22, 23, 29, 37, 45, 48] by using neural networks to parameterize the unknown solution of high dimensional PDEs. Thanks to the flexibility of the neural network approximations, such methods have achieved remarkable results for various kind of PDE problems, including eigenvalue problems for many-body quantum systems (see e.g., [7, 9, 12, 19, 24–26, 36]), where the high dimensional wave functions are parameterized by neural networks with specific

Received by the editors May 18, 2021, and, in revised form, September 30, 2021, and December 23, 2021.
2020 *Mathematics Subject Classification.* Primary 35P15, 65N25, 68T07, 35Q40.
The first author was supported in part by National Science Foundation via grants DMS-2012286 and CCF-1934964.
The second author was supported by the National Science Foundation through the award DMS-2107934.
The second author is the corresponding author.

architecture design to address the symmetry properties of many-body quantum systems.

Despite wide popularity and many successful examples of employing neural network ansatz for solving PDEs, their theoretical analyses are still sparse. In [40,41], the authors obtained the convergence error estimates for PINNs based on both strong and variational formulations in the context of solving linear elliptic and parabolic PDEs. In [35], the authors proposed a general framework to study the a-posterior-type generalization error estimates for PINNs. It is worth to note that in the aforementioned work, the error estimates were proved under the assumption that the solution belongs to Sobolev or Hölder spaces and hence those estimates suffer from the curse of dimensionality. In [27, 34, 47], the authors established dimension-explicit a priori-estimates for the generalization error of two-layer neural networks for solving elliptic PDEs assuming (but without verifying) the solution of the PDEs lie in certain Barron spaces. In our recent work [33], we proved a dimension-independent convergence rate for the generalization error bound for the deep Ritz method [18] for solving elliptic PDEs when the solutions lie in some spectral Barron space, and more importantly we also established new regularity theory in the spectral Barron space for the PDEs.

Nonetheless, to the best of our knowledge, the numerical analysis of neural network methods for high dimensional eigenvalue problems is not yet established. The goal of this paper is to provide an *a priori* generalization analysis of variational methods for computing the ground state of the Schrödinger operator in high dimension based on the two-layer neural network ansatz.

Our generalization error analysis follows largely the framework established in our previous work [33], where a priori generalization error is analyzed for deep Ritz method for solving elliptic equations. In particular, to establish approximation results that do not deteriorate as dimension increases, we will work in a spectral Barron space that is firstly defined in the seminal work of Barron [5]. It has been shown in [5, 14, 30, 33, 43] that spectral Barron functions has "lower complexity" than more familiar regularity-based functions such as Sobolev or Hölder functions in the sense that the former can be efficiently approximated by two-layer neural networks without curse of dimensionality. It is also worth mentioning that another notion of Barron space based on an integral representation is defined in [15], in which similar neural network approximation result holds. Discussions on the relationship between the two notions of Barron spaces and their properties can be found in [8] and [16].

On the other hand, as the Barron space is rather different from Sobolev or Hölder spaces, the main challenge one faces is to establish regularity theory for high dimensional PDEs in such space. Our previous work [33] established the appropriate solution theory for elliptic equations. The key contribution of the present work is to extend such novel solution theory to Schrödinger eigenvalue problems in high dimension. Since we are working in spectral Barron space, which is a general Banach space without inner product structure, Lax-Milgram or Courant-Fisher theorems are not applicable, and thus we have to rely on fixed point theorem to establish the existence of solutions. In our previous work for elliptic PDEs [33], the Fredholm alternative principle was used; while in this work, to establish the existence of nontrivial eigenfunctions, we rely on the Krein-Rutman theorem [31]. We also remark that similar regularity estimates in an

alternative Barron space [15] were obtained for some nonlinear PDEs [17] and linear elliptic PDEs [11].

The remainder of this paper is organized as follows. In Section 2 we first set up the ground state problem of the Schrödinger operator and present the main generalization results (see Theorems 2.3 and 2.4) and the new regularity estimate on the ground state in the spectral Barron space (see Theorem 2.5). In Section 3 we prove a key stability estimate on the ground state, which allows us to bound the H^1-error between the ground state and its approximation in terms of the energy excess. We present the proof of the main generalization result in Section 4 and the proof of the new regularity estimate on the ground state in Section 5.

2. Set-up and main results

2.1. **Set-up of problem.** Let $\Omega = [0,1]^d$ be the unit hypercube on $\mathbb{R}^d$ with the boundary $\partial\Omega$. Consider the Neumann eigenvalue problem for the Schrödinger operator

$$\begin{aligned} \mathcal{H}u = -\Delta u + Vu &= \lambda u \qquad \text{in } \Omega, \\ \frac{\partial u}{\partial \nu} &= 0 \qquad\quad \text{on } \partial\Omega, \end{aligned}$$

where $\mathcal{H} := -\Delta + V$ is the Schrödinger operator with the potential function V equipped with the Neumann boundary condition. We are particularly interested in computing the ground state of $\mathcal{H}$, that is the eigenfunction associated to the smallest eigenvalue of $\mathcal{H}$.

Throughout the paper we make the following minimum assumption on the potential function.

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

Note that we assume without loss of generality that $V_{\min}$ is positive, since one can always add a constant to V without changing the eigenfunctions.

It is well-known that the minimum eigenvalue λ_0 can be characterized as the minimum of Rayleigh quotient, i.e.

$$\lambda_0 = \min_{u \in H^1(\Omega)} \mathcal{R}(u) := \min_{u \in H^1(\Omega)} \frac{\langle u, \mathcal{H}u \rangle_{H^1 \times (H^1)^*}}{\langle u, u \rangle} = \min_{u \in H^1(\Omega)} \frac{\int_\Omega |\nabla u|^2 + Vu^2 dx}{\int_\Omega u^2 dx}, \tag{2.1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2(\Omega)$ and $\langle \cdot, \cdot \rangle_{H^1 \times (H^1)^*}$ denotes the dual product on $H^1 \times (H^1)^*$. Under Assumption 2.1, the minimizer of the variational problem (2.1) is achieved at the ground state $u_0 \in H^1(\Omega)$. Moreover, the ground state u_0 is unique (up to a multiplicative constant) and is strictly positive (up to a global sign) on Ω; see e.g., [20, Theorem 3.3.2]. Without loss of generality we assume further that the ground state u_0 is normalized, i.e., $\|u_0\|_{L^2(\Omega)} = 1$.

For certain results to hold, we may also need to make the following additional spectrum assumption on the Schrödinger operator $\mathcal{H}$.

Assumption 2.2. The operator $\mathcal{H}$ has discrete spectrum $\{\lambda_j\}_{j=0}^\infty$ with a positive spectral gap, i.e. $\lambda_0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_k \uparrow \infty$.

Notice that in our setting where the Schrödinger equation is defined on the compact domain, $\mathcal{H}$ always has a discrete spectrum thanks to the well-known Hilbert-Schmidt

theorem and the fact that $(-\Delta+V)^{-1}$ is a compact self-adjoint operator on L^2 (assuming the validity of Assumption 2.1). Moreover, the eigenvector of $\mathcal{H}$ forms an orthonormal basis on L^2. As for the spectral gap, it is well-known that the Schrödinger operator with Dirichlet boundary condition has a positive spectral gap under certain convexity assumption on the domain and the potential; see e.g. [2, 44]. However, less is known about the spectral gap for the Neumann Schrödinger operator; some limited results were obtained by [3, 28] in the one-dimensional setting when the potential is a single well.

To avoid any confusion with the subscript in the notation, let us denote the ground state eigenpair by $(\lambda^*, u^*) \coloneqq (\lambda_0, u_0)$, which our study focuses on.

The natural idea is to seek an approximate solution to Problem (2.1) within some hypothesis class $\mathcal{F} \subset H^1(\Omega)$ that is parameterized by neural networks. In practice, the Monte Carlo method is employed to compute the high dimensional integrals defined by the inner products in (2.1), leading to the definition of empirical loss (or risk) minimization. More concretely, let us denote by $\mathcal{P}_\Omega$ the uniform probability distribution on the domain Ω. Then the population loss $\mathcal{R}$ can be written as

$$\mathcal{R}(u) = \frac{\mathcal{E}_V(u)}{\mathcal{E}_2(u)} \coloneqq \frac{\mathbf{E}_{X\sim\mathcal{P}_\Omega}\Big[|\nabla u(X)|^2 + V(X)|u(X)|^2\Big]}{\mathbf{E}_{X\sim\mathcal{P}_\Omega}\Big[|u(X)|^2\Big]}. \tag{2.2}$$

Let $\{X_j\}_{j=1}^n$ be a sequence of random variables that are independent and identically distributed (i.i.d.) according to $\mathcal{P}_\Omega$. The population loss is approximated by the following empirical loss

$$\mathcal{R}_n(u) = \frac{\mathcal{E}_{n,V}(u)}{\mathcal{E}_{n,2}(u)}, \tag{2.3}$$

where $\mathcal{E}_{n,V}$ and $\mathcal{E}_{n,2}$ are defined by

$$\mathcal{E}_{n,V} \coloneqq \frac{1}{n}\sum_{j=1}^n \big(|\nabla u(X_j)|^2 + V(X_j)|u(X_j)|^2\big)$$

$$\mathcal{E}_{n,2} \coloneqq \frac{1}{n}\sum_{j=1}^n |u(X_j)|^2.$$

Note that we have used the fact that $|\Omega| = 1$ in deriving the Monte Carlo approximation above. Let u_n be a minimizer of $\mathcal{R}_n$ within $\mathcal{F}$, i.e., $u_n = \arg\min_{u\in\mathcal{F}} \mathcal{R}_n(u)$. Again since $\mathcal{R}_n(u)$ is scaling-invariant, we may assume that $\|u_n\|_{L^2} = 1$. Our goal is to obtain quantitative estimates for the error between u_n and u^*, and following the statistical learning literature we will call such error the generalization error.

We are interested in quantifying the error between u_n and u^* in terms of two criteria. The first one is given by the energy excess $\mathcal{R}(u_n) - \mathcal{R}(u^*)$ that quantifies the approximation of $\mathcal{R}(u_n)$ to the leading eigenvalue $\lambda^* = \mathcal{R}(u^*)$.

