EIGENVALUE PROBLEMS FOR EXPONENTIAL TYPE KERNELS

DIFENG CAI AND PANAYOT S. VASSILEVSKI

ABSTRACT. We study approximations of eigenvalue problems for integral operators associated with kernel functions of exponential type. We show convergence rate $|\lambda_k - \lambda_{k,h}| \leq C_k h^2$ in the case of lowest order approximation for both Galerkin and Nyström methods, where h is the mesh size, λ_k and $\lambda_{k,h}$ are the exact and approximate kth largest eigenvalues, respectively. We prove that the two methods are numerically equivalent in the sense that $|\lambda_{k,h}^{(G)} - \lambda_{k,h}^{(N)}| \leq Ch^2$, where $\lambda_{k,h}^{(G)}$ and $\lambda_{k,h}^{(N)}$ denote the kth largest eigenvalues computed by Galerkin and Nyström methods, respectively, and C is a eigenvalue independent constant. The theoretical results are accompanied by a series of numerical experiments.

1. INTRODUCTION

In this paper we are interested in the eigenvalue problem associated with integral operators $Af := \int_D K(x,y)f(y)dy$ defined from kernel functions $K(x,y)$ of exponential type (cf. (1.1)) and D is a bounded Lipschitz domain in $\mathbb{R}^d$.

Our approach is general, but driven by practical applications we focus on kernels $K(x,y)$, $x, y \in \mathbb{R}^d$ of the following particular (exponential) form

$$K(x,y) = e^{-\rho(x-y)} \quad \text{with} \quad \rho(x) := (|x_1|^s/\omega_1^s + \cdots + |x_d|^s/\omega_d^s)^\gamma\,, \tag{1.1}$$

where $s \in \{1,2\}$, $\gamma = 1$ or $1/s$, $\omega_i (i = 1,\ldots,d) > 0$. Examples of such kernel functions include $e^{-|x-y|^2}, e^{-|x-y|}$, etc. The kernel defines the integral operator

$$Af(x) := \int_D K(x,y)f(y)dy, \quad x \in D. \tag{1.2}$$

Of our main interest is the numerical approximation of the eigenvalue problem associated with A, namely, $A\phi = \lambda\phi$, for some λ, ϕ.

Eigenproblems of the above type, arise frequently in various research areas such as geology [8, 7], uncertainty quantification [3, 10, 19], machine learning [16], etc. The analysis of the underlying eigenvalue problem is beneficial in the derivation of the error control, algorithm design, and overall numerical practice, etc.

Mathematically, the problem is usually formulated in either the space of continuous functions or the space of L^2-integrable functions. The corresponding discretizations are the Nyström method and Galerkin method, respectively.

Date: start: June 12, 2018, today is: March 27, 2019.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-JRNL-755085. The 2nd author was partially supported by NSF under grant DMS-1619640.

For Nyström method, though various error estimates were derived, e.g., in [12, 25, 2, 18, 21, 22], it is not known if they are consistent with actual numerical results, especially when the kernel function is not smooth enough, for example, not continuously differentiable. Additionally, the proofs require the mesh size to be sufficiently small.

For Galerkin method, which generally requires to use certain quadrature rule to evaluate the double integrals when assembling the matrix, the impact of the quadrature error to the computed eigenvalues is of practical interest and needs special investigation especially when the integrand is not sufficiently smooth (for example, functions with unbounded derivatives). This is the case of the kernels of the form (1.1) studied in the present paper.

1.1. **Contributions.** Our aim is to present a comprehensive study of the eigenvalue problems for integral operators associated with kernel functions of exponential type as defined in (1.1). Those kernel functions are not necessarily smooth, i.e., they may not have continuous (partial) derivatives. Theoretically, we focus on the analysis of two formulations of the operator eigenvalue problem in terms of L^2-integrable functions and continuous functions. In the first case we use the Galerkin method whereas in the second case the Nyström method is used. We utilize piecewise constant approximation in the Galerkin method and midpoint rule in the Nyström method. Numerical experiments were conducted to illustrate and sometimes to complement the theoretical results.

The contributions are listed below (see Section 4 for details).

Firstly, we present a new framework to analyze the Nyström discretization. To obtain the Nyström discretization error, we show that it is numerically equivalent to the Galerkin discretization, and thus the error estimate for Galerkin discretization immediately carries over, which reads

$$|\lambda_k - \lambda_{k,h}| \leq C_k h^2,$$

where h is the mesh size, λ_k and $\lambda_{k,h}$ denote the kth largest exact and approximate eigenvalues (counted with multiplicity), respectively. To the best of our knowledge, it is the first result that captures the $O(h^2)$ convergence rate of the Nyström method when the kernel function is not continuously differentiable, while existing results (cf. [12, 25, 2, 18, 21, 22]) can only yield $O(h)$ convergence rate (see Section 5.3). For the nonsmooth kernel function considered in (1.1), for example, when $\rho(x-y) = |x-y|$, the $O(h^2)$ convergence rate is known as *superconvergence* in [5]. Moreover, unlike existing results, our proof does not require the mesh size to be sufficiently small.

Secondly, to the best of our knowledge, we prove for the first time that the Galerkin method and Nyström method are numerically equivalent in the sense that

$$|\lambda_{k,h}^{(G)} - \lambda_{k,h}^{(N)}| \leq Ch^2,$$

where $\lambda_{k,h}^{(G)}$ and $\lambda_{k,h}^{(N)}$ denote the kth largest eigenvalues (counted with multiplicity) computed by Galerkin and Nyström discretizations, respectively, and C is a constant independent of any eigenvalue. The estimate indicates that the convergence

rates for two methods are the same up to a generic constant independent of the eigenvalues. Also, the result guarantees that the error induced by numerical integration in the practical implementation of the Galerkin method does not affect the final convergence rate, and it provides a theoretical foundation for the use of the more implementation-friendly Nyström discretization, while maintaining the same rate of convergence. Numerical results are presented to confirm the claim.

Thirdly, we perform several numerical experiments to examine various theoretical estimates, including the convergence rate, dependence of the asymptotic constant on λ, approximation of eigenfunctions, etc. Our numerical results indicate that the eigenvalue convergence rate is quadratic with respect to the mesh size and for different eigenvalues, the approximation error is roughly independent of the eigenvalue magnitude. A detailed discussion relating our estimates and the ones from [12, 2, 18, 21] is presented.

1.2. **Outline.** The rest of the paper is organized as follows. In Section 2, we study the integral operator associated with kernel functions of exponential type and state the positive (semi-)definiteness of the operator as well as the related matrices. Section 3 presents abstract estimates for the Galerkin approximation to the underlying eigenvalue problem. The main results are presented in Section 4, including convergence rates of Galerkin and Nyström discretizations, the equivalence between the two discretizations, etc. Section 5 provides a numerical study of various theoretical results in Section 4 and in existing literature [12, 2, 18, 21]. The proof for the positive (semi-)definiteness of the operator and the related matrices is given in the appendix (Section 7).

2. Integral Operators with Kernel Functions of Exponential Type

For notational convenience, for any given bounded Lipschitz domain in $\mathbb{R}^d$, when working with Sobolev spaces $L^2(D), H^1(D) := \{f \in L^2(D) : \nabla f \in L^2(D)^d\}$, etc., we assume D is open; while for $C(D)$ - the Banach space of continuous functions with the supremum norm $\|f\|_{\sup} := \sup_{x\in D} |f(x)|$, D is assumed to be closed. In this section, unless otherwise stated, we use $\|\cdot\|$ without subscript to denote the usual L^2 norm.

2.1. **Some auxiliary estimates.** The following result is immediate using straightforward calculation.

