Numer Algor (2015) 70:709–726
DOI 10.1007/s11075-015-9970-0

ORIGINAL PAPER

On spectral distribution of kernel matrices related to radial basis functions

Andrew J. Wathen · Shengxin Zhu

Received: 13 January 2014 / Accepted: 26 January 2015 / Published online: 10 April 2015

Abstract This paper focuses on spectral distribution of kernel matrices related to radial basis functions. By relating a contemporary finite-dimensional linear algebra problem to a classical problem on infinite-dimensional linear integral operator, the paper shows how the spectral distribution of a kernel matrix relates to the smoothness of the underlying kernel function. The asymptotic behaviour of the eigenvalues of a infinite-dimensional kernel operator are studied from a perspective of low rank approximation—approximating an integral operator in terms of Fourier series or Chebyshev series truncations. Further, we study how the spectral distribution of interpolation matrices of an infinite smooth kernel with flat limit depends on the geometric property of the underlying interpolation points. In particularly, the paper discusses the analytic eigenvalue distribution of Gaussian kernels, which has important application on stably computing of Gaussian radial basis functions.

Keywords Eigenvalues · Radial basis functions · Spectral distribution · Integral equation of the first kind

Mathematics Subject Classification (2010) 42A10 · 45A25 · 45B05 · 45C05 · 47A52 · 47A75

This research is partially supported by Nature Science Foundation of China (No. 61170309, No.61472462, and No.91430218) and the Laboratory of Computational Physics.

A. J. Wathen
Numerical Analysis Group, The University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, England
e-mail: wathen@maths.ox.ac.uk

S. Zhu (✉)
Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, P.O.Box 8009, Beijing 100088, China
e-mail: zhus@maths.ox.ac.uk; sxchu@foxmail.com

1 Motivation

The increasing importance of high-dimensional problems and scattered data processing motivates us to investigate the properties of *kernel matrices* related to radial basis functions (RBFs). RBFs have been shown as an attractive approach for scattered data approximation [46]. They can guarantee invertible linear systems [25], have good approximating quality, and are promising to deal with high dimensional problems and complex geometry domain. These attractive properties make RBFs one of the foundations of many multivariate approximation based methods and techniques such as mesh-free methods, machine learning, global optimization, surface reconstruction and computing. However, research has found that linear systems related to some globally supported radial basis functions can be highly ill-conditioned for standard basis which can result in extremely large condition number. The condition number of an matrix A is defined by $\kappa(A) = \|A^{-1}\|\|A\|$, where $\|\cdot\|$ denotes the standard 2-norm of matrices. It depends on the ratio of the largest magnitude eigenvalue to the smallest magnitude eigenvalue for symmetric matrices; many RBFs interpolation matrices are of this type.

Stably computing the ill-conditioned linear system usually need to change the standard basis to a better set of basis and to explore the spectral information. For example, Gaussian radial basis functions $e^{-\varepsilon^2\|\mathbf{x}\|^2}$, $\mathbf{x} \in \mathbb{R}^d$, with *flat limit*—the cases corresponding to $\varepsilon \to 0$—can result in highly ill-conditioned linear systems [13, 22]. Stably computing such highly ill-conditioned linear systems requires spectral information of Gaussian kernel, see [9, 11, 23]. In the context of finite element method (FEM), based on the spectral distribution of Galerkin mass matrices, a simple and efficient diagonal preconditioner has been developed to speed up the computation [45]. Such work has already vividly shown that priori information on the spectral distribution of a linear system is useful and sometime essential to design a simple and efficient preconditioner for fast and stable solvers.

Pioneering results on the spectral information of RBFs interpolation matrices focused primarily on the estimation of the smallest (magnitude) eigenvalue of the underlying interpolation matrices. There are two fundamental motivations behind such work: first, for understanding the solvability of high-dimensional interpolation problem with certain radial basis functions, to prove the related matrices do not have any zero eigenvalue [1, 25]; second, for understanding the stability of theoretical computing, to estimate the upper bound of the norm of the inverse of an interpolation matrix, or to estimate the condition number of the underlying interpolation matrix, in this case, sharper estimation on the smallest magnitude eigenvalue is often required [3, 26–28, 40, 41].

Spectral distribution can further characterize the conditioning issues and reveal more information than the conditioning number does [39]. Our recent research finds that several other important computing issues are also closely related to the spectral distribution (for example, constructing preconditioners and truncated SVD based regularization). Further understanding the spectral information of RBFs interpolation matrices is still necessary.

One main idea of the paper is to relate the RBFs interpolation scheme to the discrete integral equation of the first kind, see Section 2. Instead of considering the finite-dimensional linear algebra problem directly, we investigate the corresponding classical infinite-dimensional linear operator problem. Once put a contemporary problem in a proper perspective, powerful tools are already available and many results can be applied to explore the underlying problem. Combining several scattered results, we obtain results on how the decay of eigenvalues of kernel matrices of RBFs is closely related to the smoothness of the underlying RBFs (Theorem 4 and Theorem 5). The results perhaps are not striking new, but we have not seen the results elsewhere. Several established results—Theorem 1, Theorem 3 and Theorem 6—are well known in the community of linear integral equations, but it seems that these results drew little attention from the community of RBFs. The difference and connections between these theorems are discussed.

The remaining of the paper is organised as follow. Section 2 introduce the connection between the RBFs interpolation problem and the Fredholm integral equation of the first kind. Also introduced are several relevant lemmas including Weyl-Courant minimax principle. Main results are discussed in Section 3. Finally, we discuss other relevant results in existence, the connection with RBF-QR method, and possible applications.

2 Preliminaries

Consider the following integral equation of the first kind

$$\int_{\Omega} K(\mathbf{x}, \mathbf{y})\alpha(\mathbf{y})d\mathbf{y} = f(\mathbf{x}) \tag{1}$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. We call $K(\mathbf{x}, \mathbf{y})$ a *kernel function* in $\mathbb{R}^d$. Further, if $K(\mathbf{x}, \mathbf{y})$ satisfies $\int_{\Omega} K(\mathbf{x}, \mathbf{y})\alpha(\mathbf{y})\alpha(\mathbf{x})\, d\mathbf{x}d\mathbf{y} > 0$ for any non-zero function $\alpha(\mathbf{x})$, then $K(\mathbf{x}, \mathbf{y})$ is positive definite. If a simple quadrature rule is applied to sample $\alpha(\mathbf{y})$, then we get a discrete equation

$$\sum_{j=1}^{N} K(\mathbf{x}, \mathbf{y}_j) w_j \alpha(\mathbf{y}_j) = f(\mathbf{x}). \tag{2}$$

Further collocating at $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$ gives

$$\sum_{j=1}^{N} K(\mathbf{x}_i, \mathbf{y}_j) w_j \alpha(\mathbf{y}_j) = f_i, i = 1, 2, \cdots, N, \tag{3}$$

where $f_i = f(\mathbf{x}_i)$. Let $K(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|)$, then we can write these equations as the linear system

$$\begin{pmatrix} \phi(\|\mathbf{x}_1 - \mathbf{y}_1\|) & \phi(\|\mathbf{x}_1 - \mathbf{y}_2\|) & \cdots & \phi(\|\mathbf{x}_1 - \mathbf{y}_N\|) \\ \phi(\|\mathbf{x}_2 - \mathbf{y}_1\|) & \phi(\|\mathbf{x}_2 - \mathbf{y}_2\|) & \cdots & \phi(\|\mathbf{x}_2 - \mathbf{y}_N\|) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\|\mathbf{x}_N - \mathbf{y}_1\|) & \phi(\|\mathbf{x}_N - \mathbf{y}_2\|) & \cdots & \phi(\|\mathbf{x}_N - \mathbf{y}_N\|) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}, \tag{4}$$

where $\alpha_j = w_j \alpha(\mathbf{y}_j)$, $j = 1, 2, \cdots, N$. We denote the matrix in (4) by $A_{\phi,\mathcal{X}}$, where $\mathcal{X}$ denotes the set $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$. When $\mathbf{x}_i = \mathbf{y}_i$, $1 \leq i \leq N$, the linear system (4) is the same as that obtained for interpolation matrices with the radial basis function ϕ at the points $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$. In this paper we call the matrix in (4) a *kernel matrix*.