To introduce the second quantity for measuring the error, we define the projection operator P onto the space of ground state by setting

$$Pu = \langle u, u^*\rangle u^*.$$

Let us also define the operator $P^\perp = (I - P)$, i.e.

$$P^\perp u = u - \langle u, u^*\rangle u^*.$$

Notice that if both u_0 and u are normalized, then

$$\|P^\perp u\|_{L^2}^2 = 1 - |\langle u, u^*\rangle|^2.$$

Therefore $\|P^\perp u_n\|_{L^2(\Omega)}$ quantifies the offset of the direction of u_n from that of u^*. Proposition 2.1 shows that the H^1-norm of $P^\perp u$ can be bounded above by the energy excess.

Proposition 2.1. *Assume that $\mathcal{H}$ satisfies Assumptions* 2.1 *and* 2.2. *Then for any $u \in H^1(\Omega)$,*

$$\|P^\perp u\|_{L^2(\Omega)}^2 \le \frac{\mathcal{R}(u) - \mathcal{R}(u^*)}{\lambda_1 - \lambda^*}\|u\|_{L^2(\Omega)}^2,$$
$$\|\nabla P^\perp u\|_{L^2(\Omega)}^2 \le \big(\mathcal{R}(u) - \mathcal{R}(u^*)\big)\Big(\frac{V_{\max} - V_{\min}}{\lambda_1 - \lambda^*} + 1\Big)\|u\|_{L^2(\Omega)}^2.$$

In particular, if $\|u\|_{L^2} = 1$, then

$$\|P^\perp u\|_{H^1(\Omega)}^2 \le \Big(\frac{V_{\max} - V_{\min} + 1}{\lambda_1 - \lambda^*} + 1\Big)\big(\mathcal{R}(u) - \mathcal{R}(u^*)\big).$$

2.2. **Main results.** We aim to establish quantitative generalization error estimates between the approximate ground state u_n parametrized by neural networks and the exact ground state u^*. Our particular interest is to show that under certain circumstances the generalization error of the neural network solution does not suffer from the curse of dimensionality. To this end, we will assume (and prove below) that the exact ground state u^* lies in a smaller function space than the usual Sobolev space within which the functions can be approximated by neural networks without curse of dimensionality. Specifically, we consider the *spectral Barron space* [33] defined as follows.

Recall the domain $\Omega = [0,1]^d$. Let us first define the set of cosine functions

$$\mathcal{C} = \Big\{\Phi_k\Big\}_{k\in\mathbb{N}_0^d} := \Big\{\prod_{i=1}^d \cos(\pi k_i x_i) \mid k_i \in \mathbb{N}_0\Big\}.$$

Let $\{\hat{u}(k)\}_{k\in\mathbb{N}_0^d}$ be the expansion coefficients of a function $u \in L^1(\Omega)$ under the basis $\{\Phi_k\}_{k\in\mathbb{N}_0^d}$. For $s \ge 0$, the spectral Barron space $\mathcal{B}^s(\Omega)$ on Ω is defined by

$$\mathcal{B}^s(\Omega) := \Big\{u \in L^1(\Omega) : \sum_{k\in\mathbb{N}_0^d} (1 + \pi^s|k|_1^s)|\hat{u}(k)| < \infty\Big\}, \tag{2.4}$$

which is equipped with the spectral Barron norm

$$\|u\|_{\mathcal{B}^s(\Omega)} = \sum_{k\in\mathbb{N}_0^d} (1 + \pi^s|k|_1^s)|\hat{u}(k)|.$$

Note that we use $|k|_1$ to denote the ℓ^1-norm of a vector k. It is clear that $\mathcal{B}^s(\Omega)$ is a Banach space since it can be viewed as a weighted ℓ^1 space $\ell^1_{W_s}(\mathbb{N}_0^d)$ of the cosine coefficients defined on the lattice $\mathbb{N}_0^d$ with the weight $W_s(k) = (1 + \pi^s|k|_1^s)$. Moreover, since functions in $\mathcal{B}^s(\Omega)$ have summable cosine coefficients, we have $\mathcal{B}^s(\Omega) \hookrightarrow C(\overline{\Omega})$. When $s = 2$, we adopt the short notation $\mathcal{B}(\Omega)$ for $\mathcal{B}^2(\Omega)$. Our notion of spectral Barron space is an adaptation of the Barron space defined in the seminal work [5]; see also the recent works [4, 15, 30, 42] on other variants of Barron spaces. The original Barron function f in [5] is defined on the whole space $\mathbb{R}^d$ whose Fourier transform $\hat{f}(w)$ satisfies that $\int |\hat{f}(\omega)||\omega| d\omega < \infty$. Our spectral Barron space $\mathcal{B}^s(\Omega)$ with $s = 1$,

defined on the bounded domain Ω, can be viewed as a finite domain analog of the original Barron space from [5].

Example. We provide an example of the potential function $V \in \mathcal{B}([0,1]^d)$. Consider the potential $V(x_1, \cdots, x_N)$ of N pairwise interacting particles $\{x_i\}_{i=1}^N$ with

$$V(\mathbf{X}) \coloneqq V(x_1, \cdots, x_N) = \sum_{i<j}^{N} K(x_i, x_j). \tag{2.5}$$

Here we assume that $x_i \in [0,1]^{d_0}$ with $d_0 \geq 1$. Therefore the vector of N particles is $\mathbf{X} \in [0,1]^d$ with $d = Nd_0$. The kernel $K : [0,1]^{2d_0} \to \mathbb{R}$ describes the pairwise interaction between two particles. A straightforward calculation shows that if $K \in \mathcal{B}^s([0,1]^{2d_0})$, then $V \in \mathcal{B}^s([0,1]^d)$ with

$$\|V\|_{\mathcal{B}^s([0,1]^d)} \leq \frac{N(N-1)}{2} \|K\|_{\mathcal{B}^s([0,1]^{2d_0})}.$$

In fact, note that for any $k \in \mathbb{N}_0^d$,

$$\begin{aligned}
\hat{V}(k) &= \int_{\mathbb{R}^{Nd_0}} \sum_{i<j}^{N} K(x_i, x_j) \prod_{\ell=1}^{N} \cos(\pi k_\ell \cdot x_\ell) d\mathbf{X} \\
&= \sum_{i<j}^{N} \hat{K}(k_i, k_j) \prod_{\ell \notin \{i,j\}}^{N} \delta_0(k_\ell),
\end{aligned}$$

where $\delta_0(0) = 1$ and $\delta_0(k) = 0$ if $k \neq 0$. Hence

$$\begin{aligned}
\|V\|_{\mathcal{B}^s([0,1]^d)} &= \sum_{k \in \mathbb{N}_0^d} (1 + \pi^s |k|_1^s) |\hat{V}(k)| \\
&\leq \sum_{i<j}^{N} \sum_{k_i \in \mathbb{N}_0^{d_0}} \sum_{k_j \in \mathbb{N}_0^{d_0}} (1 + \pi^s (|k_i|_1 + |k_j|_1)^s) |\hat{K}(k_i, k_j)| \\
&= \frac{N(N-1)}{2} \|K\|_{\mathcal{B}^s([0,1]^{2d_0})}.
\end{aligned}$$

Moreover, to give a concrete example of $K \in \mathcal{B}^s([0,1]^{2d_0})$, consider the kernel $K(x,y) = |x-y|$ with $x, y \in [0,1]$. A straightforward calculation shows that the cosine coefficients of K, denoted by $\{\hat{K}(m,n)\}_{(m,n) \in \mathbb{N}_0^2}$, are given by

$$\hat{K}(m,n) = \begin{cases} 0, & \text{if } m \neq n, \\ -\frac{1}{m^2 \pi^2} & \text{if } m = n. \end{cases}$$

Therefore by definition, $K(x,y) = |x-y| \in \mathcal{B}^s([0,1]^2)$ for every $s \in (0,1)$ and hence the corresponding pairwise interacting potential $V(\mathbf{X}) \in \mathcal{B}^s([0,1]^N)$ with $s \in (0,1)$. Moreover, it holds that

$$\|V\|_{\mathcal{B}^s([0,1]^N)} \lesssim N^2.$$

In particular, the Barron norm of V does not increase exponentially in N.

Remark 2.1. Notice that since the spectral Barron space embeds continuously into the space of bounded continuous functions, the interacting potential V defined by (2.5) in the example above unfortunately excludes some important physical potentials that are singular, such as the Coulomb potential. Note also that, in practical applications

to many body electron systems, when the Coulomb potential is involved, often times specific ansatz that involve cusp conditions are used for the wave function. In such practical setting, the effective potential does not contain singularities and the theory developed here can still apply. Nonetheless, how to establish analogous Barron regularity of the Schrödinger equation with singular potentials remains an interesting open question. We will leave such cases for future studies.

Functions in the spectral Barron space differ substantially from those in Sobolev or Hölder spaces; most importantly, they can be approximated with respect to the H^1-norm by two-layer neural networks without curse of dimensionality. To make this more precise, we recall an approximation result from [33]. Let us define for an activation function ϕ, a constant $\mathcal{B} > 0$ and the number of hidden neurons m the set of functions

$$\mathcal{F}_{\phi,m}(\mathcal{B}) := \Big\{ c + \sum_{i=1}^{m} \gamma_i \phi(w_i \cdot x - t_i), |c| \le 2\mathcal{B}, |w_i|_1 = 1, |t_i| \le 1, \sum_{i=1}^{m} |\gamma_i| \le 4\mathcal{B} \Big\}. \tag{2.6}$$

Following our earlier work [33], we will focus on the rescaled Softplus activation function

$$\mathrm{SP}_\tau(z) = \frac{1}{\tau}\mathrm{SP}(\tau z) = \frac{1}{\tau}\ln(1 + e^{\tau z}),$$

where $\tau > 0$ is a rescaling parameter. Observe that $\mathrm{SP}_\tau \to \mathrm{ReLU}$ pointwisely as $\tau \to 0$ (see [33, Lemma 4.6]).

Let $\mathcal{F}_{\mathrm{SP}_\tau,m}(\mathcal{B})$ be the set of neural networks defined by setting $\phi = \mathrm{SP}_\tau$ in (2.6). Lemma 2.2 shows that functions in $\mathcal{B}(\Omega)$ can be well approximated by functions in $\mathcal{F}_{\mathrm{SP}_\tau,m}(\mathcal{B})$ without curse of dimensionality.