Proposition 2.1. *The kernel function $K(x, y)$ defined in* (1.1) *satisfies*

$$|K(x, y) - K(x, y')| \le C_K|y - y'| \quad \textit{with} \quad C_K = \left(\sum_{i=1}^{d} \omega_i^{-2}\right)^{1/2}.$$

Consequently, if A is the integral operator defined in (1.2)*, then for each $f \in L^1(D)$, Af is Lipschitz continuous over $\mathbb{R}^d$ with $|Af(x) - Af(x')| \le C_K\|f\|_{L^1(D)}|x - x'|$. In particular, $Af \in H^1(D)$ and $\|\nabla Af\| \le \|\nabla_x K\|_{L^2(D\times D)}\|f\|$, $\quad \forall\, f \in L^2(D)$.*

Next we estimate the second derivatives of the kernels of our interest, which is needed in the error analysis that we provide later on. Let $K(x, y) = e^{-\rho(x-y)}$ with

$\rho = \left(\sum_{i=1}^d x_i^2/\omega_i^2\right)^{1/2}$. A direct calculation shows that the second order partial derivatives of K are unbounded at $x = y$. More specifically, we have

$$\frac{\partial^2}{\partial x_i^2}K(x,y) = -\frac{\rho^2(x-y) - (x_i-y_i)^2/\omega_i^2}{\omega_i^2\rho^3(x-y)}K(x,y) + \frac{(x_i-y_i)^2}{\omega_i^4\rho^2(x-y)}K(x,y),$$
$$\frac{\partial^2}{\partial x_j\partial x_i}K(x,y) = \frac{(x_i-y_i)(x_j-y_j)}{\omega_i^2\omega_j^2\rho^3(x-y)}K(x,y) + \frac{(x_i-y_i)(x_j-y_j)}{\omega_i^2\omega_j^2\rho^2(x-y)}K(x,y), \quad i \neq j,$$

and

$$|\partial^\alpha K(x,y)| \leq C\max\{1, 1/\rho(x-y)\}, \quad |\alpha| = 2, \tag{2.1}$$

where α is a multi-index and C is a generic constant depending only on ω_i.

2.2. **Mapping properties.** The following well-known mapping properties of integral operators associated with continuous kernel functions are collected below (e.g., [14]).

Proposition 2.2. *Let A be defined in* (1.2). *Then* (1). *$A : L^2(D) \to L^2(D)$ is compact;* (2). *$A : C(D) \to C(D)$ is compact;* (3). *$A : L^2(D) \to C(D)$ is compact.*

The above proposition ensures that the theoretical results presented in Section 3 apply to our particular case of kernels of exponential type.

2.3. **Positive definiteness.** Define $\Phi(x) = e^{-\rho(x)}$ with ρ in (1.1). Then $K(x,y) = \Phi(x-y)$. The main result in Theorem 2.1 asserts that the function $\Phi(x)$ is positive definite in the sense below (cf. [26]).

Definition 2.1 (positive definite functions)**.** *A continuous function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is called* positive (semi-)definite *if for any n distinct points $x_1, \ldots, x_n \in \mathbb{R}^d$ ($n = 1, 2, \ldots$), the matrix $a_{i,j} = \Phi(x_i - x_j)$ is* positive (semi-)definite.

For bounded continuous functions, the positive semi-definiteness is equivalent to that of the associated integral operator (cf. [26]).

Proposition 2.3. *A bounded continuous function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is positive semi-definite if and only if $\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Phi(x-y)v(x)v(y)dxdy \geq 0$ for all functions v in the Schwartz space*

$$\left\{ v \in C^\infty(\mathbb{R}^d) : \sup_{x\in\mathbb{R}^d}(1+|v|)^M \sum_{|\alpha|\leq m} |\partial^\alpha v(x)| < \infty \text{ for any integers } m, M \geq 0 \right\}.$$

Theorem 2.1. *For $\omega_i > 0$ ($i = 1, \ldots, d$) and $x \in \mathbb{R}^d$, let ρ take one of the following forms:* (1). $\rho(x) = \sum_{i=1}^d |x_i|/\omega_i$; (2). $\rho(x) = \sum_{i=1}^d x_i^2/\omega_i^2$; (3). $\rho(x) = \left(\sum_{i=1}^d x_i^2/\omega_i^2\right)^{1/2}$. *Then $\Phi(x) = e^{-\rho(x)}$ is positive definite. Namely, for any distinct points $x_1, \ldots, x_n \in \mathbb{R}^d$, the matrix $a_{i,j} = \Phi(x_i - x_j)$ is positive definite.*

The proof of Theorem 2.1 is given in the appendix (see Section 7). The result below follows immediately from Proposition 2.3 and Theorem 2.1.

Corollary 2.1. *Let $K(x,y)$ be the kernel function defined in* (1.1) *and A be the corresponding integral operator defined in* (1.2). *Then $(Av,v) \geq 0$ for all $v \in L^2(D)$.*

3. Abstract Results

In this section, $\mathcal{V}$ will be assumed to be a complex Hilbert space with inner product denoted by $(\cdot,\cdot)$. The theoretical results in this section are developed for the Galerkin discretization. We use boldface symbols to denote matrices. The norm on $\mathcal{V}$ is denoted by $\|\cdot\|$ and $\|\boldsymbol{M}\|_2$ denotes the L^2 operator norm for a matrix $\boldsymbol{M}$. We use Ker and Ran to denote the kernel(or nullspace) and the range of an operator, respectively.

We use A to denote a positive compact operator on $\mathcal{V}$, where the definition of a *positive operator* (cf. [20, 15]) is given below.

Definition 3.1. *An operator A on a Hilbert space $\mathcal{V}$ is called* positive *if $(Av,v) \geq 0$, $\forall\, v \in \mathcal{V}$.*

Note that a positive operator is necessarily self-adjoint, i.e., $A = A^*$ (cf. [20, Theorem 12.32]).

A crucial tool we use in the estimate of eigenvalues is the Courant-Fischer min-max (or max-min) principle (cf. [15]).

Theorem 3.1 (min-max principle)**.** *Let A be a compact, self-adjoint operator on $\mathcal{V}$ with nonnegative eigenvalues listed in decreasing order* (counted with multiplicity)*: $\lambda_1 \geq \cdots \geq \lambda_k \geq \cdots \geq 0$. Then*

$$\lambda_k = \max_{S_k} \min_{\substack{v \in S_k \\ \|v\|=1}} (Av,v),$$

where S_k is any linear subspace of $\mathcal{V}$ of dimension k.

In this paper, we are interested in positive eigenvalues of A, and the eigenvalue problem is to find (λ,ϕ) such that

$$\lambda \in \mathbb{R}_+, \quad \phi \in \mathcal{V}\backslash\{0\}, \quad A\phi = \lambda\phi. \tag{3.1}$$

Let V_h be a finite dimensional subspace of $\mathcal{V}$. The Galerkin method for (3.1) is to find (λ_h,ϕ_h) such that

$$\lambda_h \in \mathbb{R}_+, \quad \phi_h \in V_h\backslash\{0\}, \quad (A\phi_h, v_h) = \lambda_h(\phi_h, v_h), \quad \forall\, v_h \in V_h. \tag{3.2}$$

Let $P_h : \mathcal{V} \to V_h$ be the projection from $\mathcal{V}$ onto V_h. Then (3.2) is essentially the eigenvalue problem of the operator P_hAP_h on $\mathcal{V}$. Since A is a positive compact operator, so is P_hAP_h. We can then list the eigenvalues of the Galerkin approximation in (3.2) in decreasing order (counted with multiplicity): $\lambda_{1,h} \geq \cdots \geq \lambda_{n,h}$. Theorem 3.1 implies that the eigenvalues have the following characterization:

$$\lambda_{k,h} = \max_{S_k} \min_{\substack{v \in S_k \\ \|v\|=1}} (P_hAP_hv,v) = \max_{S_{k,h}} \min_{\substack{v \in S_{k,h} \\ \|v\|=1}} (P_hAP_hv,v) = \max_{S_{k,h}} \min_{\substack{v \in S_{k,h} \\ \|v\|=1}} (Av,v),$$

where S_k and $S_{k,h}$ are k-dimensional subspaces of $\mathcal{V}$ and V_h, respectively. The min-max characterizations of λ_k and $\lambda_{k,h}$ immediately yield the following.

Proposition 3.1. $\lambda_{k,h} \leq \lambda_k$. *Consequently,* $0 \leq \lambda_k - \lambda_{k,h} \leq \lambda_k$.

Eigenvalue approximations with Galerkin methods have been studied over the past few decades (cf. [5, 13]). The following result can be easily derived using the min-max principle (cf. [24]).

Theorem 3.2. *Let $\phi_1, \ldots, \phi_k$ be orthonormal eigenfunctions associated with eigenvalues $\lambda_1, \ldots, \lambda_k$, respectively. Then*

$$|\lambda_k - \lambda_{k,h}| \leq 2 \max_{\substack{v \in \mathcal{W}_k \\ \|v\|=1}} \|(I - P_h)Av\| \|(I - P_h)v\|, \tag{3.3}$$

where $\mathcal{W}_k := \operatorname{span}\{\phi_1\} \oplus \cdots \oplus \operatorname{span}\{\phi_k\}$.