On the other hand, consider the interpolation scheme with a sum of translates of a radial basis function

$$s(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j \phi\left(\|\mathbf{x} - \mathbf{y}_j\|\right). \tag{5}$$

If $s(\mathbf{x})$ interpolates an unknown function, $f(\mathbf{x})$, on the data set $\mathcal{X}$, and further $\mathbf{x}_j = \mathbf{y}_j$, $1 \leq j \leq N$, it results in the same linear system (4). In this case we call the matrix $A_{\phi,\mathcal{X}}$ an *interpolation matrix*. In this paper, we use kernel matrices for short.

We investigate the spectral distribution of kernel matrices related to radial basis function by studying the infinite dimensional linear operator

$$\mathcal{K}\Psi(\mathbf{x}) = \int_{\Omega} K(\mathbf{x}, \mathbf{y})\Psi(\mathbf{y})d\mathbf{y}, \tag{6}$$

where $K(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|)$ is a radial function ϕ. To investigate the spectral distribution of the operator $\mathcal{K}$ in (6), consider the following well known Weyl-Courant minimax principle.

Lemma 1 (The Weyl-Courant minimax principle) *Let $\mathcal{K}$ be a compact symmetric operator on a Hilbert space $\mathcal{H}$ with eigenvalues*

$$|\lambda_0| \geq |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq \cdots,$$

and $\mathcal{S}$ be any operator of rank $\leq n$ on $\mathcal{H}$, then $\|\mathcal{K} - \mathcal{S}\| \geq |\lambda_n|$.

The reader is directed to [24] for proof of the lemma. In [24], Little and Reade apply the Weyl-Courant minimax principle to tails of the Chebyshev expansion for analytic kernels, and conclude that the eigenvalues of an analytic kernel on a finite interval go to zero at least as fast as R^{-n} for some fixed $R > 1$. The proof depends on the estimation of Chebyshev coefficients of analytic functions. Some results used in their proof in fact date back to Bernstein [4] and relate to the so-called *Bernstein's ellipse*, denoted as $\mathcal{E}_\rho$ [44, p.56]. The ellipse, $\mathcal{E}_\rho$, has foci at ± 1, and the sum of the semi-axes equals $\rho > 1$.

Lemma 2 ([4][44, p.57]) *Let a function, f, analytic in $[-1, 1]$ be analytically continuable to the open Bernstein ellipse $\mathcal{E}_\rho$, where it satisfies $|f(z)| \leq M$ for some M. Then its Chebyshev coefficients, a_n, satisfy $a_0 \leq M$ and*

$$|a_n| \leq 2M\rho^{-n}, n \geq 1. \tag{7}$$

Its Chebyshev truncations satisfy

$$\|f - S_n\| \leq \frac{2M\rho^{-n}}{\rho - 1} \tag{8}$$

and its Chebyshev interpolants satisfy

$$\|f - p_n\| \leq \frac{4M\rho^{-n}}{\rho - 1}, \tag{9}$$

where $S_n(x) = \sum_{k=0}^{n} a_k T_k(x)$*, and* $p_n(x)$ *is the polynomial obtained by interpolation in Chebyshev points.*

The formula (7) is due to Bernstein, and the second part of the lemma can be found in [44, p.57, Thm. 8.2].

A function of *bounded variation* on $\mathbb{R}$, $f(x)$, is a real-valued function, it is an integrable function, say, $f(x) \in L(\mathbb{R})$, and, V, the supremum of $\int f(x)\, dg(x)$ over all $g(x) \in C^1(\mathbb{R})$ with $|g(x)| < 1$ is finite. If $f(x)$ is continuous, then the supremum of $\sum_{j=1}^{N} |f(x_j) - f(x_{j-1})|$ is bounded over all finite samples $x_0, x_1, \cdots, x_N$. For finitely differentiable functions with the highest derivatives of bounded variations, we consider the following result.

Lemma 3 ([44, p.52-p.53]) *For any integer* $\nu \geq 0$*, let* u *and its derivatives* $u', \ldots, u^{(\nu-1)}$ *be absolutely continuous on* $[-1, 1]$*, and* $u^{(\nu)}$ *be of bounded variation* V*, then for any* $n \geq \nu + 1$*, the Chebyshev coefficients of* u *satisfy*

$$|a_n| \leq \frac{2V}{\pi(n-\nu)^{\nu+1}},\ n > \nu. \tag{10}$$

Its Chebyshev truncations S_n *satisfy*

$$\|u - S_n\| \leq \frac{2V}{\pi\nu(n-\nu)^{\nu}},\ n > \nu \tag{11}$$

and its Chebyshev interpolants satisfy

$$\|f - p_n\| \leq \frac{4V}{\pi\nu(n-\nu)^{\nu}},\ n > \nu. \tag{12}$$

For details of the proof of Lemma 3, the reader is directed to Trefethen's new book [44, Thm 7.1 and Thm 7.2].

Besides Chebyshev truncations, we also consider the Fourier series truncations. Such an approximation is valid for a kernel function which is Lebesgue integrable, not necessarily analytic, and it can be generalized to higher dimensional space. Here, we demonstrate the idea with the simplest one dimensional case. For a Lebesgue integrable function u on Ω, say, $u \in L^1(\Omega)$, the n-th Fourier coefficient of u is defined by

$$\hat{u}_n := \hat{u}(n) = \frac{1}{T}\int_{-T/2}^{T/2} u(t)e^{-int}\, dt \tag{13}$$

for a T-periodic function, u, on $\Omega = [-T/2, T/2]$, where T could be finite or infinite. If there is no confusion we use $\hat{u}_n$ for short, otherwise we use $\hat{u}(n)$. The

Fourier series $S[u]$ of a function $u \in L^1(\Omega)$ is the trigonometric series

$$S[u] \sim \sum_{n=-\infty}^{\infty} \hat{u}_n e^{int}. \tag{14}$$

The n-th Fourier series truncation is denoted as $S_n[u] = \sum_{k=-n}^{n} \hat{u}_n e^{int}$. Remembering the Euler's formula $e^{it} = \cos t + i \sin t$, the Fourier series of a function $u \in L^1(\Omega)$ is equal to

$$\frac{A_0}{2} + \sum_{n=1}^{\infty}(A_n \cos nt + B_n \sin nt) = \frac{A_0}{2} + \sum_{n=1}^{\infty} \sqrt{A_n^2 + B_n^2} \sin(nt + \theta_n), \tag{15}$$

where $A_n = \hat{u}_n + \hat{u}_{-n}$, $B_n = i\left\{\hat{u}_n - \hat{u}_{-n}\right\}$ and $\arcsin \theta_n = B_n / \sqrt{A_n^2 + B_n^2}$. In particularly, when u is an even function, $\hat{u}_n = \hat{u}_{-n}$,

$$S_n = \hat{u}(0) + 2\sum_{k=1}^{n} \hat{u}_n \cos nt. \tag{16}$$

As discussed above, it is clear that *the Fourier series truncation S_n is a function of finite rank $n+1$*.

According to standard classical Fourier analysis, Fourier coefficients have the following properties, for example, see [20, p.3].

