Application of Fredholm integral equations inverse theory to the radial basis function approximation problem☆

Rosemary A. Renaut[a,1,*], Shengxin Zhu[b,c,2]

[a]*School of Mathematical & Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804, USA*
[b]*Oxford Center for Collaborative and Applied Mathematics, 24-29 St Giles, OX1 3LB, Oxford, UK*
[c]*Numerical Analysis Group, The University of Oxford, 24-29 St Giles, OX1 3LB, Oxford, UK*

Abstract

This paper reveals and examines the relationship between the solution and stability of Fredholm integral equations and radial basis function approximation or interpolation. The underlying system (kernel) matrices are shown to have a smoothing property which is dependent on the choice of kernel. Instead of using the condition number to describe the ill-conditioning, hence only looking at the largest and smallest singular values of the matrix, techniques from inverse theory, particularly the Picard condition, show that it is understanding the exponential decay of the singular values which is critical for interpreting and mitigating instability. Results on the spectra of certain classes of kernel matrices are reviewed, verifying the exponential decay of the singular values. Numerical results illustrating the application of integral equation inverse theory are also provided and demonstrate that interpolation weights may be regarded as samplings of a weighted solution of an integral equation. This is then relevant for mapping from one set of radial basis function centers to another set. Techniques for the solution of integral equations can be further exploited in future studies to find stable solutions and to reduce the impact of errors in the data.

Keywords: Kernel, Radial basis function, Picard condition, first kind Fredholm integral, Spectral distribution

2008 MSC: 45A15, 45A25, 65F10

1. Introduction

Radial basis functions (RBFs) are increasingly of interest for the solution of practical problems described by partial differential equations. A recent review of Flyer and Fornberg [1] demonstrates the promise of RBFs for numerically approximating sophisticated physical phenomena, particularly with the use of the novel QR-RBF algorithm [2]. Still, questions on how to effectively and efficiently determine stable solutions for RBF approximation remain. Here a new perspective on this interesting problem is presented which reemphasizes the significance of the spectral decomposition of the associated systems described by RBF interpolation/approximation.

Given the complete eigen-system $\{(\lambda_\ell, \mathbf{v}_\ell)\}_{\ell=1}^N$ for a matrix Φ, any $\mathbf{x} \in \mathbb{R}^N$ can be written as $\mathbf{x} = \sum_{\ell=1}^N \alpha_\ell \mathbf{v}_\ell$ yielding $\mathbf{y} = \Phi\mathbf{x} = \sum_{\ell=1}^N \alpha_\ell \lambda_\ell \mathbf{v}_\ell$. It is thus immediate that the contribution of $\mathbf{v}_\ell$ to $\mathbf{y}$ is less than that to $\mathbf{x}$ when $|\lambda_\ell| < 1$. Suppose now that this eigen-system is ordered with $\{|\lambda_\ell|\}_{\ell=1}^N$ arranged in a descending order, and such that this ordering also corresponds to an ordering of the basis vectors $\{\mathbf{v}_\ell\}_{\ell=1}^N$ from low to high frequency. Here, a vector is said to be a high frequency vector when, regarding the vector entries as the sampling of an underlying function, the function is highly oscillating. Then, the matrix Φ is said to have a *smoothing effect* when the vector $\mathbf{y} = \Phi\mathbf{x}$ is *smoother* than $\mathbf{x}$ in the sense that *high frequency information* in $\mathbf{x}$ is damped and, correspondingly, the low frequency content of $\mathbf{y}$ is more dominant than that of $\mathbf{x}$. Immediately, observe that Φ has a smoothing effect only if the decay of $\{|\lambda_\ell|\}_{\ell=1}^N$ is fast and the eigenvalues of small magnitude correspond to the high frequency eigenvectors. In the context of this paper, such eigenvalues are referred to as *smoothing eigenvalues*; their distribution determines the smoothing properties of Φ.

For the forward problem the smoothing properties of Φ can be useful; for example it is often beneficial to smooth a given measured data signal $\mathbf{x}$ by applying a *moving average technique* [3, 4], [5, chap 2]. The simplest moving average, in which each data point is replaced by the mean of its k nearest neighbours, generates a non-negative matrix Φ where all the non-zero entries are $1/k$ and

[☆]This research is supported by Award no KUK-C1-013-04, made by King Abdullah University of Science of Technology.