Corollary 3.1. *Under the assumptions in Theorem 3.2,*

$$|\lambda_k - \lambda_{k,h}| \leq 2 \left(\sum_{i=1}^{k} \|(I - P_h)\phi_i\|^2 \right)^{1/2} \left(\sum_{i=1}^{k} \lambda_i^2 \|(I - P_h)\phi_i\|^2 \right)^{1/2}.$$

Proof. By writing $v = \sum_{i=1}^{k} \alpha_i \phi_i \in \mathcal{W}_k$ in (3.3), we can obtain the estimate above via the triangle inequality and the Cauchy-Schwarz inequality. □

The scaling of $|\lambda_k - \lambda_{k,h}|$ and $\|(I - P_h)\phi_k\|$ will be investigated via numerical experiments in Section 5.

Remark 3.1. *Note that if the multiplicity of λ_k is greater than* 1*, then the subspace $\mathcal{W}_k$ may be different for a different choice/ordering of basis functions in Ker$(A - \lambda_k I)$.*

4. Eigenvalue Problems

For the integral operator A defined in (1.2), we present two formulations for its eigenvalue problem based on $\mathcal{V} = L^2(D)$ and $\mathcal{V} = C(D)$, respectively. It can be seen later that the two formulations are actually equivalent. Corresponding the two formulations at the continuous level, two discretizations are discussed, and it is shown later (in Section 4.4) that the two discretizations are also numerically equivalent.

4.1. Two formulations: $\mathcal{V} = L^2(D)$ and $\mathcal{V} = C(D)$.

Recall that A is compact on both $\mathcal{V} = L^2(D)$ and $\mathcal{V} = C(D)$. For the Hilbert space $\mathcal{V} = L^2(D)$, the eigenvalue problem reads:

$$\text{find} \quad (\lambda, \phi) \quad \text{such that} \quad A\phi = \lambda\phi, \quad \phi \in L^2(D)\backslash\{0\}. \tag{4.1}$$

For the Banach space $\mathcal{V} = C(D)$ with supremum norm, the eigenvalue problem reads:

$$\text{find} \quad (\lambda, \phi) \quad \text{such that} \quad A\phi = \lambda\phi, \quad \phi \in C(D)\backslash\{0\}. \tag{4.2}$$

From the mapping property of A in Proposition 2.2, it is easy to see that the two formulations in (4.1) and (4.2) are equivalent in the sense below.

Proposition 4.1. *An eigenpair* (λ, ϕ) *satisfies* (4.1) *if and only if it satisfies* (4.2).

In addition to being compact on $L^2(D)$, it was shown in Corollary 2.1 that A is a positive operator on $L^2(D)$. Therefore, we know from Proposition 4.1 and the spectral theory of self-adjoint compact operators that:

Proposition 4.2. *The eigenvalues of* A *in* (4.1) *or* (4.2) *are nonnegative and can be listed in decreasing order* (counted with multiplicity)*:*

$$\lambda_1 \geq \cdots \geq \lambda_k \geq \cdots \geq 0 \quad \textit{with} \quad \lim_{k\to+\infty} \lambda_k = 0,$$

where the algebraic multiplicity is equal to geometric multiplicity for any $\lambda_k > 0$.

4.2. **Galerkin discretization for $\mathcal{V} = L^2(D)$.** In this section, we consider the Galerkin discretization of the eigenvalue problem in (4.1). Let $\mathcal{T} = \{\tau_i\}_{i=1}^n$ be a subdivision of D of maximum mesh size $h := \max_{\tau\in\mathcal{T}} \operatorname{diam}(\tau)$, where $\operatorname{diam}(\tau)$ denotes the diameter of τ. Introduce the space of piecewise constant functions

$$V_h := \{v \in L^2(D) : v|_\tau \text{ is a constant, } \forall\, \tau \in \mathcal{T}\}$$

and let the projection $P_h : L^2(D) \to V_h$ be given by

$$P_h f|_\tau = \frac{1}{|\tau|}\int_\tau f dx, \quad \forall\, f \in L^2(D). \tag{4.3}$$

The result below is standard.

Proposition 4.3. *Let* P_h *be the projection defined in* (4.3). *Then*

$$\|(I - P_h)f\| \leq C_P h\|\nabla f\|, \quad \forall\, f \in H^1(D),$$
$$\|(I - P_h)Af\| \leq C_P\|\nabla_x K\|_{L^2(D\times D)} h\|f\|, \quad \forall\, f \in L^2(D),$$

where C_P *comes from the Poincaré constant, depending only on the shape regularity of* $\mathcal{T}$. *In particular, if* (λ, ϕ) *is an eigenpair of* A *with* $\lambda > 0$, *then*

$$\|(I - P_h)\phi\| \leq C_P\|\nabla_x K\|_{L^2(D\times D)}\lambda^{-1} h\|\phi\|.$$

Applying Proposition 4.3 to Corollary 3.1 yields the following estimate of the eigenvalue convergence rate with respect to the mesh size h.

Theorem 4.1. *Assume* P_h *is defined in* (4.3) *and* $A_h = P_h A P_h$. *Let* λ_k *and* $\lambda_{k,h}$ *be the* k^{th} *largest positive eigenvalues of* A *and* A_h (*counted with multiplicity*) , *respectively. Then*

$$|\lambda_k - \lambda_{k,h}| \leq 2C_P^2\|\nabla_x K\|^2_{L^2(D\times D)} C_k h^2, \tag{4.4}$$

where $C_k = \sqrt{k}\left(\sum_{i=1}^k \lambda_i^{-2}\right)^{1/2}$ *and* C_P *is the constant in* Proposition 4.3, *depending only on the shape regularity of* $\mathcal{T}$.

Remark 4.1. *In addition to* (4.4), *an* $O(h^2)$ *error bound can also be found in* [5, Chapter 7], *but it is an asymptotic estimate valid only for small enough mesh size* h. [13, Section 18] *provides a non-asymptotic estimate but will result in an* $O(h)$ *error bound.*

4.2.1. *The matrix eigenvalue problem.* Given a subdivision $\mathcal{T} = \{\tau_i\}_{i=1}^n$, let $\chi_{\tau_i}(x)$ denote the characteristic function on τ_i. The Galerkin method seeks a nonzero function $\phi_h(x) = \sum_{i=1}^n c_i \chi_{\tau_i}(x) \in V_h$ such that

$$(A\phi_h, v_h) = (\lambda_h^{(G)} \phi_h, v_h), \quad \forall\, v_h \in V_h$$

for some $\lambda_h^{(G)} > 0$. This yields the matrix eigenvalue problem below:

$$\boldsymbol{M}_h^{(G)} \boldsymbol{c} = \lambda_h^{(G)} \boldsymbol{D}_h \boldsymbol{c}, \tag{4.5}$$

where

$$\boldsymbol{M}_h^{(G)} = \Big[\int_{\tau_i}\int_{\tau_j} K(x,y)dydx\Big]_{i,j=1}^n, \quad \boldsymbol{D}_h = \mathrm{diag}(|\tau_1|, \ldots, |\tau_n|), \quad \boldsymbol{c} = [c_1, \ldots, c_n]^T.$$

Here diag(...) denotes a diagonal matrix with diagonal entries (...). By introducing $\boldsymbol{q} = \boldsymbol{D}_h^{1/2}\boldsymbol{c}$ and multiplying both sides of (4.5) by $\boldsymbol{D}_h^{-1/2}$ on the left, we convert (4.5) into a standard eigenvalue problem below.

$$\boldsymbol{A}_h^{(G)} \boldsymbol{q} = \lambda_h^{(G)} \boldsymbol{q} \quad \text{with} \quad \boldsymbol{A}_h^{(G)} = \boldsymbol{D}_h^{-1/2} \boldsymbol{M}_h^{(G)} \boldsymbol{D}_h^{-1/2}.$$

In practice, for the ease of implementation, we use certain quadrature rule to compute the double integral of the kernel function $K(x,y)$. For example, since $K(x,y) \in C(\mathbb{R}^d \times \mathbb{R}^d)$, we may simply use the mid-point rule to evaluate the integral on each element, i.e.,

$$\int_{\tau_i}\int_{\tau_j} K(x,y)dydx \approx K(x_i, x_j)|\tau_i||\tau_j|, \tag{4.6}$$

where x_i is the centroid of τ_i. The resulting linear system of $c_1, \ldots, c_n$ reads

$$\boldsymbol{M}_h^{(N)} \boldsymbol{c} = \lambda_h^{(N)} \boldsymbol{D}_h \boldsymbol{c} \quad \text{with} \quad \boldsymbol{M}_h^{(N)} = \Big[K(x_i, x_j)|\tau_i||\tau_j|\Big]_{i,j=1}^n. \tag{4.7}$$

Again, using $\boldsymbol{q} = \boldsymbol{D}_h^{1/2}\boldsymbol{c}$, (4.7) can be transformed into

$$\boldsymbol{A}_h^{(N)} \boldsymbol{q} = \lambda_h^{(N)} \boldsymbol{q} \quad \text{with} \quad \boldsymbol{A}_h^{(N)} = \boldsymbol{D}_h^{-1/2} \boldsymbol{M}_h^{(N)} \boldsymbol{D}_h^{-1/2}. \tag{4.8}$$

The approximation error introduced by the quadrature in (4.6) will be analyzed in Section 4.4.