Proposition 1 *If $u \in L^1(\Omega)$, then*

(a) *denote $u_\tau = u(t-\tau)$, $\tau \in \Omega$; then $\hat{u}_\tau(n) = \hat{u}_n e^{-in\tau}$;*
(b) $|\hat{u}_n| \le \frac{1}{T}\int |u(t)|\, dt = \|u\|_{L^1}$;
(c) *denote $u_\varepsilon(t) = u(\varepsilon t)$ for some $\varepsilon > 0$, then $\hat{u}_\varepsilon(n) = \frac{1}{\varepsilon}\hat{u}\left(\frac{n}{\varepsilon}\right)$;* [1]
(d) *if u has ν-the order derivative, then $\hat{u}^{(\nu)}(n) = (in)^\nu \hat{u}(n)$;*
(e) *(The Riemann-Lebesgue Lemma)* $\lim_{|n|\to\infty} \hat{u}(n) = 0$.

If u is a ν-times differentiable function and $u^{(\nu)} \in L^1(\Omega)$, then according to the Riemann-Lebesgue Lemma

$$\lim_{|n|\to\infty} \hat{u}^{(\nu)}(n) = \lim_{|n|\to\infty} (in)^\nu \hat{u}(n) = 0, \tag{17}$$

which suggest $\hat{u}(n) = o\left(\frac{1}{n^\nu}\right)$ as $|n| \to \infty$. The small "o" symbol is defined in the standard way : $u(k) = o(g(k))$ as $k \to \infty$ if $\lim_{k\to\infty} |u(k)|/|g(k)| = 0$.

If u is of bound variation V on Ω, then integrate (13) by parts

$$|\hat{u}(n)| = \left|\frac{1}{T}\int e^{-int} u(t)dt\right| = \left|\frac{1}{iTn}\int u(t)de^{-int}\right| \le \frac{2V}{T|n|}. \tag{18}$$

[1] $\hat{u}\left(\frac{n}{\epsilon}\right)$ is defined similarly as the Fourier coefficients, but note that $\frac{n}{\varepsilon}$ may not be an integer.

In this case $\hat{u}(n) = \mathcal{O}\left(\frac{1}{|n|}\right)$, i.e. as $|n| \to \infty$, there exist a constant C such that $|\hat{u}(n)| < C/|n|$. Similarly if u has up to $\nu-1$-time continuous derivatives and the ν-th derivative is of bounded variation and belongs to $L^1(\Omega)$, then $\hat{u}(n) = \mathcal{O}\left(1/n^{\nu+1}\right)$.

One case of great interest is the square integrable function u on Ω, say, $u \in L^2(\Omega)$. In this case, $u = \lim_{n\to\infty} \sum_{k=-n}^{n} \hat{u}e^{ikt}$, in the L^2 norm. And further we have

$$\sum_{n=-\infty}^{\infty} \left|\hat{u}(n)^2\right| = \frac{1}{T}\int |u(t)|^2 dt. \tag{19}$$

Because the series in (19) is convergent, further suppose $\hat{u}(n) > 0$, then $\hat{u}(n)^2 = o(1/n)$ and $\hat{u}(n) = o\left(1/n^{1/2}\right)$. Similarly, if u is k-times differentiable, and $u^{(k)} \in L^2(\Omega)$, then $\hat{u}^{(k)}(n) = o\left(1/n^{k+1/2}\right)$. By such standard analysis, see [20] for example, we have

Lemma 4 (Fourier coefficients of differentiable functions) *Let u be a square integral function on Ω with Fourier transform $\hat{u}$ and Fourier coefficients $\hat{u}_n$.*

1. *If u has $\nu - 1$ continuous derivatives in $L^2(\Omega)$ for some $\nu \geq 0$, and the ν^{th} derivative is of bounded variation, then $\hat{u}_n = \mathcal{O}\left(|n|^{-\nu-1}\right)$ as $|n| \to \infty$.*
2. *If $u \in L^2(\Omega)$ and $u \in C^\infty$, then $\hat{u}_n = \mathcal{O}(|n|^{-\nu})$ as $|n| \to \infty$ for every $\nu \geq 0$.*
3. *If u has up to ν times continuous derivatives and $u^{(\nu)} \in L^2(\Omega)$, then $\hat{u}(n) = o\left(n^{-\nu-1/2}\right)$.*

3 Main results

The eigenvalue problem of the linear integral operator (6) has been well-studied for many years since the work of Fredholm in 1903 [15]. There are many insightful results in the literature, see [8, 15, 17, 18, 21, 24, 30–38, 43, 47] for example. This paper only collects the most relevant and brief results, provides alternative simple proofs, and develops new results.

Theorem 1 (Weyl[47, p.449-450]) *If the kernel function $K(x, y) = K(y, x)$ and $\frac{\partial K^\nu(x,y)}{\partial^\nu x}$ exist and are continuous, then the magnitude of its eigenvalues decays at the order $|\lambda_n| = o\left(n^{-\nu-1/2}\right)$.*

The original result of Weyl only states this results for the case $\nu = 1$, and states it as $\lim_{n\to\infty} n^{3/2}\lambda_n = 0$ [47, p.449].

If the kernel function is positive definite, sharper results hold [7, 16, 36].

Theorem 2 (Reade-Ha[16, 36]) *If a kernel function $K(x, y)$ is positive definite, 2π-periodic in x and y, and ν times continuously differentiable, then*

1. *$\lambda_n = \mathcal{O}\left(n^{-\nu-1}\right)$;*
2. *for even ν, the sharper result $\sum_{n=1}^{\infty} n^\nu \lambda_n < \infty$ holds.*

Remark 1 As far as we know, the above theorem was first proved for $\nu = 1$ in [32] and then for the general case [33]. The second part of the theorem was initially believed to be true for all ν [16], but Reade [36] constructs a counter example showing that the second part doesn't hold for odd p. The eigenvalues of positive kernels have also been considered in [5, 6, 10] with additional constraints.

It has been noted that both Theorem 1 and Theorem 2 are sharp [35].

Without the positive definite constraint, similar results can be obtained according to the following results which state the essential connection between the Fourier coefficients of a 2π-periodic kernel function and the eigenvalues of its corresponding integral operator.

Theorem 3 (Hille-Tamarkin[17, p.10]) *If $K(x, y) = k(x-y)$ and $k(x) \in L^2[-\pi, \pi]$ is a periodic kernel, then $\lambda_n = 2\pi\hat{k}_n$, where $\hat{k}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} k(x)e^{-inx}\,dx$ is the Fourier coefficient of the kernel function.*

Applying Lemma 4 and Theorem 3, it is easy to show the following results.

Theorem 4 *Let $K(x, y) = k(x - y)$, where $k(x)$ is a 2π periodic and square integrable function on $[-\pi, \pi]$. Further let $|\lambda_n|$ be the n^{th} largest eigenvalue in magnitude of its corresponding integral operator, then*

1. *if $k(x) \in C^\nu$, for some natural number $\nu > 0$ and its ν-th derivative is of bounded variation, then $|\lambda_n| = \mathcal{O}\left(n^{-\nu-1}\right)$ as $n \to \infty$;*
2. *if $k(x) \in C^\infty$, then $|\lambda_n| = \mathcal{O}(n^{-\nu})$ as $n \to 0$ for any $\nu > 0$;*
3. *if $k(x) \in C^\nu$ has up to ν times continuous derivatives and the ν-th derivative is square integrable, then $|\lambda_n| = o\left(n^{-\nu-1/2}\right)$.*

Now, we consider methods for more general cases.