[*]Principal corresponding author

Email addresses: renaut@math.asu.edu (Rosemary A. Renaut), shengxin.zhu@maths.ox.ac.uk (Shengxin Zhu)

[1]Rosemary Renaut was supported by an OCCAM Fellowship in Spring 2012

[2]Thanks to OCCAM studentship to support Shengxin Zhu's research.

each row sum is equal to 1. According to the Perron-Frobenius theorem [6, chap 2], the lowest frequency eigenvector of Φ, with all elements identical, corresponds to the largest eigenvalue (Perron root) 1. All other higher frequency eigenvectors, have at least one entry with opposite sign to the others, and correspond to smoothing eigenvalues $|\lambda| < 1$. This smoothing property is, however, a hindrance for the inverse, or data recovery, problem. In particular, suppose that Φ^{-1} exists, then it has a complete eigen-system $\{(\frac{1}{\lambda_\ell}, \mathbf{v}_\ell)\}_{\ell=1}^{N}$ and $\mathbf{z} = \Phi^{-1}\mathbf{x} = \sum_{\ell=1}^{N} \frac{\alpha_\ell}{\lambda_\ell}\mathbf{v}_\ell$. Whereas a high frequency eigenvector $\mathbf{v}_\ell$ is smoothed by application of Φ, it will be amplified by application of Φ^{-1} particularly when $|\lambda_\ell| \ll 1$, and any perturbation in the $\mathbf{v}_\ell$ component will be correspondingly amplified rather than damped when forming $\mathbf{z}$ rather than $\mathbf{y}$.

The so-called *smoothing kernels* are typically discussed in the context of the solution of ill-posed inverse problems, and particularly in the context of the solution of first kind Fredholm integral equations [7, 8]. The focus in this paper is on the kernel matrices that arise in using RBFs for interpolation/approximation and extends results concerning the stability of the solutions of discrete first kind Fredholm integral equations to RBF approximation. For this problem, Micchelli has proved that the RBF interpolation (kernel) matrices are always invertible provided that the interpolation points are distinct [9]. But it is well recognized that the difficulty of solving systems determined by such kernel matrices, Φ, depends significantly on the condition of Φ [10–13]. Here it is emphasized how the smoothing properties of Φ, which depend on the smoothing properties of the RBFs and the location of the interpolation points, are directly related to the difficulty of solving the RBF interpolation/approximation problem. The stability of the solution found by solving the inverse problem, finding $\mathbf{z}$ given Φ and $\mathbf{x}$, is investigated using the discrete Picard condition [7, 14, p.37], which is an extension of the continuous Picard condition that is pertinent for examining the stability of the solution of first kind Fredholm integral equations [8]. To show how these results are relevant, the connection between the RBF interpolation problem and the solution of an ill-posed integral equation of the first kind is now established.

1.1. The connection between the radial basis function interpolation problem and integral equations

Given observational data $\{f_k : k = 1, \ldots, M\}$ of a function $f(\mathbf{x})$ at scattered data points $\{\mathbf{x}_k \in \mathbb{R}^d : k = 1, \ldots, M\}$, an approximating function $s(\mathbf{x}) \approx f(\mathbf{x})$ is defined by

$$s(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j \phi(\mathbf{x}, \mathbf{y}_j). \tag{1}$$

For RBFs $\phi(\mathbf{x}, \mathbf{y}) = \phi(||\mathbf{x} - \mathbf{y}||_2)$. Selecting weights $\{\alpha_j\}$ such that $s(\mathbf{x}_k) \approx f(\mathbf{x}_k)$, $k = 1, \ldots, M$ defines the *approximation* problem for finding $\{\alpha_j\}$. The *interpolation* problem, $M = N$, yields the defining system of linear equations

$$\mathbf{f} = \Phi\boldsymbol{\alpha}, \quad \Phi_{kj} = \phi(||\mathbf{x}_k - \mathbf{y}_j||), \quad 1 \leq k, j \leq N. \tag{2}$$

In matrix form

$$\begin{pmatrix} \phi(||\mathbf{x}_1 - \mathbf{y}_1||_2) & \phi(||\mathbf{x}_1 - \mathbf{y}_2||_2) & \cdots & \phi(||\mathbf{x}_1 - \mathbf{y}_N||_2) \\ \phi(||\mathbf{x}_2 - \mathbf{y}_1||_2) & \phi(||\mathbf{x}_2 - \mathbf{y}_2||_2) & \cdots & \phi(||\mathbf{x}_2 - \mathbf{y}_N||_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(||\mathbf{x}_N - \mathbf{y}_1||_1) & \phi(||\mathbf{x}_N - \mathbf{y}_2||_2) & \cdots & \phi(||\mathbf{x}_N - \mathbf{y}_N||_2) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}. \tag{3}$$

Under the assumption $\mathbf{y}_k = \mathbf{x}_k, k = 1, \ldots, N$, $\Phi_{kj} = \phi(||\mathbf{x}_k - \mathbf{x}_j||)$ and matrix Φ is symmetric.