4.3. **Nyström discretization for $\mathcal{V} = C(D)$.** In this section, we consider Nyström discretization for the eigenvalue problem in (4.2) with $\mathcal{V} = C(D)$. Based on the mid-point rule applied to the integral in (1.2), we define the finite-rank operator:

$$A_h\phi(x) := \sum_{j=1}^n K(x, x_j)|\tau_j|\phi(x_j), \quad x \in D,\ \phi \in C(D), \tag{4.9}$$

where x_j is the centroid of τ_j. The eigenvalue problem for A_h is to find $\lambda_h^{(N)}$ and $\phi_h \in C(D)\backslash\{0\}$ such that $A_h\phi_h = \lambda_h^{(N)}\phi_h$. Evaluating the equation at quadrature nodes $x_1, \ldots, x_n$ yields the following equivalent matrix eigenvalue problem (cf. [17, 2]): $\boldsymbol{D}_h^{-1}\boldsymbol{M}_h^{(N)}\boldsymbol{\phi}_h = \lambda_h^{(N)}\boldsymbol{\phi}_h$, which is identical to (4.7). The eigenfunction can then be recovered from nodal values $\boldsymbol{\phi}_h = (\phi_h(x_1), \ldots, \phi_h(x_n))$ by using (4.9).

The substitution $\boldsymbol{q} = \boldsymbol{D}_h^{1/2}\boldsymbol{\phi}_h$ transforms the above matrix problem into a standard symmetric eigenvalue problem identical to (4.8): $\boldsymbol{A}_h^{(N)}\boldsymbol{q} = \lambda_h^{(N)}\boldsymbol{q}$. Therefore, we see that, the Galerkin method coincides with the Nyström method up to quadrature errors from (4.6). We will show in Section 4.4 (see Theorem 4.3) that the quadrature error does not dominate the discretization error in the eigenvalue computation. Therefore, the convergence result for the Nyström method below can be obtained with the help of Theorem 4.1 for the Galerkin method.

Theorem 4.2. *Let $K(x,y) = e^{-(|x_1-y_1|^2/\omega_1^2+\cdots+|x_d-y_d|^2/\omega_d^2)^\gamma}$ ($\gamma = 1/2$ or 1) and $\mathcal{T}$ be a quasi-uniform subdivision in $D \subset \mathbb{R}^d$ ($d = 1,2,3$) with maximum mesh size h. With quadrature approximation A_h defined in (4.9), let λ_k and $\lambda_{k,h}$ denote the k^{th} largest positive eigenvalues of A and A_h (counted with multiplicity) , respectively. Then*

$$|\lambda_k - \lambda_{k,h}| \le C\|\nabla_x K\|^2_{L^2(D\times D)} C_k h^2, \tag{4.10}$$

where C is a constant that only depends on the shape parameter of the mesh $\mathcal{T}$ and C_k is the constant defined in Theorem 4.1.

4.4. **Equivalence of Galerkin and Nyström discretizations.** We have shown in Proposition 4.1 that at the continuous level the two formulations in (4.1) and (4.2) are equivalent. In this section, we build the discrete counterpart of such an equivalence. Namely, we estimate the error in computed eigenvalues from two discretizations discussed in Section 4.2 and Section 4.3. The main result is stated below.

Theorem 4.3. *Let $K(x,y) = e^{-(|x_1-y_1|^2/\omega_1^2+\cdots+|x_d-y_d|^2/\omega_d^2)^\gamma}$ ($\gamma = 1/2$ or 1) and $\mathcal{T} = \{\tau_i\}_{i=1}^n$ be a quasi-uniform subdivision in $D \subset \mathbb{R}^d$ ($d = 1,2,3$) with maximum mesh size h. If $\lambda_h^{(G)}$ and $\lambda_h^{(N)}$ are the i^{th} largest eigenvalues (counted with multiplicity) of $\boldsymbol{A}_h^{(G)}$ and $\boldsymbol{A}_h^{(N)}$, respectively. Then*

$$\left|\lambda_h^{(G)} - \lambda_h^{(N)}\right| \le Ch^2,$$

where the constant C is independent of any eigenvalue.

To prove Theorem 4.3, we first analyze the quadrature error in (4.6).

Lemma 4.1. *Let $\mathcal{T} = \{\tau_i\}_{i=1}^n$ be a quasi-uniform mesh with maximum mesh size h.*

- *If $K(x,y) \in C^2(\tau_i \times \tau_j)$, then*

$$\left|\int_{\tau_i}\int_{\tau_j} K(x,y)dydx - K(x^*,y^*)|\tau_i||\tau_j|\right| \le C_1|\tau_i||\tau_j|h^2 \max_{|\alpha|=2}\max_{\tau_i\times\tau_j}|\partial^\alpha K|. \tag{4.11}$$

- *If $K(x,y) \in C(D\times D)$ is Lipschitz continuous, then*

$$\left|\int_{\tau_i}\int_{\tau_j} K(x,y)dydx - K(x^*,y^*)|\tau_i||\tau_j|\right| \le C_2|\tau_i||\tau_j|h. \tag{4.12}$$

Here α is a multi-index, C_1, C_2 are generic constants independent of i, j, x^* and y^* are centroids of τ_i and τ_j, i.e., $|\tau_i|x^* = \int_{\tau_i} x dx$ and $|\tau_j|y^* = \int_{\tau_i} y dy$.

Proof. The Taylor expansion of $K(x, y)$ over $\tau_i \times \tau_j$ reads

$$K(x,y) = K(x^*,y^*)+\nabla_x K(x^*,y^*)\cdot(x-x^*)+\nabla_y K(x^*,y^*)\cdot(y-y^*)+R(x,y), \tag{4.13}$$

where the remainder satisfies

$$|R(x,y)| \le C_1 h^2 \max_{|\alpha|=2} \max_{(x,y)\in\tau_i\times\tau_j} |\partial^\alpha K(x,y)|.$$

Since x^*, y^* are centroids, it follows that

$$\int_{\tau_i} \nabla_x K(x^*,y^*)\cdot(x-x^*)dx = 0 \quad \text{and} \quad \int_{\tau_j} \nabla_y K(x^*,y^*)\cdot(y-y^*)dy = 0.$$

Hence (4.11) can be obtained by taking double integrals of the equation in (4.13) over $\tau_i \times \tau_j$. (4.12) can be proved similarly by integrating

$$K(x,y) = K(x^*,y^*) + [(K(x,y) - K(x^*,y)) + (K(x^*,y) - K(x^*,y^*))],$$

where the summands in the bracket are estimated using the Lipschitz condition. □

The error $\|\boldsymbol{A}_h^{(G)} - \boldsymbol{A}_h^{(N)}\|_2$ can be estimated as below.

Proposition 4.4. *Under the assumptions in Theorem 4.3, $\boldsymbol{E}_h := \boldsymbol{A}_h^{(G)} - \boldsymbol{A}_h^{(N)}$ satisfies*

$$\|\boldsymbol{E}_h\|_2 = O(h^2), \quad d = 1, 2, 3. \tag{4.14}$$

Proof. Without loss of generality, assume $D = [0,1]^d$. It suffices to prove (4.14) for the following three cases: (1). $d \ge 2$; (2). $d = 1$ and $K(x,y) = e^{-(x-y)^2/\omega^2}$; (3). $d = 1$ and $K(x,y) = e^{-|x-y|/\omega}$.