Lemma 2.2. [33, Theorem 2.2] *For any $u \in \mathcal{B}(\Omega)$, there exists a two-layer neural network $u_m \in \mathcal{F}_{\mathrm{SP}_\tau,m}(\|u\|_{\mathcal{B}(\Omega)})$ with $\tau = \sqrt{m}$, such that*

$$\|u - u_m\|_{H^1(\Omega)} \le \frac{\|u\|_{\mathcal{B}(\Omega)}(6\ln m + 30)}{\sqrt{m}}. \tag{2.7}$$

The proof of the approximation bound (2.7) relies on first establishing a similar bound for two-layer networks with the ReLU activation function and then replacing ReLU with SP_τ; the latter step induces the factor $\ln m$ on the right side of (2.7). We remark that the bound (2.7) does not only hold for two-layer networks with ReLU or SP_τ function, but can still be valid for other activation functions. In fact, the recent works [27,43] have shown that the convergence rate $O(m^{-1/2})$ of Lemma 2.2 can be improved to $O(m^{-\frac{1}{2}-\frac{\alpha}{d}})$ if ReLU^k is used as the activation function where $\alpha > 0$ depends on k. Since the focus of the present paper is not to achieve the sharpest convergence estimate, we are content with Lemma 2.2 since this is enough to get a dimensional-independent rate for the generalization error.

With the above approximation result in Lemma 2.2 at hand, we are ready to state the main generalization theorem as follows.

Theorem 2.3. *Assume that $\mathcal{H}$ satisfies Assumptions* 2.1 *and* 2.2. *Assume also that the ground state $u^* \in \mathcal{B}(\Omega)$. Let u_n^m be a minimizer of the empirical loss $\mathcal{R}_n$ within the set $\mathcal{F} = \mathcal{F}_{\mathrm{SP}\tau,m}(\mathcal{B})$ with $\mathcal{B} = \|u^*\|_{\mathcal{B}(\Omega)}$ and with $\tau = \sqrt{m}$. Given $\delta \in (0, \frac{1}{3})$, assume that n and m are large enough so that*

$$\xi_i(n,\delta) \le 1/2, i = 1,2,3 \text{ and } \eta(\mathcal{B},m) \le 1/2, \tag{2.8}$$

where $\xi_i(n,\delta)$ are defined in (4.3), (4.7) *and* (4.12), *and $\eta(\mathcal{B},m)$ is defined in* (4.16). *Then with probability at least* $1-3\delta$,

$$\mathcal{R}(u_n^m)-\mathcal{R}(u^*)\le \frac{C_1\ln(\frac{1}{\delta})\sqrt{m(\ln m+1)}}{\sqrt{n}}+\frac{C_2(\ln m+1)}{\sqrt{m}}, \tag{2.9}$$

where C_1 depends on $\|u^*\|_{\mathcal{B}(\Omega)}, d, V_{\max}$ polynomially and C_2 depends on $\|u^*\|_{\mathcal{B}(\Omega)}$ linearly. In particular, with the choice $m=\sqrt{n}$, we have that there exists $C_3>0$ such that with probability at least $1-3\delta$,

$$\mathcal{R}(u_n^m)-\mathcal{R}(u^*)\le C_3\ln\Big(\frac{1}{\delta}\Big)\cdot n^{-\frac{1}{4}}\ln n.$$

The proof of Theorem 2.3 relies on decomposing the generalization error into the sum of the approximation error (the second term on the right side of (2.9)) and statistical error (the first term on the right side of (2.9)) arising from the Monte Carlo approximation. The statistical error is further bounded by controlling the Rademacher complexity of certain neural network classes associated to the loss formulation (see Section 4.3). It is worth to comment that the numerator of the statistical error scales like $O(\sqrt{m(\ln m)})$ for $m\gg 1$, which seems worse than the bound proved in [15] where the statistical error (or the Rademacher complexity) scales like $O(\frac{\|u\|_{\mathcal{B}}}{n^{1/2}})$. This is mainly because [15] only considers the Rademacher complexity bound of Barron functions with finite Barron norm while we need to bound the Rademacher complexities of several neural network classes with m-dependent network parameters.

Let us also comment about the largeness assumption on m and n. In fact, by tracking the proof of Theorem 2.3, the condition (2.8) holds as long as m and n are larger than $Poly(\|u^*\|_{\mathcal{B}(\Omega)})$. Therefore m and n need not be exponentially in d if $\|u^*\|_{\mathcal{B}(\Omega)}$ does not grow exponentially in d.

Thanks to Proposition 2.1, the generalization error in terms of the energy excess translates directly to that in terms of the H^1-norm of $P^\perp u_n^m$.

Theorem 2.4. *Suppose that the assumption of Theorem* 2.3 *holds and suppose further that $\mathcal{H}$ has a spectral gap. Then there exist positive constants C_4 and C_5 depending polynomially on $\|u^*\|_{\mathcal{B}(\Omega)}, d, V_{\min}, V_{\max}$ and $\lambda_1-\lambda_0$ such that with probability at least* $1-3\delta$,

$$\|P^\perp u_n^m\|_{H^1(\Omega)}^2\le \frac{C_4\ln(\frac{1}{\delta})\sqrt{m(\ln m+1)}}{\sqrt{n}}+\frac{C_5(\ln m+1)}{\sqrt{m}}.$$

Setting $m=\sqrt{n}$ in the above leads to that the following holds for some $C_6>0$ with probability at least $1-3\delta$:

$$\|P^\perp u_n^m\|_{H^1(\Omega)}^2\le C_6\ln\Big(\frac{1}{\delta}\Big)\cdot n^{-\frac{1}{4}}\ln n.$$

Theorem 2.3 and Theorem 2.4 show that with high probability the convergence rate of the generalization error of two-layer network for approximating the ground state u^* and the corresponding leading eigenvalue $\lambda^*=\mathcal{R}(u^*)$ does not suffer from the curse of dimensionality provided that the ground state $u^*\in\mathcal{B}(\Omega)$.

Finally we justify the regularity assumption on the ground state in Theorem 2.5. This gives a novel solution theory of high dimensional eigenvalue problems in Barron type spaces.

Theorem 2.5. *Assume that $V \in \mathcal{B}^s(\Omega)$ with $s \geq 0$ and V satisfies Assumption* 2.1. *Then the ground state $u^* \in \mathcal{B}^{s+2}(\Omega)$.*

Our idea of proving Theorem 2.5 differs from the standard proof of Sobolev regularity of eigenfunctions, which usually relies on bootstrapping estimates on the weak derivative of the eigenfunctions. Instead, we prove Theorem 2.5 by reformulating the ground state problem as a fixed point problem on the spectral Barron space $\mathcal{B}^s(\Omega)$; the existence of a nontrivial fixed point is proved by employing the celebrated Krein-Rutman theorem [31]. See Section 5 for a complete proof.

3. Stability estimate of the ground state (Proof of Proposition 2.1)

In this section, we show that the offset $\|P^\perp u\|_{L^2(\Omega)}$ of any $u \in H^1(\Omega)$ can be bounded by the energy excess $\mathcal{R}(u) - \mathcal{R}(u^*)$.

Proof. Note that by the spectral gap assumption, for any $u \in H^1(\Omega)$,

$$\langle \mathcal{H}(P^\perp - \lambda_0 P^\perp)u, u\rangle \geq (\lambda_1 - \lambda_0)\|P^\perp u\|^2_{L^2(\Omega)}.$$

Let us decompose

$$u = Pu + P^\perp u =: \alpha u_0 + u_\perp,$$

where $\alpha = \langle u, u_0\rangle$ and $\langle u_\perp, u_0\rangle = 0$. Substituting above into the Rayleigh quotient, we have

$$\begin{aligned}
\mathcal{R}(u) &= \frac{|\alpha|^2\lambda_0 + \langle u_\perp, \mathcal{H}u_\perp\rangle}{\|u\|^2} \\
&\geq \frac{\lambda_0\|u\|^2 + (\lambda_1 - \lambda_0)\|u_\perp\|^2}{\|u\|^2} \\
&= \lambda_0 + (\lambda_1 - \lambda_0)\frac{\|u_\perp\|^2}{\|u\|^2} \\
&= \mathcal{R}(u^*) + (\lambda_1 - \lambda_0)\frac{\|u_\perp\|^2}{\|u\|^2}.
\end{aligned}$$

Thus, the L^2-norm of $u_\perp \equiv P^\perp u$ can be bounded as

$$\|u_\perp\|^2 \leq \frac{\mathcal{R}(u) - \mathcal{R}(u_0)}{\lambda_1 - \lambda_0}\|u\|^2. \tag{3.1}$$

Note that this bound cannot be improved as can be seen by taking a linear combination of the first and second eigenfunctions of $\mathcal{H}$.

To obtain the bound on $\nabla u^\perp$, we notice that

$$\mathcal{R}(u) = \lambda_0 + \frac{\langle u_\perp, (\mathcal{H} - \lambda_0)u_\perp\rangle}{\|u\|^2}.$$

Thus

$$\begin{aligned}
\langle u_\perp, (\mathcal{H} - \lambda_0)u_\perp\rangle &= \|\nabla u_\perp\|^2_{L^2} + \langle u_\perp, (V - \lambda_0)u_\perp\rangle \\
&\leq \big(\mathcal{R}(u) - \mathcal{R}(u_0)\big)\|u\|^2.
\end{aligned}$$

Rearranging the terms, we arrive at

$$\begin{aligned}
\|\nabla u_\perp\|^2 &\leq -\langle u_\perp, (V - \lambda_0)u_\perp\rangle + \big(\mathcal{R}(u) - \mathcal{R}(u_0)\big)\|u\|^2 \\
&\leq \big(V_{\max} - V_{\min}\big)\|u_\perp\|^2 + \big(\mathcal{R}(u) - \mathcal{R}(u_0)\big)\|u\|^2 \\
&\overset{(3.1)}{\leq} \big(\mathcal{R}(u) - \mathcal{R}(u_0)\big)\Big(\frac{V_{\max} - V_{\min}}{\lambda_1 - \lambda_0} + 1\Big)\|u\|^2.
\end{aligned}$$

□

4. Proof of Theorem 2.3

4.1. **Oracle inequality for the generalization error.** We first introduce an oracle inequality for the empirical loss that decomposes the generalization error into the sum of approximation error and statistical error. Recall the population loss $\mathcal{R}$ and the empirical loss $\mathcal{R}_n$ defined in (2.2) and (2.3) respectively. Consider the minimization of $\mathcal{R}_n$ in a function class $\mathcal{F}$ and we denote by u_n a minimizer of $\mathcal{R}_n$ within $\mathcal{F}$, i.e. $u_n = \arg\min_{u\in\mathcal{F}} \mathcal{R}_n(u)$.