Suppose that $\hat{\phi}(\boldsymbol{\omega})$ is the Fourier transformation of $\phi(||\mathbf{x}||)$ of dimension d, then

$$\phi(||\mathbf{x}||) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \hat{\phi}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \mathbf{x}} d\boldsymbol{\omega}. \tag{4}$$

For any $\boldsymbol{\alpha} \in \mathbb{C}^N$, using the standard notation, $\bar{\alpha}$, for the complex conjugate of α, the following relation holds [15, p.67]:

$$\boldsymbol{\alpha}^H \Phi \boldsymbol{\alpha} = \sum_{j,k=1}^{N} \bar{\alpha}_k \alpha_j \phi(||\mathbf{x}_k - \mathbf{x}_j||) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \hat{\phi}(\boldsymbol{\omega}) \left| \sum_{j=1}^{N} \alpha_j e^{i\boldsymbol{\omega}^T \mathbf{x}_j} \right|^2 d\boldsymbol{\omega}. \tag{5}$$

Therefore if the Fourier transform $\hat{\phi}$ is positive, Φ is symmetric and positive definite (SPD) and thus *non-singular*. Examples of RBFs which have a positive Fourier transform include the Gaussian basis functions $\phi(||\mathbf{x}||_2) = e^{-\epsilon^2 ||\mathbf{x}||_2^2}$, $\epsilon > 0$ which are globally supported [15, p.74] and

Wendland's functions [15, p.129] which are compactly supported. Notice the dependence of the RBF on the *shape* parameter $\epsilon > 0$, explicitly denoted throughout this paper where relevant by $\phi = \phi_\epsilon(\|\mathbf{x}\|_2)$. Non-singular interpolation matrices are also guaranteed under other mild conditions, see for example Micchelli [9]. Indeed it has been demonstrated over recent years that RBFs are of considerable use for the approximation of high dimensional scattered data because the resulting interpolation matrices can often be shown to be non-singular [9, 15]. Despite such theoretical results the resulting linear systems may be highly ill-conditioned, dependent on the shape parameter defining ϕ_ϵ. Discussions and numerical investigations of the ill-conditioning can be found in significant references, for example, [10–13]. In particular, such results demonstrate that Φ may exhibit ill-conditioning of the kind often seen when solving inverse problems; hence theory on the solution of ill-posed integral equations of the first kind, [7, 8, 16–18], can be relevant in the context of the RBF interpolation problem.

Consider the solution of the Fredholm integral equation of the first kind

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

from which we wish to determine the underlying source function $\beta(\mathbf{y})$ given the measurements $f(\mathbf{x})$ and a kernel function $K(\mathbf{x}, \mathbf{y})$. Given the observed data vectors $\mathbf{x}$, $\mathbf{f}$, and applying numerical quadrature at the nodes $\mathbf{y}_j$ yields the discrete interpolation conditions

$$\sum_{j=1}^{N} K(\mathbf{x}_k, \mathbf{y}_j)\beta(\mathbf{y}_j)w_j = f(\mathbf{x}_k), \quad k = 1, \ldots, N \tag{7}$$

for quadrature weights w_j. Defining $\alpha(\mathbf{y}) = w(\mathbf{y})\beta(\mathbf{y})$ leads to a system of equations (2) where $\Phi_{kj} = K(\mathbf{x}_k, \mathbf{y}_j) = K(\mathbf{x}_k, \mathbf{x}_j)$ when the nodes and interpolation points are the same. *The connection between the solution of the RBF approximation problem which defines the weights* $\{\alpha_j\}$ *and the solution of* (6) *which defines the source function* $\alpha(\mathbf{y})$ *is now established for kernel function* $K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}, \mathbf{y})$. Explicitly for ϕ_ϵ the kernel in (6) also depends on ϵ, and is denoted by $K(\epsilon\mathbf{x}, \epsilon\mathbf{y})$, with resulting kernel matrix $\Phi(\epsilon)$.

The paper is organised as follows. The Picard condition, which is commonly used to investigate properties of the solution to (2) for discrete ill-posed problems, is introduced in Section 2.