Case 1. In this case, we illustrate the proof for a uniform rectangular mesh and the same idea applies to the general case. For $\boldsymbol{E}_h = [e_{i,j}]_{i,j}$ with

$$e_{i,j} = |\tau_i|^{-\frac{1}{2}}|\tau_j|^{-\frac{1}{2}} \left(\int_{\tau_i}\int_{\tau_j} K(x,y)dydx - K(x_i,x_j)|\tau_i||\tau_j| \right),$$

we estimate for each fixed i the quantity $\sum_{j=1}^n |e_{i,j}|$. By first partitioning the elements into consecutive layers centered at τ_i, we can evaluate the contribution layer by layer.

The 0th layer is τ_i itself. The 1st layer contains elements that share a vertex with τ_i. In general, the kth($k \ge 1$) layer is composed of elements outside layer $k-1$ that share a vertex with layer $k-1$. See Figure 1 for an illustration in one and two dimensions.

Next we estimate $|e_{i,j}|$ layer by layer. We use C to denote a generic constant independent of i, j. The assumption on $\mathcal{T} = \{\tau_i\}_{i=1}^n$ yields

$$n = O(h^{-d}) \quad \text{and} \quad |\tau| = O(h^d), \quad \forall\, \tau \in \mathcal{T}. \tag{4.15}$$

Note that for each τ_j in layer $k \geq 2$, $K(x,y) \in C^2(\tau_i \times \tau_j)$ and (2.1) implies that

$$\max_{|\alpha|=2} \max_{\tau_i \times \tau_j} |\partial^\alpha K| \leq C(kh)^{-1}, \quad \forall\, \tau_j \text{ in layer } k \geq 2.$$

Together with (4.15) and Lemma 4.1, it can be deduced that

$$|e_{i,j}| \leq Ch^{d+1}, \ \forall\, \tau_j \text{ in layer } 0, 1, \quad \text{and} \quad |e_{i,j}| \leq Ck^{-1}h^{d+1}, \ \forall\, \tau_j \text{ in layer } k \geq 2.$$

The number of elements in layer $k \geq 1$ is $(2k+1)^d - (2k-1)^d \leq 26k^{d-1}$. Therefore,

$$\sum_{j=1}^{n} |e_{i,j}| \leq Ch^{d+1} + C \sum_{\text{layer } k=1}^{L} 26k^{d-1}k^{-1}h^{d+1} = O(h^2), \quad d = 2, 3,$$

where L denotes the maximal number of layers and obviously $L \leq 1/h$.

Since $\boldsymbol{E}_h$ is symmetric, it follows from Gershgorin's Circle Theorem that

$$\|\boldsymbol{E}_h\|_2 = \max_i |\lambda_i(\boldsymbol{E}_h)| \leq \max_i \sum_{j=1}^{n} |e_{i,j}| = O(h^2),$$

which completes the proof of Case 1. The inequality above can also be shown via the following argument: since $\boldsymbol{E}_h$ is symmetric, $\|\boldsymbol{E}_h\|_2$ is equal to its spectral radius, which is bounded by any matrix norm (cf. [11, Theorem 5.6.9]), and the quantity $\max_i \sum_{j=1}^{n} |e_{i,j}|$ is the l_∞ matrix norm of $\boldsymbol{E}_h$.

Case 2. In this case, $K \in C^\infty(D \times D)$ and there exists a constant C such that

$$\max_{|\alpha|=2} \|\partial^\alpha K\|_{L^\infty(D\times D)} \leq C.$$

Hence Lemma 4.1 implies

$$|e_{i,j}| \leq Ch^3, \quad \forall\, i, j.$$

Then the same argument as in Case 1 yields the desired estimate:

$$\|\boldsymbol{E}_h\|_2 \leq \max_i \sum_{j=1}^{n} |e_{i,j}| = O(h^2).$$

Case 3. In this case, $K(x,y) = e^{-\frac{|x-y|}{\omega}}$, $x, y \in [0,1]$. Let h_i denote the length of the ith interval $\tau_i = [t_{i-1}, t_i]$ and recall that x_i is the center of τ_i. We estimate $|e_{i,j}|$ as follows. When $i = j$, it can be computed that

$$\int_{t_{i-1}}^{t_i} \int_{t_{j-1}}^{t_j} e^{-\frac{|x-y|}{\omega}} dydx = \omega(2h_i + 2\omega e^{-\frac{h_i}{\omega}} - 2\omega) = h_i^2 + O(h^3),$$

where the last identity follows from the Taylor expansion:

$$e^{-\frac{h_i}{\omega}} = 1 - \frac{h_i}{\omega} + \frac{h_i^2}{2\omega^2} + O(h^3).$$

Then

$$|e_{i,i}| = h_i^{-1}(h_i^2 + O(h^3) - K(x_i, x_i)|\tau_i|^2) = O(h^2). \tag{4.16}$$

When $i \neq j$, since $K(x,y) = K(y,x)$, assume without loss of generality that $t_{i-1} \geq t_j$. Then $K(x,y) = e^{\frac{y-x}{\omega}}$ for $(x,y) \in \tau_i \times \tau_j$ and we deduce that

$$\begin{aligned}\int_{t_{i-1}}^{t_i}\int_{t_{j-1}}^{t_j} e^{\frac{y-x}{\omega}}dydx &= -\omega^2(e^{\frac{h_i}{\omega}}-1)(e^{-\frac{h_j}{\omega}}-1)e^{\frac{t_j-t_i}{\omega}} \\ &= -\omega^2\left(\frac{h_i}{\omega}+\frac{h_i^2}{2\omega^2}+O(h^3)\right)\left(-\frac{h_j}{\omega}+\frac{h_j^2}{2\omega^2}+O(h^3)\right)e^{\frac{t_j-t_i}{\omega}} \\ &= h_ih_j(1+\frac{h_i-h_j}{2\omega})e^{\frac{t_j-t_i}{\omega}}+O(h^4)\end{aligned}$$

and

$$K(x_i,x_j)|\tau_i||\tau_j| = h_ih_je^{\frac{h_i-h_j}{2\omega}}e^{\frac{t_j-t_i}{\omega}} = h_ih_j(1+\frac{h_i-h_j}{2\omega})e^{\frac{t_j-t_i}{\omega}}+O(h^4).$$

Therefore,

$$|e_{i,j}| = h_i^{-1/2}h_j^{-1/2}O(h^4) = O(h^3), \quad \forall\, i \neq j. \tag{4.17}$$

We conclude from (4.16) and (4.17) that

$$\|\boldsymbol{E}_h\|_2 \leq \max_i \sum_{j=1}^{n} |e_{i,j}| = O(h^2).$$

The proof of the theorem is complete. □

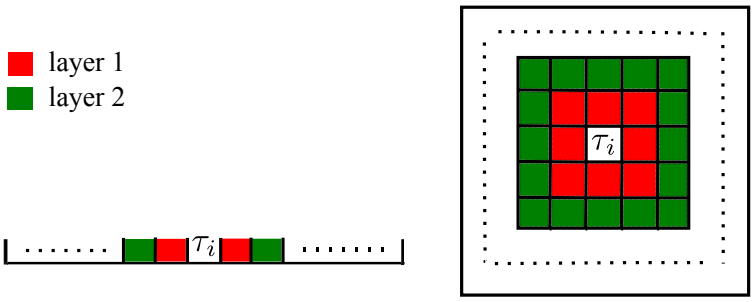

FIGURE 1. Partition of D into layers with respect to τ_i (left: 1D; right: 2D).

Theorem 4.3 follows readily from Proposition 4.4 and Weyl's inequality [27, 4, 23].

Lemma 4.2 (Weyl's inequality)**.** *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be n-by-n Hermitian matrices with eigenvalues $\lambda_1^{(A)} \geq \cdots \geq \lambda_n^{(A)}$ and $\lambda_1^{(B)} \geq \cdots \geq \lambda_n^{(B)}$, respectively. Then*

$$\max_{i=1,\ldots,n}\left|\lambda_i^{(A)} - \lambda_i^{(B)}\right| \leq \|\boldsymbol{A}-\boldsymbol{B}\|_2.$$

4.4.1. *Numerical illustration.* To show that the $O(h^2)$ error bounds in Proposition 4.4 and Theorem 4.3 are attainable, we perform a numerical experiment with $K(x,y) = e^{-|x-y|}$, $x, y \in D = (0,1)$, so the integrals in the Galerkin matrix $\boldsymbol{A}_h^{(G)}$ can be evaluated exactly. Uniform meshes are used and by varying the mesh size h, we compute the corresponding eigenvalue errors measured by $\max_{k\le 1000}\left|\lambda_{k,h}^{(G)} - \lambda_{k,h}^{(N)}\right|$. The matrix errors $\|\boldsymbol{A}_h^{(G)} - \boldsymbol{A}_h^{(N)}\|$ are also computed. It can be seen from Figure 2 that both errors are $O(h^2)$.