3.1 First method: truncated Fourier series approximation

For simplicity, we consider the compactly supported radial basis functions; such functions are even, though themselves are not necessarily periodic but can be extended to be periodic on the real line. Let $k(x)$ be a square integrable even function in $\mathbb{R}$. Its Fourier series truncation has the form in (16). Furthermore,

$$\lim_{n\to\infty} \|k - S_n\|_{L^2} = 0. \tag{20}$$

This implies $\sum_n^\infty |\hat{k}(n)|^2$ is convergent. According to the translation property of Fourier coefficient, Proposition 1(a),

$$k(x - y) \sim \hat{k}(0) + \sum_{n=1}^{\infty} \hat{k}(n)e^{-iny} \cos nx, \tag{21}$$

since $|\hat{k}_y(n)|^2 = |\hat{k}(n)e^{-iny}|^2 = |\hat{k}(n)|^2$, and thus

$$\|k(x-y) - S_n[k_y]\|_{L^2}^2 = \sum_{m=n+1}^{\infty} |\hat{k}_y(m)|^2 = \sum_{m=n+1}^{\infty} |\hat{k}(m)|^2. \tag{22}$$

According to the Weyl-Courant minimax principle, the $n+1$-st eigenvalue of the kernel function $k(x, y)$ satisfies

$$|\lambda_{n+1}|^2 \leq \left\|k(x-y) - S_n\left[k_y\right]\right\|_{L^2}^2 = \sum_{m=n+1}^{\infty} |\hat{k}(m)|^2. \tag{23}$$

Suppose $k(x)$ satisfies Lemma 4.1, then we get

$$\begin{aligned}|\lambda_{n+1}|^2 &\leq \sum_{m=n+1}^{\infty} |\hat{k}(m)|^2 \leq \sum_{m=n+1}^{\infty} \left|Cm^{-2\nu-2}\right| \\ &\leq \int_{n+1}^{\infty} \frac{C}{t^{2\nu+2}} dt = \frac{C}{(2\nu+1)} \frac{1}{(n+1)^{2\nu+1}}.\end{aligned}$$

Thus, $|\lambda_n| = \mathcal{O}\left(n^{-\nu-1/2}\right)$. Similarly for $k^{(\nu)} \in L^2(\Omega)$, we have $|\lambda_n| = o(n^{-\nu})$.

Theorem 5 *Let ϕ be a radial basis function in $L^2(\Omega)$, then*

1. *if ϕ has $\nu - 1$ continuous derivatives for some $\nu > 0$ and the ν^{th} derivative is of bounded variation, then its eigenvalues decay at least in the order $\mathcal{O}\left(n^{-\nu-1/2}\right)$.*
2. *if ϕ has $\nu - 1$ continuous derivatives and $\phi^{(\nu)} \in L^2(\Omega)$, then the eigenvalues of the corresponding linear operator decay in the order $o(n^{-\nu})$.*

With Proposition 1.(c) and a few additional computations, we obtain the following results for scaled radial basis functions.

Corollary 1 *Let $\phi_\epsilon(x) = \phi(\epsilon x)$, $\epsilon > 0$ be a scaled radial basis function in $L^2(\Omega)$, then*

1. *if ϕ has $\nu - 1$ continuous derivatives for some $\nu > 0$ and the ν^{th} derivative is of bounded variation, then its eigenvalues in (4) decay at least in the order $\mathcal{O}\left((\frac{\epsilon}{n})^{\nu+1/2}\right)$.*
2. *ϕ has ν continuous derivatives and $\phi^{(\nu)} \in L^2(\Omega)$, then the eigenvalues of the corresponding linear operator decay in the order $o\left((\frac{\epsilon}{n})^{\nu}\right)$.*

The following method employing Chebyshev truncation brings sharper results.

3.2 Second method: truncated Chebyshev series approximation

Using truncated Chebyshev series approximations to an analytic function has been studied in [24]. As mentioned, two key points in their poof are the Weyl-Courant minimax principle and Bernstein's results (7). A slightly different formula to (8) is needed. The Chebyshev expansion of a kernel function is

$$K(x, y) = \frac{1}{2}a_0(y) + \sum_{\ell=1}^{\infty} a_\ell(y)T_\ell(x) \tag{24}$$

Thus the estimation of $a_\ell(y)$ is needed. For details see [24].

Theorem 6 (Little-Reade [24]) *If $K(x, y) = K(y, x) \in C[-1, 1]^2$, and for each $y \in [-1, 1]$ there is an analytic continuation to $K(z, y)$ for z inside the ellipse $\mathcal{E}_\rho$, which is uniformly bounded in z, y in this range, then the eigenvalues corresponding to the kernel function $K(x, y)$ decay in the order $|\lambda_{n+1}| = \mathcal{O}(\rho^{-n})$.*

Proof See [24], or one can directly apply Lemma 2. □

Now consider the scaled kernel function $K(\epsilon x, \epsilon y)$ in the unit square,

$$\mathcal{K} f(\epsilon x) = \int_{|\epsilon t| \le 1} K(\epsilon x, \epsilon t) f(\epsilon t) d(\epsilon t). \tag{25}$$

Then the operator (25) is equivalent to the following

$$\mathcal{K} f(X) = \int_{-1/\epsilon}^{1/\epsilon} K(X, T) f(T) d(T). \tag{26}$$

Suppose $\mathcal{E}_{\rho_\epsilon}$ denotes the ellipse with foci at $\pm\frac{1}{\epsilon}$ and semi-axis sum ρ_ϵ, then $\rho_\epsilon = \frac{\rho}{\epsilon}$. If $\epsilon < 1$ and the kernel function $K(x, y)$ is analytic in $\mathcal{E}_{\rho_\epsilon}$, we have the shaper result.

Theorem 7 *The scaled eigenvalues of the scaled kernel $K(\epsilon x, \epsilon y)$, $\epsilon < 1$, decay in the order $|\lambda_{n+1}| = \mathcal{O}(\rho_\epsilon^{-n}) = \mathcal{O}(\epsilon^n \rho^{-n})$.*

It is observed that the case $\epsilon \to 0$ corresponds to the basis functions tending to flatness; the smallest eigenvalue of the kernel matrix will become smaller.

Similarly, Lemma 3 can be used to obtain similar results as Theorem 5 according to Chebyshev series truncation. The proof involves additional complex analysis results, the reader is directed to [24, 44] for details.

3.3 Separable kernels in $\mathbb{R}^d$

A separable kernel in $\mathbb{R}^d$ can be expressed as the product of multiple kernels, say

$$K(\mathbf{x}, \mathbf{y}) = K_1(x_1, y_1)K_2(x_2, y_2) \cdots K_d(x_d, y_d), \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, x_i, y_i \in \mathbb{R}.$$

Such separable kernels exist and have been considered for a long time, see [43] for example. The famous example is the Gaussian radial basis functions $\exp\left(-\|\mathbf{x} - \mathbf{y}\|^2\right)$.

Consider the simplest separable kernel in the case $d = 2$, if there exist an analytical eigenfunction expansions for each K_i, $i = 1, 2$,

$$K_1(x_1, y_1) = \sum_{m=1}^{\infty} \lambda_m \varphi_m(x_1)\varphi_m^*(y_1) \text{ and } K_2(x_2, y_2) = \sum_{n=1}^{\infty} \mu_n \psi_n(x_2)\psi_n^*(y_2), \tag{27}$$

and further $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\mu_k\}_{k=1}^{\infty}$ are absolute convergent, then according to Mercer theorem [29, p.96]

$$\begin{aligned} & \int\int K_1(x_1, y_1)K_2(x_2, y_2)\varphi_p(y_1)\psi_q(y_2)dy_1dy_2 \\ = & \int\int \varphi_p(y_1)\left(\sum_{m=1}^{\infty}\lambda_m\varphi_m(x_1)\varphi_m^*(y_1)\right)\psi_q(y_2)\left(\sum_{n=1}^{\infty}\mu_n\psi_n(x_2)\psi_n^*(y_2)\right)dy_1dy_2 \\ = & \lambda_p\mu_q\varphi_p(x_1)\psi_q(x_2). \end{aligned}$$

Note that when the kernel is symmetric, then $\varphi_m^* = \varphi_m$ and $\psi_n^* = \psi_n$. In this way, one can show that if λ is an eigenvalue of $K_1(x_1, y_1)$ and μ is an eigenvalue of $K_2(x_2, y_2)$, then $\lambda\mu$ is an eigenvalue of $K(\mathbf{x}, \mathbf{y})$.