We aim to bound the energy excess $\Delta\mathcal{R}_n := \mathcal{R}(u_n) - \mathcal{R}(u^*)$ where u^* is the exact ground state. Let us first decompose $\Delta\mathcal{R}_n$ as follows:

$$\Delta\mathcal{R}_n = \mathcal{R}(u_n)-\mathcal{R}_n(u_n)+\mathcal{R}_n(u_n)-\mathcal{R}_n(u_{\mathcal{F}})+\mathcal{R}_n(u_{\mathcal{F}})-\mathcal{R}(u_{\mathcal{F}})+\mathcal{R}(u_{\mathcal{F}})-\mathcal{R}(u^*). \tag{4.1}$$

Here $u_{\mathcal{F}} = \arg\min_{u\in\mathcal{F}} \mathcal{R}(u)$. Note that $\mathcal{R}_n(u_n)-\mathcal{R}_n(u_{\mathcal{F}}) \le 0$ since u_n is the minimizer of $\mathcal{R}_n$. Therefore

$$\Delta\mathcal{R}_n \le \big(\mathcal{R}(u_n) - \mathcal{R}_n(u_n)\big) + \big(\mathcal{R}_n(u_{\mathcal{F}}) - \mathcal{R}(u_{\mathcal{F}})\big) + \big(\mathcal{R}(u_{\mathcal{F}}) - \mathcal{R}(u^*)\big) =: T_1 + T_2 + T_3.$$

Note that the first term T_1 is the statistical error arising from the random approximation of integrals, the second term T_2 is the Monte Carlo error and the third term T_3 is the approximation error term due to restricting the minimization of $\mathcal{E}$ from over the set $H^1(\Omega)$ to $\mathcal{F}$; see an upper bound of T_3 in Theorem 4.5 when $\mathcal{F}$ is chosen as the set of two-layer neural networks. To control the statistical errors, we employ the well-known tool of Rademacher complexity, for which we recall its definition as follows.

Definition 4.1. We define for a set of random variables $\{Z_j\}_{j=1}^n$ independently distributed according to $\mathcal{P}_\Omega$ and a function class $\mathcal{S}$ the random variable

$$\hat{R}_n(\mathcal{S}) := \mathbf{E}_\sigma\Big[\sup_{g\in\mathcal{S}}\Big|\frac{1}{n}\sum_{j=1}^n \sigma_j g(Z_j)\Big| \;\Big|\; Z_1,\cdots,Z_n\Big],$$

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 $\mathcal{S}$ is defined by $R_n(\mathcal{S}) = \mathbf{E}_{\mathcal{P}_\Omega^n}[\hat{R}_n(\mathcal{S})]$.

Now we are ready to bound T_1 and T_2 in terms of Rademacher complexities of suitable function classes.

Bounding T_1. Thanks to the scale invariance of $\mathcal{R}$ and $\mathcal{R}_n$, we can assume that u_n is normalized, i.e. $\|u_n\|_2 = 1$. Hence we have

$$\begin{aligned} T_1 &\le \Big|\frac{\mathcal{E}_{n,V}(u_n)}{\mathcal{E}_{n,2}(u_n)} - \frac{\mathcal{E}_V(u_n)}{\mathcal{E}_2(u_n)}\Big| \\ &\le \frac{\big|\mathcal{E}_{n,V}(u_n) - \mathcal{E}_V(u_n)\big|}{\mathcal{E}_{n,2}(u_n)} + \frac{\mathcal{E}_V(u_n)}{\mathcal{E}_2(u_n)\cdot\mathcal{E}_{n,2}(u_n)}\big|\mathcal{E}_2(u_n) - \mathcal{E}_{n,2}(u_n)\big| \\ &=: T_{11} + T_{12}. \end{aligned}$$

To bound T_{11} and T_{12}, let us define two sets of functions

$$\begin{aligned} \mathcal{G}_1 &:= \{g : g = u^2 \text{ where } u \in \mathcal{F}\}, \\ \mathcal{G}_2 &:= \{g : g = |\nabla u|^2 + V|u|^2 \text{ where } u \in \mathcal{F}\}. \end{aligned} \tag{4.2}$$

We assume that the set $\mathcal{F}$ satisfies $\sup_{u\in\mathcal{F}} \|u\|_{L^\infty} \le M_{\mathcal{F}} < \infty$ so that $\sup_{g\in\mathcal{G}_1} \|g\|_{L^\infty} \le M_{\mathcal{F}}^2$. Assume further that $\sup_{g\in\mathcal{G}_2} \|g\|_{L^\infty} \le M_{\mathcal{G}_2} < \infty$. Now let us first derive a high-probability lower bound for $\mathcal{E}_{n,2}(u_n)$. For doing so, we define for $n \in \mathbb{N}$ and $\delta > 0$ the constant

$$\xi_1(n,\delta) := 2R_n(\mathcal{G}_1) + 4M_{\mathcal{F}}^2 \cdot \sqrt{\frac{2\ln(4/\delta)}{n}}, \tag{4.3}$$

where $R_n(\mathcal{G}_1)$ denotes the Rademacher complexity of the set $\mathcal{G}_1$. We also define the event

$$A_1(n,\delta) := \left\{ |\mathcal{E}_{n,2}(u_n) - \mathcal{E}_2(u_n)| \le \xi_1(n,\delta) \right\}.$$

Then applying Lemma 4.3 to $\mathcal{G}_1$ we have that

$$\mathbf{P}\big[A_1(n,\delta)\big] \ge 1-\delta. \tag{4.4}$$

Since by assumption $\mathcal{E}_2(u_n) = 1$, within the event $A_1(n,\delta)$ we have $\mathcal{E}_{n,2}(u_n) \ge 1 - \xi_1(n,\delta)$ and hence

$$\mathbf{P}\big[\mathcal{E}_{n,2}(u_n) \ge 1-\xi_1(n,\delta)\big] \ge 1-\delta. \tag{4.5}$$

Notice that $\mathcal{E}_V(u_n) \le M_{\mathcal{G}_2}$ by the assumption on $\mathcal{G}_2$. Therefore if $\xi_1(n,d) < 1$, then

$$\mathbf{P}\Big[T_{12} \le \frac{M_{\mathcal{G}_2}\cdot\xi_1(n,\delta)}{1-\xi_1(n,\delta)}\Big] \ge \mathbf{P}\big[A_1(n,\delta)\big] \ge 1-\delta. \tag{4.6}$$

Next to bound T_{11}, let us define the constant

$$\xi_2(n,\delta) := 2R_n(\mathcal{G}_2) + 4M_{\mathcal{G}_2} \cdot \sqrt{\frac{2\ln(4/\delta)}{n}} \tag{4.7}$$

and the event

$$A_2(n,\delta) := \Big\{ \sup_{u\in\mathcal{F}} \big|\mathcal{E}_{n,V}(u) - \mathcal{E}_V(u)\big| \le \xi_2(n,\delta) \Big\}.$$

Then applying again Lemma 4.3 to $\mathcal{G}_2$ leads to

$$\mathbf{P}\big[A_2(n,\delta)\big] \ge 1-\delta. \tag{4.8}$$

As a result of (4.5) and (4.8), one has that if $\xi_1(n,d) < 1$ then

$$\mathbf{P}\Big[T_{11} \le \frac{\xi_2(n,d)}{1-\xi_1(n,d)}\Big] \ge \mathbf{P}\Big[A_1(n,\delta)\cap A_2(n,\delta)\Big] \ge 1-2\delta. \tag{4.9}$$

Therefore it follows from (4.6) and (4.9) that

$$\mathbf{P}\left[T_1 \le \frac{M_{\mathcal{G}_2}\cdot\xi_1(n,d) + \xi_2(n,d)}{1-\xi_1(n,d)}\right] \ge \mathbf{P}\Big[A_1(n,\delta)\cap A_2(n,\delta)\Big] \ge 1-2\delta. \tag{4.10}$$

Bounding T_2. Similar to the process of bounding T_1, by assuming $\|u_{\mathcal{F}}\|_{L^2} = 1$ we first bound T_2 as follows

$$\begin{aligned} T_2 &\le \Big| \frac{\mathcal{E}_{n,V}(u_{\mathcal{F}})}{\mathcal{E}_{n,2}(u_{\mathcal{F}})} - \frac{\mathcal{E}_V(u_{\mathcal{F}})}{\mathcal{E}_2(u_{\mathcal{F}})} \Big| \\ &\le \frac{\big|\mathcal{E}_{n,V}(u_{\mathcal{F}}) - \mathcal{E}_V(u_{\mathcal{F}})\big|}{\mathcal{E}_{n,2}(u_{\mathcal{F}})} + \frac{\mathcal{E}_V(u_{\mathcal{F}})}{\mathcal{E}_2(u_{\mathcal{F}})\cdot\mathcal{E}_{n,2}(u_{\mathcal{F}})}\big|\mathcal{E}_2(u_{\mathcal{F}}) - \mathcal{E}_{n,2}(u_{\mathcal{F}})\big| \\ &=: T_{21} + T_{22}. \end{aligned}$$

Since $u_{\mathcal{F}}$ does not depend on the sample points $\{X_i\}$, applying Hoeffding's inequality from Lemma 4.2 yields that

$$\mathbf{P}[A_3(n,\delta)] \coloneqq \mathbf{P}\Big[|\mathcal{E}_{n,2}(u_{\mathcal{F}}) - \mathcal{E}_2(u_{\mathcal{F}})| \le \xi_3(n,\delta)\Big] \ge 1-\delta, \tag{4.11}$$

where

$$\xi_3(n,\delta) \coloneqq M_{\mathcal{F}}^2 \cdot \sqrt{\frac{\ln(2/\delta)}{2n}}. \tag{4.12}$$

Since by assumption $\mathcal{E}_2(u_{\mathcal{F}}) = 1$, this implies further that

$$\mathbf{P}\Big[\mathcal{E}_{n,2}(u_{\mathcal{F}}) \ge 1 - \xi_3(n,\delta)\Big] \ge \mathbf{P}[A_3(n,\delta)] \ge 1-\delta. \tag{4.13}$$

Combining (4.11) and (4.13) implies that if $\xi_3(n,\delta) < 1$, then

$$\mathbf{P}\left[T_{22} \le \frac{M_{\mathcal{G}_2}\cdot\xi_3(n,\delta)}{1-\xi_3(n,\delta)}\right] \ge \mathbf{P}[A_3(n,\delta)] \ge 1-\delta.$$

In addition, as a consequence of (4.13) and (4.8), we have

$$\mathbf{P}\left[T_{21} \le \frac{\xi_2(n,\delta)}{1-\xi_3(n,\delta)}\right] \ge \mathbf{P}[A_2(n,\delta)\cap A_3(n,\delta)] \ge 1-2\delta.$$

Therefore it holds that

$$\mathbf{P}\left[T_2 \le \frac{\xi_2(n,\delta) + M_{\mathcal{G}_2}\cdot\xi_3(n,\delta)}{1-\xi_3(n,\delta)}\right] \ge 1-2\delta. \tag{4.14}$$

Lemma 4.2 (Hoeffding). *Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. random variables with $a_i \le Z_i \le b_i$ a.s. Then for any $t > 0$,*

$$\mathbf{P}\Big(\Big|\frac{\sum_{i=1}^n Z_i}{n} - \mathbf{E}Z\Big| \ge t\Big) \le 2\exp\Big(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\Big).$$

We recall the following useful PAC-type generalization bound via the Rademacher complexity.