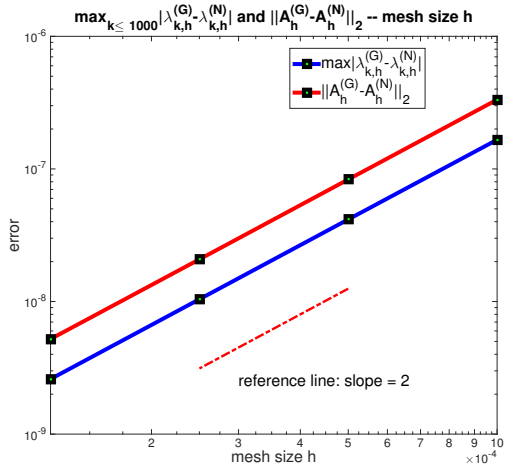

FIGURE 2. $\max_{k\le 1000} |\lambda_{k,h}^{(G)} - \lambda_{k,h}^{(N)}|$ (blue line) and $\|\boldsymbol{A}_h^{(G)} - \boldsymbol{A}_h^{(N)}\|$ (red line) v.s. h.

5. Numerical Experiments

We perform various numerical tests for the integral operator $Af := \int_D K(x,y)f(y)dy$. We use piecewise constant approximation in the Galerkin method and midpoint rule in the Nyström method. Also, uniform triangular meshes are used in the two dimensional case. In Section 5.1, the actual eigenvalue convergence rates computed by Nyström method are shown. Section 5.2 investigates the eigenfunction approximation. Section 5.3 presents a comparison of our error bounds with the ones from [12, 2, 18, 21], etc.

Example 1. We first consider an example with known eigenpairs from [9]

$$K(x,y) = e^{-\|x-y\|_1}, \quad x, y \in D = (0,1)^d \quad (d = 1, 2).$$

If $d = 1$, the exact eigenpairs of the integral operator A are given by

$$\lambda_k = \frac{2}{w_k^2 + 1}, \quad \phi_k(x) = B_k(\sin(w_k x) + w_k \cos(w_k x)), \tag{5.1}$$

where $w_k (k = 1, 2, \dots)$ are positive solutions of the equation $\tan(w) = \frac{2w}{w^2-1}$ and B_k is chosen such that $\|\phi_k\|_{L^2(D)} = 1$. The decay rate of eigenvalues is known

to be $\lambda_k = O(k^{-2})$. If $d = 2$, the exact eigenvalues/eigenfunctions are the tensor products of eigenvalues/eigenfunctions in one dimensions, i.e.,

$$\lambda_{k_1,k_2} = \lambda_{k_1}\lambda_{k_2}, \quad \phi_{k_1,k_2}(x) = \phi_{k_1}(x_1)\phi_{k_2}(x_2), \quad x = (x_1, x_2) \in \mathbb{R}^2.$$

Example 2. We consider in this example the kernel function associated with the L^2 norm in two dimensions.

$$K(x, y) = e^{-\|x-y\|_2}, \quad x, y \in D = (0, 1)^2.$$

Since the exact eigenvalues are not known, we use the computed eigenvalues over a finer mesh with mesh size $h = \sqrt{2}/200$ as reference eigenvalues to evaluate the errors of approximate eigenvalues derived from much coarser meshes($h = \sqrt{2}/25, \sqrt{2}/50$).

5.1. **Rate of convergence.** The results for Example 1 are shown in Figure 3 – 4. The results for Example 2 are shown in Figure 5 – 6, which are similar to those in Example 1.

The log-log plots in Figure 3 and Figure 6 indicate the convergence rate:

$$|\lambda - \lambda_h| = O(h^2).$$

From Figure 4 and Figure 5 (with fixed mesh size in each plot), we see that (for leading eigenvalues) the error $|\lambda - \lambda_h|$ is roughly independent of λ. Hence we deduce that there is a constant C independent of λ such that

$$|\lambda - \lambda_h|/h^2 \le C, \quad \forall\, \lambda. \tag{5.2}$$

We then examine the magnitude of the constant C. For the four problems shown in Figure 4 and Figure 5, the maximal approximation errors $\max_k |\lambda_k - \lambda_{k,h}|$ are bounded by $5 \times 10^{-8}, 3 \times 10^{-4}, 3 \times 10^{-5}, 2 \times 10^{-4}$, respectively. Hence it can be computed that the constant $C \le 0.1$. That is to say, we have $|\lambda - \lambda_h| \le 0.1h^2, \quad \forall\, \lambda$, for the above four experiments.

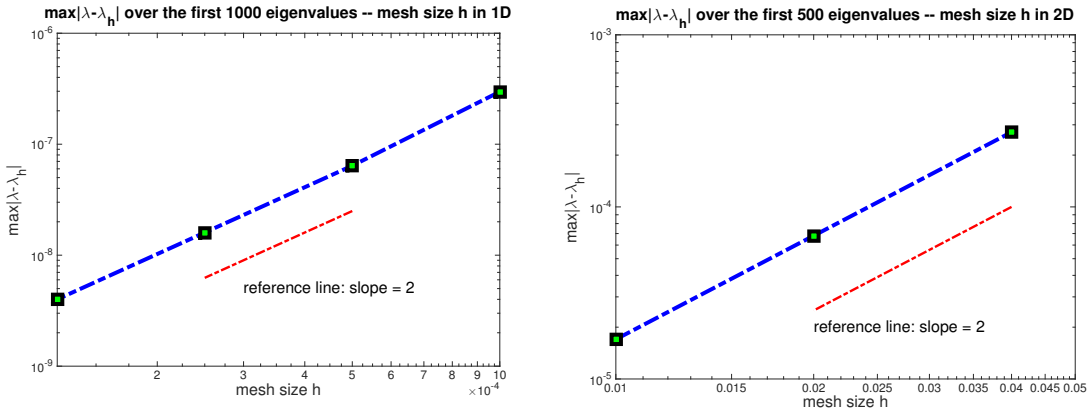

FIGURE 3. Example 1 - Rate of convergence. Left: 1D; Right: 2D

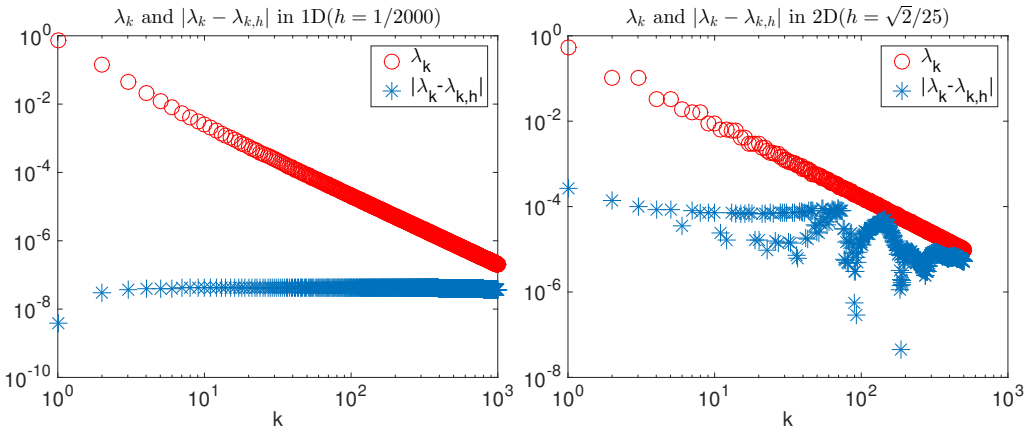

FIGURE 4. Example 1 - λ_k and $|\lambda_k - \lambda_{k,h}|$ for $1 \le k \le m$. Left: 1D, $m = 1000$, $h = 1/2000$; Right: 2D, $m = 500$, $h = \sqrt{2}/25$