Suppose the eigenvalues of $K_i(x_i, y_i)$, $i = 1, 2$ are in the order $\mathcal{O}(1)$, $\mathcal{O}\left(R^{-1}\right)$, $\mathcal{O}\left(R^{-2}\right)$, ..., for some $R > 1$, then the eigenvalues of $K(\mathbf{x}, \mathbf{y})$ are expected in the following order: 1 in order $\mathcal{O}(1)$, 2 in order $\mathcal{O}\left(R^{-1}\right)$, 3 in order $\mathcal{O}\left(R^{-2}\right)$ and $m+1$ in order $\mathcal{O}(R^{-m})$. See Table 1 for illustration. Such a case can happen, for example, with a separable kernel on a square with an equally spaced $N \times N$ regular mesh. If $K_1(x_1, y_1)$ is the kernel in the horizontal direction, and $K_2(x_2, y_2)$ is the kernel in the vertical direction, then their corresponding discrete kernel matrices should have the same eigenvalues because their relative distances in each direction are identical. Similarly, one can show in $\mathbb{R}^d$ that, on a $d \times d$ cube with a regular equally spaced mesh, the discrete kernel matrix of a separable kernel are supposed to have $\binom{k+d}{d} = \frac{(k+d)!}{k!(d)!}$ eigenvalues in similar order. This number is equal to the number of terms in the expansions of

$$(x_1 + x_2 + \ldots + x_d)^k \tag{28}$$

and will become more clear as our discussion proceeds. We first use the analytic and numerical results on Gaussian radial basis functions to verify the result.

Example 1 Consider a weighted inner product defined by

$$(u(x), v(x)) = \int u(x)v(x)w(x)dx,$$

where $w(x) > 0$ is a density function, then the eigenvalues of a kernel are defined by

$$\int K(x, y)w(y)\psi(y)dy = \lambda\psi(x).$$

With the weighted function $w(x) = \frac{\alpha}{\sqrt{\pi}}e^{-\alpha^2x^2}$, $\alpha > 0$, the analytic eigenvalues of the Gaussian kernel $e^{-\varepsilon^2(x-y)^2}$ in $\mathbb{R}$ are given for $n = 1, 2, 3, \cdots$ by

Table 1 The order of eigenvalues of tensor product Kernel in R^2

	1	R^{-1}	R^{-2}	R^{-3}	R^{-4}	R^{-5}	$\cdots$
1	1	R^{-1}	R^{-2}	R^{-3}	R^{-4}	R^{-5}	$\cdots$
R^{-1}	R^{-1}	R^{-2}	R^{-3}	R^{-4}	R^{-5}	R^{-6}	$\iddots$
R^{-2}	R^{-2}	R^{-3}	R^{-4}	R^{-5}	R^{-6}	$\iddots$	$\iddots$
R^{-3}	R^{-3}	R^{-4}	R^{-5}	R^{-6}	$\iddots$	$\iddots$	$\iddots$
R^{-4}	R^{-4}	R^{-5}	R^{-6}	$\iddots$	$\iddots$	$\iddots$	$\iddots$
R^{-5}	R^{-5}	R^{-6}	$\iddots$	$\iddots$	$\iddots$	$\iddots$	$\iddots$
$\vdots$	$\iddots$	$\iddots$	$\iddots$	$\iddots$	$\iddots$	$\iddots$	$\iddots$

$$\lambda_n = \sqrt{\frac{\alpha^2}{\alpha^2+\delta^2+\varepsilon^2}}\left(\frac{1}{1+\delta^2/\varepsilon^2+\alpha^2/\varepsilon^2}\right)^{n-1},$$

where

$$\delta^2 = \frac{\alpha^2}{2}\left(\sqrt{1+\left(\frac{2\varepsilon}{\alpha}\right)^2}-1\right).$$

See [48][29, p.97] for details. Clearly, $1+\delta^2/\varepsilon^2+\alpha^2/\varepsilon^2 > 1$ is the parameter ρ in Theorem 6. For the multivariate Gaussian kernel

$$K(\mathbf{x},\mathbf{y}) = e^{-\varepsilon_1^2(x_1-y_1)^2-\cdots-\varepsilon_d^2(x_d-y_d)^2},$$

we only consider the case $\varepsilon_1 = \cdots \varepsilon_d = \varepsilon$, according to formula (3.6a) in [9, p.A742], the eigenvalues of multivariate Gaussian radial basis function under the weighted inner product with the above weight function can be written as

$$\lambda_{\mathbf{n}} = \prod_{j=1}^{d}\lambda_{n_j} = \left(\frac{\alpha^2}{\alpha^2+\delta^2+\varepsilon^2}\right)^d\left(\frac{1}{1+\delta^2/\varepsilon^2+\alpha^2/\varepsilon^2}\right)^{\sum_{j=1}^{d} n_j-d}. \qquad (29)$$

Then if $\sum_{j=1}^{d} n_j = k$, then there are $\binom{k+d}{d} = \frac{(k+d)!}{k!d!}$ possible combinations of $n_1,\cdots,n_d$.

3.4 Infinite smooth kernels in $\mathbb{R}^d$ with flat limit

The discussion above is primarily based on the properties of the continuous integral operator and pays little attention to the location of the underlying interpolation points. This section concentrates on the discrete form, in particular, on those cases where the shape parameter of the underlying radial basis function tends to 0. These radial basis functions with shape parameter $\varepsilon \to 0$ are called radial basis function with flat limit, and have received a lot attention in recent years [11–13]. The flatness brings the advantage of asymptotic techniques in ε.

Another technique used to investigate the spectral distribution of high dimensional kernels with flatness limit is due to Schaback [42, Theorem 6,p.307]; it doesn't require the kernel function to be separable. In [42], the author focused on a geometric property of the interpolation points and has not concerned explicit results on the distribution of eigenvalues. The basic idea is to construct a sequence of nested subspaces, and investigate quadratic forms corresponding to the interpolation matrix (4). We combine these techniques with other linear algebra results to show the spectral distribution of scaled infinitely smooth radial functions. To prove the results, we introduce the following notation and lemma.

Let $\{p_1, p_2, \cdots, p_Q\}$ be a basis set for $\pi_{k-1}\left(\mathbb{R}^d\right)$, the multivariate polynomial space of degree at most $k-1$, and denote by $\boldsymbol{P}_k$ the matrix with entries $\left(p_j\left(\mathbf{x}_i\right)\right)_{1\le i\le N, 1\le j\le Q}$, then for any given data set $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\} \subset \Omega \subset \mathbb{R}^d$, we have $\ker\left(\boldsymbol{P}_{k+1}^T\right) \subseteq \ker\left(\boldsymbol{P}_k^T\right)$ for $k = 1, 2, 3, \cdots$. Further denote $\ker(\boldsymbol{P}_0^T) = \mathbb{R}^N$, then for any finite set $\mathcal{X}$, there is a positive integer $\mu(\mathcal{X})$ such that

$$\emptyset = \ker\left(\boldsymbol{P}_{\mu(\mathcal{X})}^T\right) \subseteq \cdots \subseteq \ker\left(\boldsymbol{P}_{k+1}^T\right) \subseteq \ker\left(\boldsymbol{P}_k^T\right) \subseteq \cdots \subseteq \ker\left(\boldsymbol{P}_1^T\right) \subset \ker\left(\boldsymbol{P}_0^T\right). \tag{30}$$

$\mu(\mathcal{X})$ can be viewed as a geometric property of the data set $\mathcal{X}$, see [42] for details.