Lemma 4.3. [39, Theorem 26.5] *Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. random variables. Let $\mathcal{G}$ be a function class such that $\sup_{g\in\mathcal{G}} \|g\|_{L^\infty(\Omega)} \le C_{\mathcal{G}}$ and that $\mathcal{G}$ is symmetric, i.e. $\mathcal{G} = -\mathcal{G}$. Then with probability at least $1-\delta$,*

$$\sup_{g\in\mathcal{G}} \Big|\frac{1}{n}\sum_{i=1}^n g(Z_i) - \mathbf{E}g(Z)\Big| \le 2R_n(\mathcal{G}) + 4C_{\mathcal{G}}\sqrt{\frac{2\ln(4/\delta)}{n}}.$$

Combining the bounds derived above leads to the following oracle inequality. Recall the quantities $\xi_i(n,\delta), i = 1,2,3$ defined in (4.3), (4.7) and (4.12).

Theorem 4.4. *Let $u_n = \arg\min_{u\in\mathcal{F}} \mathcal{R}_n(u)$. Let $\delta \in (0, \frac{1}{3})$ be fixed. Assume that $\xi_i(n,\delta) < 1, i = 1,2,3$.*

Then with probability at least $1-3\delta$,

$$\begin{aligned}\mathcal{R}(u_n) - \mathcal{R}(u^*) \le{}& \frac{M_{\mathcal{G}_2}\cdot\xi_1(n,d) + \xi_2(n,d)}{1-\xi_1(n,d)} + \frac{M_{\mathcal{G}_2}\cdot\xi_3(n,\delta) + \xi_2(n,\delta)}{1-\xi_3(n,\delta)} \\ &+ \big(\inf_{u\in\mathcal{F}} \mathcal{R}(u) - \mathcal{R}(u^*)\big).\end{aligned} \tag{4.15}$$

4.2. **Bounding the approximation error.** Recall the spectral Barron space $\mathcal{B}^s(\Omega)$ defined in (2.4) and the set of two-layer networks $\mathcal{F}_{\mathrm{SP}_\tau,m}(B)$ defined by setting $\phi = \mathrm{SP}_\tau$ in (2.6). Theorem 4.5 bounds the approximation error (the third term) in (4.15) when $u^* \in \mathcal{B}(\Omega)$ and $\mathcal{F} = \mathcal{F}_{\mathrm{SP}_\tau,m}$.

Theorem 4.5. *Let the ground state $u^* \in \mathcal{B}(\Omega)$ with $\|u^*\|_{L^2(\Omega)} = 1$. Let $u_m \in \mathcal{F}_{\mathrm{SP}_\tau,m}(\|u\|_{\mathcal{B}(\Omega)})$ be defined in Lemma 2.2. Assume that V satisfies Assumption 2.1. Assume in addition that*

$$\eta(\|u^*\|_{\mathcal{B}(\Omega)}, m) := \frac{\|u^*\|_{\mathcal{B}(\Omega)} \cdot (6\ln m + 30)}{\sqrt{m}} \le \frac{1}{2}. \tag{4.16}$$

Then

$$\mathcal{R}(u_m) - \mathcal{R}(u^*) \le \left(2(1+V_{\max})\Big(\sqrt{\frac{\lambda^*}{\min(1,V_{\min})}}+1\Big) + 3\lambda^*\right)\eta(\|u^*\|_{\mathcal{B}(\Omega)}, m). \tag{4.17}$$

Proof. By assumption $\mathcal{E}_2(u^*) = \|u^*\|^2_{L^2(\Omega)} = 1$. Then $\mathcal{R}(u^*) = \mathcal{E}_V(u^*) = \lambda^*$. Since $V \ge V_{\min} > 0$, this implies that

$$\|u^*\|^2_{H^1(\Omega)} \le \frac{\mathcal{E}_V(u^*)}{\min(1,V_{\min})} = \frac{\lambda^*}{\min(1,V_{\min})}. \tag{4.18}$$

Now observe that

$$\begin{aligned}\mathcal{R}(u_m) - \mathcal{R}(u^*) &= \frac{\mathcal{E}_V(u_m)}{\mathcal{E}_2(u_m)} - \frac{\mathcal{E}_V(u^*)}{\mathcal{E}_2(u^*)} \\ &= \frac{\mathcal{E}_V(u_m) - \mathcal{E}_V(u^*)}{\mathcal{E}_2(u_m)} + \frac{\mathcal{E}_2(u^*) - \mathcal{E}_2(u_m)}{\mathcal{E}_2(u_m)} \cdot \mathcal{R}(u^*).\end{aligned} \tag{4.19}$$

Thanks to Lemma 2.2, $\|u_m - u^*\|_{H^1(\Omega)} \le \eta(\|u^*\|_{\mathcal{B}(\Omega)}, m) < 1$. This implies that

$$1 - \eta(\|u^*\|_{\mathcal{B}(\Omega)}, m) \le \|u_m\|_{L^2(\Omega)} \le 1 + \eta(\|u^*\|_{\mathcal{B}(\Omega)}, m)$$

and that

$$\begin{aligned}|\mathcal{E}_2(u_m) - \mathcal{E}_2(u^*)| &= (\|u^*\|_{L^2(\Omega)} + \|u_m\|_{L^2(\Omega)})|\|u^*\|_{L^2(\Omega)} - \|u_m\|_{L^2(\Omega)}| \\ &\le (2 + \eta(\|u^*\|_{\mathcal{B}(\Omega)}, m))\eta(\|u^*\|_{\mathcal{B}(\Omega)}, m) \\ &\le 3\eta(\|u^*\|_{\mathcal{B}(\Omega)}, m).\end{aligned}$$

In addition, it follows from the boundedness of V that

$$\begin{aligned}|\mathcal{E}_V(u_m) - \mathcal{E}_V(u^*)| &\le (1+V_{\max})(\|u^*\|_{H^1(\Omega)} + \|u_m\|_{H^1(\Omega)})\|u^* - u_m\|_{H^1(\Omega)} \\ &\le (1+V_{\max})(2\|u^*\|_{H^1(\Omega)} + \eta(\|u^*\|_{\mathcal{B}(\Omega)}, m))\eta(\|u^*\|_{\mathcal{B}(\Omega)}, m) \\ &\le (1+V_{\max})\Big(\sqrt{\frac{\lambda^*}{\min(1,V_{\min})}}+1\Big)\eta(\|u^*\|_{\mathcal{B}(\Omega)}, m),\end{aligned}$$

where we have used (4.18) in the last inequality. Finally, the estimate follows simply by substituting the last two estimates into (4.19). □

4.3. **Bounding the statistical error.** In this section we proceed to bound the statistical errors (the first two terms on the right side of (4.15)). This is achieved by controlling the Rademacher complexities of the function classes $\mathcal{G}_1$ and $\mathcal{G}_2$ defined in (4.2). More specifically, since we set the trial functions $\mathcal{F} = \mathcal{F}_{\mathrm{SP}_\tau,m}$, we need to bound the Rademacher complexities of the following

$$\begin{aligned} &\mathcal{G}_{\mathrm{SP}_\tau,m,1}(\mathcal{B}) \coloneqq \{g = u^2 : u \in \mathcal{F}_{\mathrm{SP}_\tau,m}(\mathcal{B})\}, \\ &\mathcal{G}_{\mathrm{SP}_\tau,m,2}(\mathcal{B}) \coloneqq \{g = |\nabla u|^2 + V|u|^2 : u \in \mathcal{F}_{\mathrm{SP}_\tau,m}(\mathcal{B})\}. \end{aligned} \tag{4.20}$$

Theorem 4.6. *Assume that $\|V\|_{L^\infty(\Omega)} \le V_{\max}$, consider the sets $\mathcal{G}_{\mathrm{SP}_\tau,m,1}(\mathcal{B})$ and $\mathcal{G}_{\mathrm{SP}_\tau,m,2}(\mathcal{B})$ with $\tau = \sqrt{m}$ and $\mathcal{B} > 0$. Then there exist positive constants $C_1(\mathcal{B}, d, F)$ and $C_2(\mathcal{B}, d, V_{\max})$ depending polynomially on $\mathcal{B}, d, V_{\max}$ such that*

$$R_n(\mathcal{G}_{\mathrm{SP}_\tau,m,1}(\mathcal{B})) \le \frac{C_1(\mathcal{B},d)\sqrt{m}(\sqrt{\ln m}+1)}{\sqrt{n}}, \tag{4.21}$$

$$R_n(\mathcal{G}_{\mathrm{SP}_\tau,m,2}(\mathcal{B})) \le \frac{C_2(\mathcal{B},d,V_{\max})\sqrt{m}(\sqrt{\ln m}+1)}{\sqrt{n}}. \tag{4.22}$$

To prove Theorem 4.6, we rely on the celebrated Dudley's theorem [13], which bounds the Rademacher complexities in terms of the metric entropy. Below we restate the Dudley's theorem by following [46, Theorem 1.19].

Theorem 4.7 (Dudley's theorem [46, Theorem 1.19])**.** *Let $\mathcal{F}$ be a function class such that $\sup_{f\in\mathcal{F}} \|f\|_\infty \le M$. Then the Rademacher complexity $R_n(\mathcal{F})$ satisfies that*

$$R_n(\mathcal{F}) \le \inf_{0\le\delta\le M}\Big\{4\delta + \frac{12}{\sqrt{n}}\int_\delta^M \sqrt{\ln \mathcal{N}(\varepsilon,\mathcal{F},\|\cdot\|_\infty)}\,d\varepsilon\Big\}, \tag{4.23}$$

where $\mathcal{N}(\varepsilon,\mathcal{F},\|\cdot\|_\infty)$ denotes the ε-covering number of $\mathcal{F}$ w.r.t the L_∞-norm.