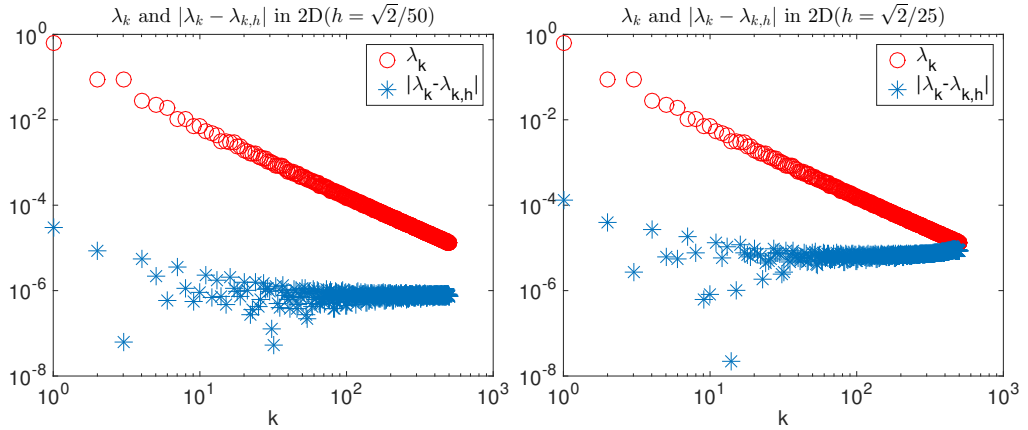

FIGURE 5. Example 2 - λ_k and $|\lambda_k - \lambda_{k,h}|$ for $1 \le k \le 500$. Left: $h = \sqrt{2}/50$; Right: $h = \sqrt{2}/25$

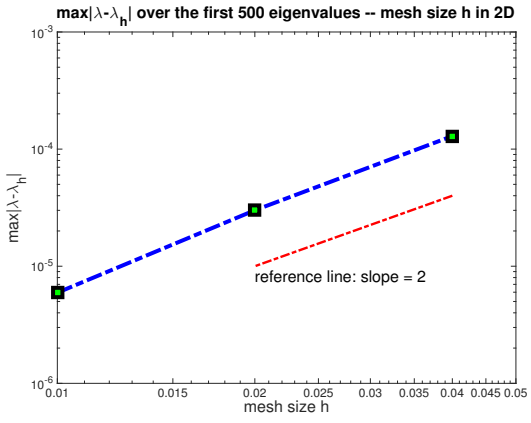

FIGURE 6. Example 2 - Rate of convergence.

5.2. **Eigenfunction approximation.** Since all theoretical error bounds for eigenvalues are expressed in terms of certain approximation errors of eigenfunctions, we compute the actual approximation error of the eigenfunctions in this section using Example 1 with $d = 1$. In the Section 5.3, we insert it into the eigenvalue estimates (in [2, 18] and (4.4)) to investigate the scalings with respect to λ (exact eigenvalue) and h (mesh size).

Numerical observations: (5.1) and Figure 7 imply that

(5.3)
$$\|\phi_k'\|_{\sup} = O(\lambda_k^{-1}), \quad \|\phi_k'\| = O(\lambda_k^{-1}) \quad \text{and} \quad \|(I - P_h)\phi_k\| = O(kh) = O(\lambda_k^{-1/2}h).$$

Theoretical estimates: Proposition 4.3 implies that $\|(I - P_h)\phi_k\| \leq C\lambda_k^{-1}h$, which differs from the numerical observation in (5.3). This may indicate that using Poincaré's inequality to estimate the approximation error $\|(I - P_h)\phi\|$ is not accurate enough.

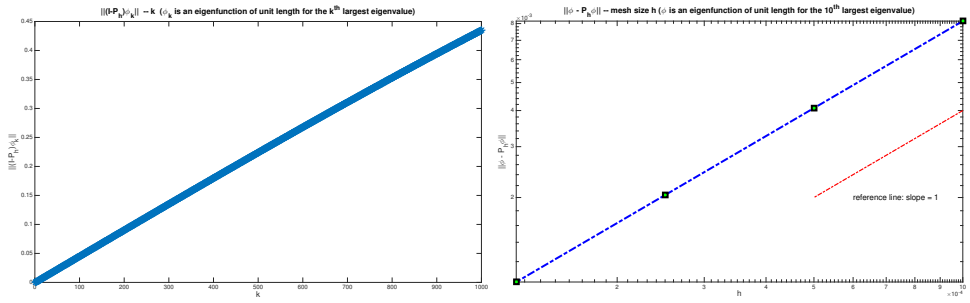

FIGURE 7. $\|\phi_k - P_h\phi_k\|$ with respect to k (left) and h (right)

5.3. **Comparison of existing theoretical estimates.** Using the exact eigenpairs in Example 1, we compare different error estimates, e.g., in [12, 2, 18, 21] and (4.10), to true errors in the eigenvalue computations. It will be seen that all theoretical error bounds overestimate the true error by a large margin of various degrees and the error bound in (4.10) is more accurate.

Estimates in (4.10)**.** Now we compute the error bound in Theorem 4.2 (or Theorem 4.1). With $\lambda_k = O(k^{-2})$, it can be computed that the constant C_k in Theorem 4.1 is $C_k = O(k^3) = O(\lambda_k^{-3/2})$. Hence the error estimate is

$$|\lambda - \lambda_h| = O(\lambda^{-3/2}h^2), \tag{5.4}$$

where the scaling h^2 is correct while the factor $\lambda^{-3/2}$ is redundant compared to (5.2).

Estimates from [12, 2, 18, 21]**.** Existing estimates for the Nyström discretization are all asymptotic and more or less of the form: $|\lambda - \lambda_h| = O(\text{quadrature error})$. For example, the estimates in [12, 2, 18] roughly say that

$$|\lambda - \lambda_h| \leq C_* \max_{\phi} \|A\phi - A_h\phi\|_{\sup} \quad \text{if } h \text{ is sufficiently small,}$$

where A_h is the quadrature operator in (4.9), $\phi \in \text{Ker}(A - \lambda I)$ is an eigenfunction of unit length, C_* is a constant that may (in [2, 18]) or may not (in [12]) depend on λ. The quadrature error $A - A_h$ corresponds to the operator Q_n in [21] and

was used in [21] to obtain the convergence rate. In Example 1, we deduce that $\|A\phi - A_h\phi\|_{\sup} = O(\|\phi'\|_{\sup} h) = O(\lambda^{-1}h)$. Hence those estimates give rise to the convergence rate

$$|\lambda - \lambda_h| \le C_* \lambda^{-1} h \quad \text{if } h \text{ is sufficiently small,}$$

which is inconsistent with the $O(h^2)$ scaling observed in Section 5.1. Moreover, in [2, 18], due to the use of spectral projection operator and an estimate of approximate resolvent in [1, Theorem 1], it can be deduced for Example 1 that $C_* = O(\lambda^{-7/2})$, which gives $|\lambda - \lambda_h| = O(\lambda^{-9/2}h)$.

Remark 5.1. *For smooth kernel functions like $e^{-|x-y|^2}$ and the piecewise constant Galerkin discretization, the numerical results in* [24] *indicate that $|\lambda - \lambda_h| = O(\lambda h^2)$.*

5.4. **A conjecture of a sharp bound.** Following the investigation in Section 5.2 on the actual approximation error of eigenfunctions, we derive a similar estimate in two dimensions and then propose a conjecture concerning the actual convergence rate. With $\phi_{k_1,k_2}(x_1, x_2) = \phi_{k_1}(x_1)\phi_{k_2}(x_2)$ in Example 1, for simplicity, we consider a tensor product mesh in $[0,1]^2$. Let $P_h^{1\mathrm{D}}$ and $P_h^{2\mathrm{D}}$ denote the projections defined in (4.3) over $[0,1]$ and $[0,1]^2$, respectively. It follows that

$$P_h^{2\mathrm{D}}\phi_{k_1,k_2} = P_h^{1\mathrm{D}}\phi_{k_1} P_h^{1\mathrm{D}}\phi_{k_2} \quad \text{and} \quad \|P_h^{2\mathrm{D}}\phi_{k_1,k_2}\|_{[0,1]^2}^2 = \|P_h^{1\mathrm{D}}\phi_{k_1}\|_{[0,1]}^2 \|P_h^{1\mathrm{D}}\phi_{k_2}\|_{[0,1]}^2.$$

Using the one dimensional result in (5.3), we deduce that

$$\begin{aligned}
\|(I - P_h^{2\mathrm{D}})\phi_{k_1,k_2}\|_{[0,1]^2}^2 &= \|\phi_{k_1}\|_{[0,1]}^2 \|\phi_{k_2}\|_{[0,1]}^2 - \|P_h^{1\mathrm{D}}\phi_{k_1}\|_{[0,1]}^2 \|P_h^{1\mathrm{D}}\phi_{k_2}\|_{[0,1]}^2 \\
&= \|(I - P_h^{1\mathrm{D}})\phi_{k_1}\|^2 \|\phi_{k_2}\|^2 + \|P_h^{1\mathrm{D}}\phi_{k_1}\|^2 \|(I - P_h^{1\mathrm{D}})\phi_{k_2}\|^2 \\
&= O\left((\lambda_{k_1}^{-1} + \lambda_{k_2}^{-1})h^2\right) \le C\lambda_{k_1,k_2}^{-1} h^2
\end{aligned}$$

in accordance with the one dimensional counterpart in (5.3).