Let $\boldsymbol{\Delta}$ be the Euclidean distance matrix with entries $\left(\|\mathbf{x}_i - \mathbf{x}_j\|\right)_{1\le i,j\le N}$ and $\boldsymbol{\Delta}^k = \left(\|\mathbf{x}_i - \mathbf{x}_j\|^k\right)_{1\le i,j\le N}$ denote the element-wise power of $\boldsymbol{\Delta}$, then there is a well known result due to Micchelli.

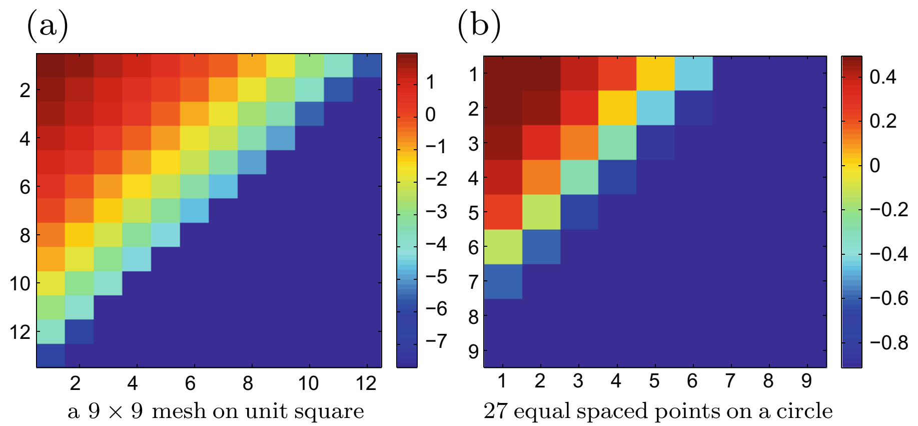

Fig. 1 Eigenvalues of kernel matrices of Gaussian radial basis function on different meshes. The shape parameter $\varepsilon = 5$. We first sort the eigenvalues in descend order, then map the 1 dimensional array to an 2 dimensional diagram by the inverting Cantor pairing function (http://mathworld.wolfram.com/PairingFunction.html). The figure shows the log scale of the eigenvalues by Matlab function `imagesc`. The dark blue area in the bottom-right corners is empty with `NaN`. Panel 1a demonstrates that, on the regular mesh, the eigenvalues in the same order are grouped by $1, 2, 3, 4, \cdots$; while panel 1b shows the eigenvalues are grouped by $1, 2, 2, 2, \cdots$

Lemma 5 (Micchelli [25]) *If* $\sum_{i=1}^{n} \alpha_i p(\mathbf{x}_i) = 0$ *for all* $p \in \pi_{k-1}\left(\mathbb{R}^d\right)$ *, then the following quadratic form satisfies* $(-1)^k \alpha^T \boldsymbol{\Delta}^{2k} \alpha \geq 0$*, where equality holds if and only if* $\sum_{i=1}^{n} \alpha_i p(x_i) = 0$*, for all* $p \in \pi_k\left(\mathbb{R}^d\right)$.

For convenience, we denote the interpolation matrix (4) corresponding to the scaled radial basis function $\phi(\varepsilon r) = g\left(\varepsilon^2 r^2\right)$ by $A_{\phi_\varepsilon, \mathcal{X}} = \phi(\varepsilon \boldsymbol{\Delta}) = g\left(\varepsilon^2 \boldsymbol{\Delta}^2\right)$. The function g is taken to be infinitely smooth and has a Taylor expansion at the origin.

Theorem 8 *Let* $\boldsymbol{A}_{\phi_\varepsilon, \mathcal{X}}$ *be an interpolation matrix corresponding to an infinitely smooth radial basis function* $\phi_\varepsilon(r) = \phi(\varepsilon r) = g\left((\varepsilon r)^2\right)$*. If* g *has a convergent Taylor expansion near the origin in the real line, then, when* $\varepsilon \to 0$*, there are exactly* $\dim\left(\ker\left(\boldsymbol{P}_k^T\right)\right) - \dim\left(\ker\left(\boldsymbol{P}_{k+1}^T\right)\right)$ *eigenvalues behaving like* ε^{2k}*, for* $0 \leq k \leq \mu(\mathcal{X}) - 1$*, where* $\boldsymbol{P}_k^T$ *and* $\mu(\mathcal{X})$ *are defined in* (30)*. Furthermore, if each* $\boldsymbol{P}_k$ *is full rank, then there are exactly* $\binom{k+d-1}{d-1} = \frac{(k+d-1)!}{k!(d-1)!}$ *eigenvalues behaving like* ε^{2k}.

Proof Since the function g has a convergent Taylor expansion near the origin, then we can write the entries of the interpolation matrix as a sum of the element powers of the distance matrix $\boldsymbol{\Delta}$,

$$\boldsymbol{A}_{\phi_\varepsilon, \mathcal{X}} = g\left((\varepsilon \boldsymbol{\Delta})^2\right) = \sum_{\ell=0}^{\infty} \frac{g^{(\ell)}(0)}{\ell!} (\varepsilon \boldsymbol{\Delta})^{2\ell}. \tag{31}$$

Let $m = \mu(\mathcal{X}) - 1$, $\boldsymbol{\alpha} \in \ker\left(\boldsymbol{P}_m^T\right)$, then by Lemma 5 $\boldsymbol{\alpha}^T \boldsymbol{\Delta}^{2\ell} \boldsymbol{\alpha} = 0$, for $\ell \leq m$. Therefore we have the following quadratic form

$$\boldsymbol{\alpha}^T \boldsymbol{A}_{\phi_\varepsilon, \mathcal{X}} \boldsymbol{\alpha} = \varepsilon^{2m} \frac{f^{(m)}(0)}{m!} \boldsymbol{\alpha}^T \boldsymbol{\Delta}^{2m} \boldsymbol{\alpha} + \sum_{\ell=\mu(X)}^{\infty} \varepsilon^{2\ell} \frac{f^{(\ell)}(0)}{\ell!} \boldsymbol{\alpha}^T \boldsymbol{\Delta}^{2\ell} \boldsymbol{\alpha} \tag{32}$$

decays like ε^{2m} as $\varepsilon \to 0$. According the Courant-Fischer's minimum-maximum principle [19, p.179], $A_{\phi_\varepsilon, \mathcal{X}}$ has at least $\dim\left(\ker(\boldsymbol{P}_m^T)\right)$ eigenvalues which decay at least as fast as ε^{2m}.

Further denote the space $\mathcal{M}_k \subseteq \ker\left(\boldsymbol{P}_k^T\right)$ and $\mathcal{M}_k \perp \ker\left(\boldsymbol{P}_{k+1}^T\right)$, then

$$\dim(\mathcal{M}_k) = \dim\left(\ker\left(\boldsymbol{P}_k^T\right)\right) - \dim\left(\ker\left(\boldsymbol{P}_{k+1}^T\right)\right).$$

Applying the same argument on the space $\mathcal{M}_k$, then we can find there are at least $\dim(\mathcal{M}_k)$ eigenvalues decaying at least as ε^{2k} in the subspace $\mathcal{M}_k$. Since $\sum_{k=1}^{m} \dim(\mathcal{M}_k) = N$, thus there are exactly $\dim(\mathcal{M}_k)$ eigenvalues decaying like ε^{2k} when $\varepsilon \to 0$.