In order to apply Dudley's theorem, we need to bound the δ-covering numbers of the sets $\mathcal{G}_{\mathrm{SP}_\tau,m,i}(\mathcal{B}), i = 1, 2$. To this end, it will be convenient to introduce the following functions

$$\mathcal{M}(\delta,\Lambda,m,d) \coloneqq \frac{4\mathcal{B}\Lambda}{\delta}\cdot\Big(\frac{12\mathcal{B}\Lambda}{\delta}\Big)^m\cdot\Big(\frac{3\Lambda}{\delta}\Big)^{dm}\cdot\Big(\frac{3\Lambda}{\delta}\Big)^m, \tag{4.24}$$

$$\begin{aligned} \mathcal{Z}(M,\Lambda,d) \coloneqq M\Big(\sqrt{(\ln(4\mathcal{B}\Lambda))_+} + \sqrt{(\ln(12\mathcal{B}\Lambda) + d\ln(3\Lambda) + \ln(3\Lambda))_+}\Big) \\ + \sqrt{d+3}\int_0^M \sqrt{(\ln(1/\varepsilon))_+}\,d\varepsilon. \end{aligned} \tag{4.25}$$

Lemma 4.8. *Consider the two sets $\mathcal{G}_{\mathrm{SP}_\tau,m,i}(\mathcal{B}), i = 1, 2$ defined in (4.2) with $\mathcal{B} > 0$. Then for any $\delta > 0$,*

$$\mathcal{N}(\delta,\mathcal{G}_{\mathrm{SP}_\tau,m,i}(\mathcal{B}),\|\cdot\|_\infty) \le \mathcal{M}(\delta,\Lambda_i,m,d), \quad i = 1, 2,$$

where the constants $\Lambda_i, i = 1, 2$ satisfy that

$$\Lambda_1 \le 36\mathcal{B}(5+8\mathcal{B}), \quad \Lambda_2 \le 64\mathcal{B}^2\sqrt{m} + 8\mathcal{B} + 36V_{\max}\mathcal{B}(5+8\mathcal{B}). \tag{4.26}$$

Proof. The proof follows directly from [33, Lemma 5.5] and [33, Lemma 5.7] by slightly adjusting the constants. We thus omit the details. □

Proof of Theorem 4.6. First from the definition of $\mathcal{F}_{\mathrm{SP}_\tau,m}(\mathcal{B})$ and the fact that $\|\mathrm{SP}_\tau\|_{W^{1,\infty}(\Omega)} \le 3 + \frac{1}{\tau}$ (see [33, Lemma 4.6]), one has the following uniform bound

$$\sup_{u\in\mathcal{F}_{\mathrm{SP}_\tau,m}(B)} \|u\|_{W^{1,\infty}(\Omega)} \le 2\mathcal{B} + 4\mathcal{B}\|\mathrm{SP}_\tau\|_{W^{1,\infty}(\Omega)} \le 16\mathcal{B}.$$

This implies that

$$\sup_{f\in\mathcal{F}_{\mathrm{SP}_\tau,m}(\mathcal{B})} \|f\|_{L^\infty(\Omega)} \le 16\mathcal{B} = M_{\mathcal{F}},$$

$$\sup_{g\in\mathcal{G}_{\mathrm{SP}_\tau,m,1}(\mathcal{B})} \|g\|_{L^\infty(\Omega)} \le (16\mathcal{B})^2 =: M_1,$$

$$\sup_{g\in\mathcal{G}_{\mathrm{SP}_\tau,m,2}(\mathcal{B})} \|g\|_{L^\infty(\Omega)} \le (1+V_{\max})(16\mathcal{B})^2 =: M_2.$$

Applying Theorem 4.7 with $\delta = 0$, we obtain from Lemma 4.8 and the estimates in the last line that

$$R_n(\mathcal{G}_{\mathrm{SP}_\tau,m,i}(\mathcal{B})) \le \mathcal{Z}(M_i,\Lambda_i,d)\sqrt{\frac{m}{n}}.$$

Moreover, after plugging the bounds on Λ_i (see (4.26)) into $\mathcal{Z}(M_i,\Lambda_i,d)$, it is easy to see that

$$\mathcal{Z}(M_1,\Lambda_1,d) \le C_1(\mathcal{B},d),$$

$$\mathcal{Z}(M_1,\Lambda_1,d) \le C_2(\mathcal{B},d,V_{\max})\sqrt{\ln m},$$

where the positive constants $C_1(\mathcal{B},d)$. and $C_2(\mathcal{B},d,V_{\max})$ depend on B and d polynomially. Combining the estimates finishes the proof of Theorem 4.6. □

4.4. **Proof of Theorem 2.3.** The proof follows directly by combining Theorem 4.4, Theorem 4.5 and Theorem 4.6.

5. Regularity of the ground state of Schrödinger operator in the spectral Barron space (Proof of Theorem 2.5)

In this section we aim to prove the regularity of the ground state u^* in the spectral Barron space as shown in Theorem 2.5. Since our proof relies heavily on the spectrum theory of positive linear operators on ordered Banach spaces (especially the Krein-Rutman theorem), we first recall some relevant terminologies and useful facts from linear functional analysis.

5.1. **A simple lemma on spectral radius.** Let E be a Banach space. Given a bounded linear operator T on E, we recall the resolvent set of T defined by

$$\rho_E(T) := \{\lambda\in\mathbb{C} \mid (\lambda I - T) \text{ is bijective on } E\}.$$

The spectrum $\sigma_E(T)$ of T is the set $\mathbb{C}\setminus\rho_E(T)$, where we have used the subscript E to indicate the dependence on the Banach space E. We further denote by $\sigma_{p,E}(T)$ the point spectrum of T, i.e.

$$\sigma_{p,E}(T) = \{\lambda\in\mathbb{C} \mid \text{ there exists } v\in E \text{ and } v\neq 0 \text{ such that } (\lambda I - T)v = 0\} \subset \sigma_E(T),$$

where we call $\lambda\in\sigma_{p,E}(T)$ an eigenvalue of T and v an eigenvector of T. The spectral radius of T is given by

$$r_E(T) := \sup\{|\lambda| \mid \lambda\in\sigma_E(T)\}.$$

Recall the Beurling-Gelfand formula $r_E(T) = \lim_{n\to\infty} \|T^n\|^{1/n}$. We also note that if $T : E \to E$ is compact, then $\sigma_E(T) \setminus \{0\} = \sigma_{p,E}(T) \setminus \{0\}$.

Lemma 5.1. *Let $T : E \to E$ be a linear compact operator on a Hilbert space E equipped with an inner product $(\cdot,\cdot)_E$ and the associated norm $\|\cdot\|_E$. Let $F \subsetneq E$ be a dense subspace of E. Assume that F is a Banach space equipped with the norm $\|\cdot\|_F$ and that $T : F \to F$ is also compact. Then $r_E(T) = r_F(T)$.*

Proof. First thanks to $F \subset E$, we claim that $r_E(T) \geq r_F(T)$. In fact, if $r_F(T) = 0$, then this holds trivially. In addition, if $r_E(T) = 0$, then $r_F(T) = 0$. Indeed, if assume otherwise that $r_F(T) > 0$, then since $\sigma_E(T) \setminus \{0\} = \sigma_{p,E}(T) \setminus \{0\}$ due to the compactness of T, there exists $\lambda \neq 0$ and $v \in F \setminus \{0\}$ such that $(\lambda I - T)v = 0$. Since $F \subset E$, we have $v \in E$ and hence $\lambda \in \sigma_E(T)$. This contradicts with the assumption that $r_E(T) = 0$ and thus proves $r_F(T) = 0$. Finally, assume that both $r_E(T)$ and $r_F(T)$ are positive. We claim that $\sigma_F(T) \setminus \{0\} \subset \sigma_E(T) \setminus \{0\}$ which immediately implies that $r_E(T) \geq r_F(T)$. In fact, for any $\lambda \in \sigma_F(T) \setminus \{0\} = \sigma_{p,F}(T) \setminus \{0\}$, there exists $v \in F \setminus \{0\}$ such that $(\lambda I - T)v = 0$. Since $F \subset E$, we have $\lambda \in \sigma_E(T) \setminus \{0\}$.

Next we show that $r_E(T) = r_F(T)$. Assume otherwise that $r_E(T) > r_F(T)$. By the definition of $r_E(T)$ and the assumption that $T : E \to E$ is a compact operator, there exists an eigenvalue $\lambda \in \rho_{p,E}(T)$ such that $r_E(T) \geq |\lambda| > r_F(T)$. This implies that $\mathrm{Ker}_F(\lambda I - T) \coloneqq \{u \in F : (\lambda I - T)u = 0\} = \{0\}$, i.e., $\mathrm{Ker}_E(\lambda I - T) \cap F = \{0\}$. Since F is a dense subset of E, this implies that $\mathrm{Ker}_E(\lambda I - T) \subset \bar{F}^{\perp} = E^{\perp} = \{0\}$. This contradicts with the fact that λ is an eigenvalue of T on E and completes the proof of the lemma. □

5.2. **Krein-Rutman theorem and the leading eigenvalue.** In this section, we recall the famous Krein-Rutman theorem [31] on the leading eigenvalue and eigenfunction of positive operators on ordered Banach spaces. To this end, let us first recall some terminologies on ordered Banach spaces. Given a Banach space E, a closed convex subset $K \subset E$ is called a *cone* on E if $\alpha K \subset K$ for every $\alpha > 0$ and $K \cap \{-K\} = \{0\}$. A cone K induces a natural partial ordering $\leq$ on the Banach space E: $x \leq y$ if and only if $y - x \in K$. Therefore a Banach space E with a cone K is called an *ordered Banach space*, denoted by (E, K). If the cone K satisfies that $\overline{K - K} = E$, then the cone K is called a *total* cone. We define $\dot{K} = K \setminus \{0\}$ and denote by K° the interior of K. If K has nonempty interior K°, then K is called a *solid* cone. It is not hard to see that a solid cone is total.

Example 5.1. Consider the Banach space $C(\overline{\Omega})$ of continuous functions on a bounded domain $\Omega \subset \mathbb{R}^d$. The space $C(\overline{\Omega})$ is an ordered Banach space with cone $C_+(\overline{\Omega})$ consisting of nonnegative functions in $C(\overline{\Omega})$. This cone is solid since any strictly positive function is an interior point.

Consider two ordered Banach spaces E and F, with cones P and Q respectively. A linear operator $T : E \to F$ is called *positive* if $T(P) \subset Q$, and *strictly positive* if $T(\dot{P}) \subset \dot{Q}$. If in addition Q is solid, then T is called *strongly positive* if $T(\dot{P}) \subset Q^\circ$.