The numerical results lead us to the following conjecture:

> Let λ and λ_h denote the exact and approximate kth largest eigenvalue, respectively. Then
>
> $$\|(I - P_h)\phi\| \le C_1 \lambda^{-1/2} h, \quad \forall\, \phi \in \mathrm{Ker}(A - \lambda I), \quad \|\phi\| = 1,$$
>
> and
>
> $$|\lambda - \lambda_h| \le C_2 \lambda \max_{\substack{\phi \in \mathrm{Ker}(A-\lambda I) \\ \|\phi\|=1}} \|(I - P_h)\phi\|^2$$
>
> where the constants C_1, C_2 are independent of λ or h.

6. Conclusion

We obtain eigenvalue error estimates of second order for the lowest order Galerkin and Nyström discretizations. The equivalence between the two discretizations is established, which makes the analysis of the Nyström method a consequence of the Galerkin one. The resulting estimates appear more accurate than the previously available ones. Numerical experiments illustrate and complement the new and previously existing theoretical results.

7. Appendix

To prove Theorem 2.1, some technical tools are needed, where the Schoenberg Interpolation Theorem relates positive definiteness to *completely monotone functions* (cf. [6]).

Definition 7.1 (completely monotone functions)**.** *A function f is called* completely monotone *on $[0,\infty)$ if $f \in C[0,+\infty) \cap C^\infty(0,+\infty)$ and $(-1)^k f^{(k)}(t) \geq 0, \quad \forall t > 0,\ k = 0, 1, \ldots$.*

Theorem 7.1 (Schoenberg Interpolation Theorem)**.** *Let $\|\cdot\|$ denote a norm induced by an inner product on $\mathbb{R}^d$. If f is completely monotone but not constant on $[0,+\infty)$, then for any n distinct points $x_1, \ldots, x_n \in \mathbb{R}^d$, the matrix $a_{i,j} = f(\|x_i - x_j\|^2)$ is symmetric positive definite.*

The Bernstein-Widder Theorem shows that the Laplace transform of a nonnegative $L^1(\mathbb{R}_+)$ function is completely monotone (cf. [6]).

Theorem 7.2 (Bernstein-Widder Theorem)**.** *A function $f : [0,+\infty) \to [0,+\infty)$ is completely monotone if and only if there is a nondecreasing bounded function ξ such that $f(t) = \int_0^{+\infty} e^{-st} d\xi(s)$.*

Proposition 7.1 lists two completely monotone functions that are needed in the proof.

Proposition 7.1. *The following two functions are completely monotone on $[0,\infty)$: (1). $f(t) = e^{-t}$; (2). $f(t) = e^{-\sqrt{t}}$.*

Proof. $f(t) = e^{-t}$ is completely monotone from the definition. For $f(t) = e^{-\sqrt{t}}$, we show that $f(t)$ satisfies the assumption in Theorem 7.2 with $\xi(s) = -\mathrm{erf}(\frac{1}{2\sqrt{s}})$, where $\mathrm{erf}(x)$ denotes the error function. In fact, it can be computed that

$$\int_0^\infty e^{-st} d\xi(s) = e^{-\sqrt{t}} = f(t).$$

Hence $f(t)$ is completely monotone according to Theorem 7.2 and the proof is complete.

□

Now we are in a position to carry out the proof of Theorem 2.1.

Proof of Theorem 2.1. If ρ is the weighted L^1 norm, then $\Phi(x) = e^{-\rho(x)}$ can be written as the inverse Fourier transform of a positive function in $L^1(\mathbb{R}^d)$. In fact, we have

$$\Phi(x) = C \int_{\mathbb{R}^d} \left(\Pi_{k=1}^d \frac{1}{\omega_k y_k^2 + \omega_k^{-1}} \right) e^{ixy} dy = C \int_{\mathbb{R}^d} \hat{\Phi}(y) e^{ixy} dy$$

where $C > 0$ and $\hat{\Phi}(y) = \Pi_{k=1}^d \frac{1}{\omega_k y_k^2 + \omega_k^{-1}}$. For n points $x_1, \ldots, x_n$ in $\mathbb{R}^d$ and a nonzero vector $(c_1, \ldots, c_n)$, we have

$$\sum_{k=1}^n \sum_{j=1}^n c_j \overline{c_k} \Phi(x_j - x_k) = C \int_{\mathbb{R}^d} \hat{\Phi}(y) \left| \sum_{j=1}^n c_j e^{ix_j y} \right|^2 dy > 0.$$

Thus the matrix $a_{i,j} = \Phi(x_i - x_j)$ is positive definite.

For the rest two forms of ρ, the result follows from the Schoenberg Interpolation Theorem and Proposition 7.1. □

References

[1] K. E. Atkinson. The numerical solution of the eigenvalue problem for compact integral operators. *Transactions of the American Mathematical Society*, 129(3):458–465, 1967.

[2] K. E. Atkinson. Convergence rates for approximate eigenvalues of compact integral operators. *SIAM Journal on Numerical Analysis*, 12(2):213–222, 1975.

[3] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM review*, 52(2):317–355, 2010.

[4] R. Bhatia. *Perturbation Bounds for Matrix Eigenvalues*, volume 53. SIAM, 1987.

[5] F. Chatelin. *Spectral approximation of linear operators*. SIAM, 2011.

[6] E.W. Cheney and W.A. Light. *A Course in Approximation Theory*. Graduate studies in mathematics. American Mathematical Society, 2009.

[7] G. Christakos. *Modern spatiotemporal geostatistics*, volume 6. Oxford University Press, 2000.

[8] L. W. Gelhar. Stochastic subsurface hydrology from theory to applications. *Water Resources Research*, 22(9S), 1986.

[9] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[10] M. D. Gunzburger, C. G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numerica*, 23:521–650, 2014.

[11] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[12] H. B. Keller. On the accuracy of finite difference approximations to the eigenvalues of differential and integral operators. *Numerische Mathematik*, 7(5):412–419, 1965.

[13] M. A. Krasnosel'skii, G. M. Vainikko, R. P. Zabreyko, Ya. B. Ruticki, and V. Ya. Stetsenko. *Approximate solution of operator equations*. Springer Science & Business Media, 2012.

[14] R. Kress. *Linear Integral Equations*. Applied Mathematical Sciences. Springer New York, 2013.

[15] P. D. Lax. *Functional analysis*. Pure and applied mathematics. Wiley, 2002.

[16] N. M. Nasrabadi. Pattern recognition and machine learning. *Journal of electronic imaging*, 16(4):049901, 2007.

[17] E. J. Nyström. Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben. *Acta Mathematica*, 54(1):185–204, 1930.

[18] J. E. Osborn. Spectral approximation for compact operators. *Mathematics of computation*, 29(131):712–725, 1975.

[19] S. Osborn, P. S. Vassilevski, and U. Villa. A multilevel, hierarchical sampling technique for spatially correlated random fields. *SIAM Journal on Scientific Computing*, 39(5):S543–S562, 2017.

[20] W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991.

[21] A. Spence. Error bounds and estimates for eigenvalues of integral equations. *Numerische Mathematik*, 29(2):133–147, 1978.

[22] A. Spence. Error bounds and estimates for eigenvalues of integral equations. ii. *Numerische Mathematik*, 32(2):139–146, 1979.

[23] G. W. Stewart. *Matrix Algorithms Volume 2: Eigensystems*, volume 2. SIAM, 2001.

[24] R. A. Todor. Robust eigenvalue computation for smoothing operators. *SIAM journal on numerical analysis*, 44(2):865–878, 2006.

[25] G. M. Vainikko. On the speed of convergence of approximate methods in the eigenvalue problem. *USSR Computational Mathematics and Mathematical Physics*, 7(5):18–32, 1967.

[26] H. Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

[27] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

Department of Mathematics, Purdue University, West Lafayette, IN 47907-2067, U.S.A. (cai92@purdue.edu)

Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR 97207, U.S.A. (panayot@pdx.edu), and Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94550, U.S.A. (panayot@llnl.gov).