If every $\boldsymbol{P}_k$ is full rank, then $\mathrm{rank}(\boldsymbol{P}_k) = \dim\left(\pi_{k-1}\left(\mathbb{R}^d\right)\right) = \binom{k-1+d}{d}$, and

$$\dim(\mathcal{M}_k) = \dim\left(\pi_k\left(\mathbb{R}^d\right)\right) - \dim\left(\pi_{k-1}\left(\mathbb{R}^d\right)\right) = \frac{(k+d-1)!}{k!(d-1)!}.$$

□

Note that the number $\binom{k-1+d}{d}$ is the number of terms in the expansion of $(x_1 + \cdots + x_d)^{k-1}$, which is consistent with the discussion above, whereas the term $\dim\left(\pi_k\left(\mathbb{R}^d\right)\right) - \dim\left(\pi_{k-1}\left(\mathbb{R}^d\right)\right)$ depends on the geometric property of the interpolation data points. The geometric property is not easy to identify in general. For example, when the interpolation points in a circle with radius 1, then $\dim\left(\pi_k\left(\mathbb{R}^2\right)\right) - \dim\left(\pi_{k-1}\left(\mathbb{R}^2\right)\right) = 2$ for $k = 1, 2, 3$, because $\{1, x, y\}$ are independent, while for $\left\{x^2, xy, y^2\right\}$, we have $y^2 = 1 - x^2$, for $\left\{x^3, y^3, x^2y, y^2x\right\}$, we have $x^2y = y - y^3, xy^2 = x - x^2$. It was observed that when the eigenvalues kernel matrices of Gaussian radial basis function on a circle are grouped in different orders, the numbers of each group in descending order are $1, 2, 2, 2, 2, \cdots$. The reader is directed to [14, p.389] for more scenarios. Figure 1b illustrates the eigenvalues of a kernel matrix of Gaussian radial basis function on a circle.

4 Discussion

It is noted that there are existing results on the conditioning issue of kernel matrices related to radial basis function. As mentioned, they mainly focused on the smallest eigenvalues and on the condition number. By contrast, here we give simple methods to estimate every eigenvalue, or the distribution of the eigenvalues.

4.1 Comparison with other results

Noteworthy results on lower bounds for the smallest eigenvalue include [1, 3, 27, 28, 40, 41]. The technique employed to prove these results can be summarized as using Fourier transform techniques to estimate a quadratic form in a subspace. There is also one paper which discusses an upper bound on the smallest eigenvalues of the kernel matrices [2], in which the authors construct a special vector, and then use the vector to estimate the inverse norm of the interpolation matrix via some sophisticated results on divided difference formula. Here, we use quite a different way to prove the upper bound of every eigenvalue of the continuous operator. This technique is based on the Weyl-Courant minimax principle via approximating an infinite dimensional square integral operator by finite Fourier or Chebyshev truncations. After finding the connections between the integral equations and the interpolation problem, the main results are quickly derived. This method can be used in high dimensional space by employing multivariate Fourier approximation.

We mention that the results of Theorem 5 are consistent with previous results. It is likely we get more accurate estimates on the smallest eigenvalue of the kernel matrix. For example for the compactly supported Wendland function, $\phi_{d,k} \in C^{2k}$, the lower bound of the smallest eigenvalues is $\lambda_{\min} \geq Cq^{2k+1}$ [46, p.214] for some constant C, where the q is the so-called separation distance. For equal space sampling in the interval $[-1, 1]$ with the sampling points arranged as $-1 = x_0 < x_1 < \cdots < x_n = 1$, $q = \frac{1}{2}\min_{i=j}|x_i - x_j| = \frac{1}{n}$, so that $\lambda_{\min} \geq Cn^{-2k-1}$. The proof of Theorem 5 shows that $\lambda_{\min} = o\left(n^{-2k}\right)$. Therefore, we obtain both a lower and upper bound on the smallest eigenvalue for large n, say, $C_1n^{-2k-1} \leq \lambda_{min} \leq C_2n^{-2k}$.

4.2 Connection with QR-RBF

As seen, especially in Theorem 3, the eigenvalues of a kernel have a close relationship with coefficients of the orthogonal expansions of the kernel functions. The eigenvalues of analytic kernels are in the same order as Chebyshev coefficients in the Chebyshev truncations. The recent RBF-QR method which aims to compute Gaussian radial basis function interpolants with flat limit is also based on orthogonal expansions [11]. In this method there arises a diagonal scaling matrix with entries which are a part of the coefficients in the orthogonal expansion. Recalling that the eigenvalues of integral operators with periodic kernel are closely related to the Fourier coefficients, for multivariate cases, one might expect, the number of the eigenvalues in the same order is the same as the number of the coefficients of orthogonal expansions with the same (polynomial) order. The number of entries in the order ε^{2k} in the scaling matrix in the RBF-QR methods happens to be $\dim\left(\pi_k\left(\mathbb{R}^d\right)\right) - \dim\left(\pi_{k-1}\left(\mathbb{R}^d\right)\right)$, which is the number of the eigenvalues in the corresponding order. This is unlikely to be a coincidence.

The spectral distribution can supply information on how to choose the right diagonal scaling matrix in the RBF-QR method. It can also be used to investigate the smoothing effects of kernel matrices [39] and is closely related to several other computing issues which go beyond discussion in this paper.

Acknowledgements We would like to thank the reviewers for their valuable suggestions and comments which improve the shape of this paper.

Appendix

Hints of proof of theorem 3

Consider

$$\lambda u(x) = \int_{-\pi}^{\pi} K(x-y)u(y)dy \tag{33}$$

where $K(x)$ is integrable and periodic of 2π. Let

$$K(x) \sim \sum_{-\infty}^{\infty} \hat{K}_n e^{-inx}, \text{ where } \hat{K}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} K(x)e^{-inx}dx,$$
$$u(x) \sim \sum_{-\infty}^{\infty} \hat{u}_n e^{-inx}, \text{ where } \hat{u}_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(x)e^{-inx}dx.$$

Multiply e^{-inx} on both side of (33), we have

$$\begin{aligned}\lambda\hat{u}_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} u(x)e^{-inx}dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} K(x-y)e^{-in(x-y)}e^{-iny}u(y)dydx = 2\pi\hat{K}_n\hat{u}_n.\end{aligned}$$