Theorem 5.2 (Krein-Rutman)**.** *Let E be an ordered Banach space with a total cone $K \subset E$. Let $T : E \to E$ be a linear compact positive operator with spectral radius $r(T) > 0$. Then $r(T)$ is an eigenvalue of T and of the dual T^* with corresponding eigenvectors $u \in K \setminus \{0\}$ and $u^* \in K^* \setminus \{0\}$.*

As an important consequence of the Krein-Rutman theorem, Theorem 5.3 establishes the simplicity of the leading eigenvalue of a strongly positive compact operator on an ordered Banach space.

Theorem 5.3 ([1, Theorem 3.2]). *Let E be an ordered Banach space with a solid cone K. Let $T : E \to E$ be a strongly positive compact operator. Then*

(i) The spectral radius $r(T) > 0$;

(ii) $r(T)$ is a simple eigenvalue with an eigenvector $u \in K^\circ$ and there is no other eigenvalue with a positive eigenvector.

5.3. **Regularity of ground state.** Consider the Schrödinger operator $\mathcal{H} = -\Delta + V$. Since $\inf_x V(x) \geq V_{\min} > 0$, the standard Sobolev estimate on Lipschitz domains (see e.g. [21, Chapter 4]) implies that the inverse of $\mathcal{H}$ (with respect to the Neumann boundary condition), denoted by $\mathcal{S} := \mathcal{H}^{-1}$, is bounded from $L^2(\Omega)$ to $H^2(\Omega)$ and hence compact on $L^2(\Omega)$. Moreover, $\mathcal{S}$ has countable many eigenvalues $\{\mu_j\}_{j=0}^\infty$ with $\mu_j \downarrow 0$ as $j \to \infty$ and with $r(\mathcal{S}) = \mu_0 = \frac{1}{\lambda_0}$.

Recall the spectral Barron space $\mathcal{B}^s(\Omega)$ defined in (2.4). We also recall from [33] the next important lemma which shows that the operator $\mathcal{S} : \mathcal{B}^s(\Omega) \to \mathcal{B}^{s+2}(\Omega)$ is bounded.

Lemma 5.4 ([33, Theorem 5]). *Assume that $V \in \mathcal{B}^s(\Omega)$ with $s \geq 0$ and $\inf_{x\in\Omega} \geq V_{\min} > 0$. Then the operator $\mathcal{S} : \mathcal{B}^s(\Omega) \to \mathcal{B}^{s+2}(\Omega)$ is bounded.*

The estimates in Barron spaces of Lemma 5.4 are different from the standard proofs for the well-posedness and the regularity estimates in Sobolev spaces. In fact, the lack of Hilbert structure of the spectral Barron space prevents us using the Lax-Milgram theory to obtain the existence and uniqueness. Instead, we rewrite the stationary Schrödinger equation with a source term as an equivalent Fredholm integral equation of the second kind for the cosine coefficients of the solution in the weighted space $\ell^1_{W_s}(\mathbb{N}_0^d)$. Thanks to the celebrated theorem of Fredholm alternative, the existence and stability estimate of the solution then follow from uniqueness where the latter holds as a result of the standard energy estimate. A detailed proof of Lemma 5.4 can be found in [33, Appendix D.2].

Notice that the inclusion $\mathcal{I} : \mathcal{B}^{s+2}(\Omega) \hookrightarrow \mathcal{B}^s(\Omega)$ is compact. In fact, by definition the space $\mathcal{B}^s(\Omega)$ can be viewed as a weighted ℓ^1 space $\ell^1_{W_s}(\mathbb{N}_0^d)$ of the cosine coefficients defined on the lattice $\mathbb{N}_0^d$ with the weight $W_s(k) = (1 + \pi^s |k|_1^s)$. Therefore the inclusion satisfies that

$$\|\mathcal{I}u\|_{\mathcal{B}^s(\Omega)} = \sum_{k\in\mathbb{N}_0^d} W_s(k)|\hat{u}(k)| = \sum_{k\in\mathbb{N}_0^d} \frac{W_s(k)}{W_{s+2}(k)} W_{s+2}(k)|\hat{u}(k)|.$$

Since $\frac{W_s(k)}{W_{s+2}(k)} \to 0$ as $|k| \to \infty$, by a similar argument as used in the proof of [33, Lemma 7.2], one can conclude that $\mathcal{I}$ is compact from $\ell^1_{W_{s+2}}(\mathbb{N}_0^d)$ to $\ell^1_{W_s}(\mathbb{N}_0^d)$ and hence from $\mathcal{B}^{s+2}(\Omega)$ to $\mathcal{B}^s(\Omega)$. Corollary 5.5 is then a direct consequence of Lemma 5.4 and compactness of the inclusion $\mathcal{I}$ from $\mathcal{B}^{s+2}(\Omega)$ to $\mathcal{B}^s(\Omega)$.

Corollary 5.5. *Under the same assumption of Lemma 5.4, the operator $\mathcal{S} : \mathcal{B}^s(\Omega) \to \mathcal{B}^s(\Omega)$ is compact.*

Note that by definition, we have $\mathcal{B}^s(\Omega) \hookrightarrow C(\overline{\Omega})$. Let us define the cone $\mathcal{B}^s_+(\Omega)$ of $\mathcal{B}^s(\Omega)$ by setting

$$\mathcal{B}^s_+(\Omega) \coloneqq \big\{f \in \mathcal{B}^s(\Omega) \,:\, f \geq 0\big\}.$$

Lemma 5.6. *The operator $\mathcal{S} : \mathcal{B}^s(\Omega) \to \mathcal{B}^s(\Omega)$ is strongly positive, i.e. for any non-zero $f \in \mathcal{B}^s_+(\Omega)$, we have that $\mathcal{S}f(x) > 0$ for all $x \in \overline{\Omega}$.*

Proof. Let $f \in \mathcal{B}^s_+(\Omega) \subset C(\overline{\Omega})$. For any fixed $t > 0$ and $x \in \Omega$, it follows by the Lie-Trotter splitting that

$$e^{-t\mathcal{H}} f(x) = \lim_{n\to\infty} [e^{-\frac{t}{n}V} e^{\frac{t}{n}\Delta}]^n f(x).$$

Since $0 < V_{\min} \leq V \leq V_{\max} < \infty$, we have $e^{-sV_{\max}} g \leq e^{-sV} g \leq e^{-sV_{\min}} g$ for any non-negative $g \in L^\infty(\Omega)$. Thanks to the fact that the heat semigroup $e^{t\Delta}$ is positivity-preserving, this implies that

$$e^{-tV_{\max}} e^{t\Delta} f(x) \leq e^{-t\mathcal{H}} f(x) \leq e^{-tV_{\min}} e^{t\Delta} f(x). \tag{5.1}$$

Moreover, as a result of the semigroup property, the solution operator $\mathcal{S}$ can be written as

$$\mathcal{S}f(x) = \int_0^\infty e^{-t\mathcal{H}} f(x) dt.$$

Note that owing to the upper bound of (5.1) the integral is finite. It follows from the last identity and the lower bound of (5.1) that

$$\mathcal{S}f(x) \geq \int_0^\infty e^{-tV_{\max}} e^{t\Delta} f(x) dt \geq \int_1^2 e^{-tV_{\max}} e^{t\Delta} f(x) dt.$$

Now since f is non-negative, continuous and non-zero on Ω, there exists a set $A \subset \Omega$ with $\mathrm{Leb}(A) > 0$ and a constant $c > 0$ such that $f \geq c > 0$. Thanks to the Gaussian lower bound of the Neumann heat kernel estimate (see e.g. [6, Theorem 3.4]), there exist positive constants c_1 and c_2 such that

$$\begin{aligned} e^{t\Delta} f(x) &= \int_\Omega p_t(x,y) f(y) dy \\ &\geq \int_\Omega \frac{c_1}{t^{\frac{d}{2}}} e^{-\frac{|x-y|^2}{c_2 t}} f(y) dy \\ &\geq \frac{c\, c_1}{t^{\frac{d}{2}}} e^{-\frac{\mathrm{diam}(\Omega)}{c_2 t}} \mathrm{Leb}(A) > 0. \end{aligned}$$

Multiplying above with $e^{-tV_{\max}}$ and then integrating on $[1,2]$ with respect to t yields that

$$\mathcal{S}f(x) \geq c\, c_1 \mathrm{Leb}(A) \int_1^2 \frac{1}{t^{\frac{d}{2}}} e^{-\frac{\mathrm{diam}(\Omega)}{c_2 t}} dt > 0. \qquad \square$$

Now with the above preparations we are ready to present the proof of Theorem 2.5.

Proof of Theorem 2.5. It is clear that the ground state u^* of $\mathcal{H}$ is identical to the eigenfunction that corresponds to the spectral radius $r(\mathcal{S}) = 1/\lambda^*$ of the inverse operator $\mathcal{S} = \mathcal{H}^{-1}$. In order to show that $u^* \in \mathcal{B}^{s+2}(\Omega)$, it suffices to show that $u^* \in \mathcal{B}^s(\Omega)$. In fact, since (u^*, λ^*) solves the Neumann eigenvalue problem

$$\mathcal{H}u^* = -\Delta u^* + Vu^* = \lambda^* u^*,$$

we have $u^* = \lambda^* \mathcal{S} u^*$. An application of Lemma 5.4 implies that $u^* \in \mathcal{B}^{s+2}(\Omega)$ if and only if $u^* \in \mathcal{B}^s(\Omega)$. To show the latter, let us consider the operator $\mathcal{S}$ defined on the ordered Banach space $\mathcal{B}^s(\Omega)$ with the solid cone $\mathcal{B}^s_+(\Omega)$. Observe that $\mathcal{S} : L^2(\Omega) \to L^2(\Omega)$ is compact and that by Corollary 5.5 $\mathcal{S} : \mathcal{B}^s(\Omega) \to \mathcal{B}^s(\Omega)$ is also compact. Therefore by Lemma 5.1 the spectral radii of $\mathcal{S}$ are identical when viewed as operators on $\mathcal{B}^s(\Omega)$ and $L^2(\Omega)$ respectively. It follows from Theorem 5.3 and the strongly positivity of $\mathcal{S}$ established in Lemma 5.6 that there exists a unique (up to a multiplicative constant) eigenfunction $u^* \in \mathcal{B}^s(\Omega)$ corresponding to the spectral radius $r(\mathcal{S}) = 1/\lambda^*$. Moreover, u^* is strictly positive on $\overline{\Omega}$. This completes the proof. □

REFERENCES

[1] Herbert Amann, *Fixed point equations and nonlinear eigenvalue problems in ordered Banach spaces*, SIAM Rev. **18** (1976), no. 4, 620–709, DOI 10.1137/1018114. MR415432

[2] Ben Andrews and Julie Clutterbuck, *Proof of the fundamental gap conjecture*, J. Amer. Math. Soc. **24** (2011), no. 3, 899–916, DOI 10.1090/S0894-0347-2011-00699-1. MR2784332

[3] Ben Andrews, Julie Clutterbuck, and Daniel Hauer, *The fundamental gap for a one-dimensional Schrödinger operator with Robin boundary conditions*, Proc. Amer. Math. Soc. **149** (2021), no. 4, 1481–1493, DOI 10.1090/proc/15140. MR4242306

[4] Francis Bach, *Breaking the curse of dimensionality with convex neutral networks*, J. Mach. Learn. Res. **18** (2017), Paper No. 19, 53. MR3634886

[5] Andrew R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inform. Theory **39** (1993), no. 3, 930–945, DOI 10.1109/18.256500. MR1237720

[6] Richard F. Bass and Pei Hsu, *Some potential theory for reflecting Brownian motion in Hölder and Lipschitz domains*, Ann. Probab. **19** (1991), no. 2, 486–508. MR1106272

[7] Cai, Zi and Jinguo Liu, 2018. *Approximating quantum many-body wave functions using artificial neural networks*. Phys. Rev. B **97**, 035116.

[8] Caragea, Andrei, Philipp Petersen, and Felix Voigtlaender, 2020. *Neural network approximation and estimation of classifiers with classification boundary in a barron class*. `arXiv:2011.09363`

[9] Giuseppe Carleo and Matthias Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science **355** (2017), no. 6325, 602–606, DOI 10.1126/science.aag2302. MR3642415

[10] Chen, Fan, Jianguo Huang, Chunmei Wang, and Haizhao Yang, 2020. *Friedrichs learning: weak solutions of partial differential equations via deep learning*. `arXiv:2012.08023`.

[11] Chen, Zhiang, Jianfeng Lu, and Yulong Lu, 2021. *On the representation of solutions to elliptic PDEs in Barron spaces*. Adv. Neural Info. Processing Syst. **34**.

[12] Choo, Kenney, Antonio Mezzacapo, and Giuseppe Carleo. *Fermionic neural-network states for ab-initio electronic structure*. Nat. Commun. **11**, no. 1, 1–7.

[13] R. M. Dudley, *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*, J. Functional Analysis **1** (1967), 290–330, DOI 10.1016/0022-1236(67)90017-1. MR0220340

[14] E. Weinan, Chao Ma, Stephan Wojtowytsch, and Lei Wu, 2020. *Towards a mathematical understanding of neural network-based machine learning: what we know and what we don't*. `arXiv:2009.10713`.

[15] E. Weinan, Chao Ma, and Lei Wu, 2019.. *Barron spaces and the compositional function spaces for neural network models*. `arXiv:1906.08039`.

[16] E. Weinan and Stephan Wojtowytsch, 2020. *Representation formulas and pointwise properties for barron functions*. `arXiv:2006.05982`.

[17] E. Weinan and Stephan Wojtowytsch, 2020. *Some observations on partial differential equations in Barron and multi-layer spaces*, `arXiv:2012.01484`.

[18] Weinan E and Bing Yu, *The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems*, Commun. Math. Stat. **6** (2018), no. 1, 1–12, DOI 10.1007/s40304-018-0127-z. MR3767958

[19] Gao, Xun and Lu-Ming. Duan, 2017. *Efficient representation of quantum many-body states with deep neural networks*, Nat. Commun. **8**, 662, 2017.

[20] James Glimm and Arthur Jaffe, *Quantum physics*, 2nd ed., Springer-Verlag, New York, 1987. A functional integral point of view, DOI 10.1007/978-1-4612-4728-9. MR887102

[21] Pierre Grisvard, *Elliptic problems in nonsmooth domains*, Classics in Applied Mathematics, vol. 69, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Reprint of the 1985 original [MR0775683]; With a foreword by Susanne C. Brenner, DOI 10.1137/1.9781611972030.ch1. MR3396210

[22] Yiqi Gu, Haizhao Yang, and Chao Zhou, *SelectNet: self-paced learning for high-dimensional partial differential equations*, J. Comput. Phys. **441** (2021), Paper No. 110444, 18, DOI 10.1016/j.jcp.2021.110444. MR4265654

[23] Jiequn Han, Arnulf Jentzen, and Weinan E, *Solving high-dimensional partial differential equations using deep learning*, Proc. Natl. Acad. Sci. USA **115** (2018), no. 34, 8505–8510, DOI 10.1073/pnas.1718942115. MR3847747

[24] Jiequn Han, Jianfeng Lu, and Mo Zhou, *Solving high-dimensional eigenvalue problems using deep neural networks: a diffusion Monte Carlo like approach*, J. Comput. Phys. **423** (2020), 109792, 13, DOI 10.1016/j.jcp.2020.109792. MR4156932

[25] Jiequn Han, Linfeng Zhang, and Weinan E, *Solving many-electron Schrödinger equation using deep neural networks*, J. Comput. Phys. **399** (2019), 108929, 8, DOI 10.1016/j.jcp.2019.108929. MR4013818

[26] Hermann, Jan, Zeno Schätzle, and Frank Noé, 2020. Deep-neural-network solution of the electronic Schrödinger equation. Nature Chemistry **12**, no. 10, 891–897.

[27] Hong, Qingguo, Jonathan W Siegel, and Jinchao Xu, 2021. *A priori analysis of stable neural network solutions to numerical PDEs*. `arXiv:2104.02903`.

[28] Kerner, Joachim, 2021. *A lower bound on the spectral gap of one-dimensional Schrödinger operators*. `arXiv:2102.03816`.

[29] Yuehaw Khoo, Jianfeng Lu, and Lexing Ying, *Solving for high-dimensional committor functions using artificial neural networks*, Res. Math. Sci. **6** (2019), no. 1, Paper No. 1, 13, DOI 10.1007/s40687-018-0160-2. MR3887217

[30] Jason M. Klusowski and Andrew R. Barron, *Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ^1 and ℓ^0 controls*, IEEE Trans. Inform. Theory **64** (2018), no. 12, 7649–7656, DOI 10.1109/tit.2018.2874447. MR3883687

[31] M. G. Kreĭn and M. A. Rutman, *Linear operators leaving invariant a cone in a Banach space*, Amer. Math. Soc. Translation **1950** (1950), no. 26, 128. MR0038008

[32] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, 2015. *Deep learning*. Nature, **521**, no. 7553, 436–444.

[33] Lu, Yulong, Jianfeng Lu, and Min Wang, 2021. *A priori generalization analysis of the deep ritz method for solving high dimensional elliptic partial differential equations*. In Conference on Learning Theory, PMLR, pp. 3196–3241.

[34] Luo, Tao and Haizhao Yang, 2020. *Two-layer neural networks for partial differential equations: Optimization and generalization theory*. `arXiv:2006.15733`.

[35] Mishra, Siddhartha and Roberto Molinaro, *Estimates on the generalization error of physics informed neural networks (PINNs) for approximating PDEs*, 2020. `arXiv:2006.16144`.

[36] Pfau, David, James S Spencer, Alexander GDG Matthews, and W Matthew C Foulkes, 2020. *Ab initio solution of the many-electron schrödinger equation with deep neural networks*. Phys. Rev. Res. **2**, no. 3, 033429.

[37] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys. **378** (2019), 686–707, DOI 10.1016/j.jcp.2018.10.045. MR3881695

[38] schmidhuber, Jürgen, 2015. *Deep learning in neural networks: an overview*. Neural Netw. **61** 85–117.

[39] Ohad Shamir and Shai Shalev-Shwartz, *Matrix completion with the trace norm: learning, bounding, and transducing*, J. Mach. Learn. Res. **15** (2014), 3401–3423. MR3277164

[40] Yeonjong Shin, Jérôme Darbon, and George Em Karniadakis, *On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs*, Commun. Comput. Phys. **28** (2020), no. 5, 2042–2074, DOI 10.4208/cicp.oa-2020-0193. MR4188529

[41] Shin, Yeonjong, Zhongqiang Zhang, and George Em Karniadakis. *Error estimates of residual minimization using neural networks for linear PDEs*, `arXiv:2010.08019`, 2020.

[42] Siegel, Jonathan W and Jinchao Xu, 2020. *Approximation rates for neural networks with general activation functions*. Neural Netw.

[43] Jonathan W. Siegel and Jinchao Xu, *High-order approximation rates for shallow neural networks with cosine and* ReLU^k *activation functions*, Appl. Comput. Harmon. Anal. **58** (2022), 1–26, DOI 10.1016/j.acha.2021.12.005. MR4357282

[44] I. M. Singer, Bun Wong, Shing-Tung Yau, and Stephen S.-T. Yau, *An estimate of the gap of the first two eigenvalues in the Schrödinger operator*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **12** (1985), no. 2, 319–333. MR829055

[45] Justin Sirignano and Konstantinos Spiliopoulos, *DGM: a deep learning algorithm for solving partial differential equations*, J. Comput. Phys. **375** (2018), 1339–1364, DOI 10.1016/j.jcp.2018.08.029. MR3874585

[46] Wolf, Michael M. 2020. *Mathematical Foundations of Supervised Learning*, `https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA4801_2020S/ML_notes_main.pdf`. Last visited on 2020/12/5.

[47] Jinchao Xu, *Finite neuron method and convergence analysis*, Commun. Comput. Phys. **28** (2020), no. 5, 1707–1745, DOI 10.4208/cicp.oa-2020-0191. MR4188519

[48] Yaohua Zang, Gang Bao, Xiaojing Ye, and Haomin Zhou, *Weak adversarial networks for high-dimensional partial differential equations*, J. Comput. Phys. **411** (2020), 109409, 14, DOI 10.1016/j.jcp.2020.109409. MR4079373

Departments of Mathematics, Physics, and Chemistry, Duke University, Box 90320, Durham, North Carolina 27708.

Email address: `jianfeng@math.duke.edu`

Department of Mathematics and Statistics, Lederle Graduate Research Tower, University of Massachusetts, 710 N. Pleasant Street, Amherst, Massachusetts 01003.

Email address: `lu@math.umass.edu`