References

1. Ball, K.: Eigenvalues of Euclidean distance matrices. J. Approx. Theory **68**(1), 74–82 (1992). doi:10.1016/0021-9045(92)90101-S
2. Ball, K., Sivakumar, N., Ward, J.D.: On the sensitivity of radial basis interpolation to minimal data separation distance. Constr. Approx. **8**(4), 401–426 (1992). doi:10.1007/BF01203461
3. Baxter, B.J.C.: Norm estimates for inverses of Toeplitz distance matrices. J. Approx. Theory **79**(2), 222–242 (1994). doi:10.1006/jath.1994.1126
4. Bernstein, S.: Sur la valeur les recherches récentes relatives à la meilleure approximation des fonctions continues par des polynômes. In: Proceedings 5th. Intern. Math. Congress, v. 1, vol. 1, pp. 256–266 (1912)
5. Buescu, J., Paixão, A.: Eigenvalue distribution of positive definite kernels on unbounded domains. Integr. Equ. Oper. Theory **57**(1), 19–41 (2007). doi:10.1007/s00020-006-1445-1
6. Buescu, J., Paixão, A.C.: Eigenvalue distribution of Mercer-like kernels. Math. Nachr. **280**(9-10), 984–995 (2007). doi:10.1002/mana.200510530
7. Chang, C.H., Ha, C.W.: On eigenvalues of differentiable positive definite kernels. Integr. Equ. Oper. Theory **33**(1), 1–7 (1999). doi:10.1007/BF01203078
8. Cochran, J.A.: The analysis of linear integral equations. McGraw-Hill Book Co., New York (1972). McGraw-Hill Series in Modern Applied Mathematics
9. Fasshauer, G.E., McCourt, M.J.: Stable evaluation of Gaussian radial basis function interpolants. SIAM J. Sci. Comput. **34**(2), A737–A762 (2012). doi:10.1137/110824784
10. Ferreira, J.C., Menegatto, V.A.: Eigenvalues of integral operators defined by smooth positive definite kernels. Integr. Equ. Oper. Theory **64**(1), 61–81 (2009). doi:10.1007/s00020-009-1680-3
11. Fornberg, B., Larsson, E., Flyer, N.: Stable computations with Gaussian radial basis functions. SIAM J. Sci. Comput. **33**(2), 869–892 (2011). doi:10.1137/09076756X
12. Fornberg, B., Piret, C.: A stable algorithm for flat radial basis functions on a sphere. SIAM J. Sci. Comput. **30**(1), 60–80 (2007/08). doi:10.1137/060671991
13. Fornberg, B., Wright, G., Larsson, E.: Some observations regarding interpolants in the limit of flat radial basis functions. Comput. Math. Appl. **47**(1), 37–55 (2004). doi:10.1016/S0898-1221(04)90004-1
14. Fornberg, B., Zuev, J.: The Runge phenomenon and spatially variable shape parameters in RBF interpolation. Comput. Math. Appl. **54**(3), 379–398 (2007). doi:10.1016/j.camwa.2007.01.028
15. Fredholm, I.: Sur une classe d'équations fonctionnelles. Acta Math. **27**(1), 365–390 (1903). doi:10.1007/BF02421317
16. Ha, C.W.: Eigenvalues of differentiable positive definite kernels. SIAM J. Math. Anal **17**(2), 415–419 (1986). doi:10.1137/0517031
17. Hille, E., Tamarkin, J.D.: On the characteristic values of linear integral equations. Acta Math. **57**(1), 1–76 (1931). doi:10.1007/BF02403043
18. de Hoog, F.R.: Review of Fredholm equations of the first kind, Application and numerical solution of integral equations (Proc. Sem., Australian Nat. Univ., Canberra, 1978), *Monographs Textbooks Mech. Solids Fluids: Mech, Anal.*, vol. 6, pp. 119–134. Sijthoff & Noordhoff, Alphen aan den Rijn (1980)
19. Horn, R.A., Johnson, C.R.: Matrix analysis. Cambridge University Press, Cambridge (1990). Corrected reprint of the 1985 original
20. Katznelson, Y.: An introduction to harmonic analysis. John Wiley & Sons Inc., New York (1968)
21. Kress, R. Linear integral equations, Applied Mathematical Sciences, 2nd edn., vol. 82. Springer, New York (1999)
22. Larsson, E., Fornberg, B.: Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. Comput. Math. Appl. **49**(1), 103–130 (2005). doi:10.1016/j.camwa.2005.01.010
23. Larsson, E., Lehto, E., Heryudono, E., Fornberg, B.: Stable computation of differentiation matrices and scattered node stencils based on Gaussian radial basis functions, Tech. Rep. 2012-020, Department of Information Technology, Uppsala University (2012)
24. Little, G., Reade, J.B.: Eigenvalues of analytic kernels. SIAM J. Math. Anal. **15**(1), 133–136 (1984). doi:10.1137/0515009
25. Micchelli, C.A.: Interpolation of scattered data: distance matrices and conditionally positive definite functions. Constr. Approx. **2**(1), 11–22 (1986). doi:10.1007/BF01893414

26. Narcowich, F., Sivakumar, N., Ward, J.: On condition numbers associated with radial-function interpolation. J. Math. Anal. Appl. **186**(2), 457–485 (1994). doi:10.1006/jmaa.1994.1311
27. Narcowich, F., Ward, J.: Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices. J. Approx. Theory **69**(1), 84–109 (1992). doi:10.1016/0021-9045(92)90050-X
28. Narcowich, F.J., Ward, J.D.: Norms of inverses for matrices associated with scattered data. In: Curves and surfaces (Chamonix-Mont-Blanc, 1990), pp. 341–348. Academic Press, Boston (1991)
29. Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning, vol. 1. MIT press Cambridge, MA (2006)
30. Reade, J.B.: Asymptotic behaviour of eigenvalues of certain integral equations. Proc. Edinburgh Math. Soc. (2) **22**(2), 137–144 (1979). doi:10.1017/S0013091500016254
31. Reade, J.B.: Eigenvalues of Lipschitz kernels. Math. Proc. Cambridge Philos. Soc. **93**(1), 135–140 (1983). doi:10.1017/S0305004100060412
32. Reade, J.B.: Eigenvalues of positive definite kernels. SIAM J. Math. Anal. **14**(1), 152–157 (1983). doi:10.1137/0514012
33. Reade, J.B.: Eigenvalues of positive definite kernels. II. SIAM J. Math. Anal. **15**(1), 137–142 (1984). doi:10.1137/0515010
34. Reade, J.B.: Eigenvalues of smooth kernels. Math. Proc. Cambridge Philos. Soc. **95**(1), 135–140 (1984). doi:10.1017/S0305004100061375
35. Reade, J.B.: On the sharpness of Weyl's estimate for eigenvalues of smooth kernels. SIAM J. Math. Anal. **16**(3), 548–550 (1985). doi:10.1137/0516040
36. Reade, J.B.: Positive definite $C^p p$ kernels. SIAM J. Math. Anal. **17**(2), 420–421 (1986). doi:10.1137/0517032
37. Reade, J.B.: On the sharpness of Weyl's estimates for eigenvalues of smooth kernels. II. SIAM J. Math. Anal. **19**(3), 627–631 (1988). doi:10.1137/0519044
38. Reade, J.B.: Eigenvalues of smooth positive definite kernels. Proc. Edinburgh Math. Soc. (2) **35**(1), 41–45 (1992). doi:10.1017/S0013091500005307
39. Renaut, R.A., Zhu, S.: Application of fredholm integral equations inverse theory to the radial basis function approximation problem. Tech Rep. 1630. The University of Oxford (2012)
40. Schaback, R.: Lower bounds for norms of inverses of interpolation matrices for radial basis functions. J. Approx. Theory **79**(2), 287–306 (1994). doi:10.1006/jath.1994.1130
41. Schaback, R.: Error estimates and condition numbers for radial basis function interpolation. Adv. Comput. Math. **3**(3), 251–264 (1995). doi:10.1007/BF02432002
42. Schaback, R.: Multivariate interpolation by polynomials and radial basis functions. Constr. Approx. **21**(3), 293–317 (2005). doi:10.1007/s00365-004-0585-2
43. Smithies, F.: The eigenvalue and singular values of integral equations. Proc. London Math. Soc. **43**, 255–279 (1937)
44. Trefethen, L.: Approximation Theory and Approximation Practice. Applied Mathematics. Society for Industrial and Applied Mathematics (2013). URL http://books.google.co.uk/books?id=En41UGQ6YXsC
45. Wathen, A.J.: Realistic eigenvalue bounds for the Galerkin mass matrix. IMA J. Numer. Anal. **7**(4), 449–457 (1987). doi:10.1093/imanum/7.4.449
46. Wendland, H.: Scattered data approximation, *Cambridge Monographs on Applied and Computational Mathematics*, vol. 17. Cambridge University Press, Cambridge (2005)
47. Weyl, H.: Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). Mathematische Annalen **71**, 441–479 (1912).
48. Zhu, H., Williams, C., Rohwer, R., Morciniec, M.: Gaussian regression and optimal finite dimensional linear models. In: Bishop, C. (ed.) Neural Networks and Machine Learning. Springer, Berlin (1998)