This analysis depends on the use of the singular value expansion of the kernel $K(\mathbf{x}, \mathbf{y})$ in the continuous case and the singular value decomposition (SVD) for the operator Φ in the discrete case. For a square integrable kernel the stability of the solution $\alpha(\mathbf{y})$ depends on the decay properties of the kernel singular values and on the projection of the given data to the singular value space. Error analysis of the approximation is facilitated also through the use of the SVD in Section 2.3. The singular value spectrum for the underlying operator is thus significant. Section 3 explores the dependence of the spectrum of Φ on the smoothness of the underlying RBFs, demonstrating that the decay rate of the discrete singular values impacts the smoothing properties of Φ. Such results indicate why well-recognised difficulties with solving the RBF interpolation problem are closely related to the underlying shape parameters. Results presented in Section 4 illustrate these properties in the context of solving (2), with respect to both the underlying spectra for some presented examples, but also in terms of the underlying basis for the solution. Moreover, these results suggest that techniques which are standard in regularizing the solutions for (6) can be applied to obtain feasible solutions for (2). This is demonstrated in Section 5 in the novel context of examining the underlying $\boldsymbol{\alpha}$ as a function which thus permits mapping the centers of the RBFs from one set to another whilst maintaining a good approximation of the data. Equivalently, this amounts to using the functional approximation for $\boldsymbol{\alpha}$ to resample for a different set of quadrature weights in (1). Conclusions of the paper and topics for future areas of investigation are given in Section 6.

2. Stability and Error Analysis of the RBF Interpolation Problem

Given the association between the solutions of (2) and (6), we now appeal to standard techniques from the analysis of discrete ill-posed linear problems to study the properties of the solution $\boldsymbol{\alpha}$ and the resulting impact on the the quality of the approximant $s(\mathbf{x})$. To make the paper more self-contained a brief review of the Picard condition now follows; for more details, the reader is directed to [7].

2.1. The Continuous Picard Condition

The Picard condition, proved first by Picard in 1910, [19, p.160], is a necessary and sufficient condition for the existence of a square integrable solution $\beta(\mathbf{y})$ to the Fredholm integral equation of

the first kind (6). It reveals how the solution of an ill-posed problem depends both on the spectrum of the kernel function and on the spectral content of the provided functional data, and is commonly used to evaluate the degree of ill-posedness of an ill-posed problem [7, 8, 16–18]. In the context of the RBF approximation problem, the Picard condition is particularly pertinent in illustrating how the smoothness of a RBF, which is directly related to its shape parameter, immediately impacts the smoothing properties of Φ and thus determines the ill-conditioning of (2).

The well-posedness of solutions to (6) is analyzed using the *singular value expansion* (SVE) [20], [7, p.10]. Consider the one dimensional case: given two functions u, v defined on a domain $\Omega \in \mathbb{R}$, their inner product is defined by

$$\langle u, v \rangle = \int_{\Omega} u(x)v(x)dx. \tag{8}$$

Then a singular value system $\{\sigma_\ell, v_\ell, u_\ell\}$ of the *square integrable* kernel $K(x, y)$ is defined by

$$\begin{cases} \int_{\Omega} K(x, y)v_\ell(y)dy = \sigma_\ell u_\ell(x), \\ \langle u_\ell, u_j \rangle = \langle v_\ell, v_j \rangle = \delta_{\ell j}, \end{cases} \tag{9}$$

and yields the SVE for K

$$K(x, y) = \sum_{\ell=1}^{\infty} \sigma_\ell u_\ell(x)v_\ell(y). \tag{10}$$

Now suppose (10) is an absolutely and uniformly convergent expansion for $K(x, y)$, then

$$\int_{\Omega} K(x, y)\beta(y)dy = \int_{\Omega} \sum_{\ell=1}^{\infty} \sigma_\ell u_\ell(x)v_\ell(y)\beta(y)dy = \sum_{\ell=1}^{\infty} \sigma_\ell \langle v_\ell(y), \beta(y) \rangle u_\ell(x) = f(x). \tag{11}$$

On the other hand $\{u_\ell(x)\}_{\ell=1}^{\infty}$ is an orthogonal basis in $L_2(\mathbb{R})$ and thus

$$f(x) = \sum_{\ell=1}^{\infty} \langle u_\ell(\mathbf{x}), f(x) \rangle u_\ell(x). \tag{12}$$

Comparing coefficients of $u_\ell(x)$ in (11) and (12) gives $\sigma_\ell \langle v_\ell(y), \beta(y) \rangle = \langle u_\ell(x), f(x) \rangle$ which, with the requirement $\langle u_\ell(x), f(x) \rangle = 0$ when $\sigma_\ell = 0$, gives the solution to (6) as

$$\beta(y) = \sum_{\ell=1}^{\infty} \langle v_\ell(y), \beta(y) \rangle v_\ell(y) = \sum_{\ell=1}^{\infty} \frac{\langle u_\ell(x), f(x) \rangle}{\sigma_\ell} v_\ell(y). \tag{13}$$

It is immediate that a necessary condition for the convergence of (13) is the faster decay of the term $|\langle u_\ell(x), f(x) \rangle|$ in the numerator than the term σ_ℓ in the denominator. This leads to the well-known *Picard condition* (PC):

Condition 1 (Picard Condition [19, III.§11][7, p.12],[8, p.279])**.** *An integral equation of the first kind,* (6)*, has an unique square integrable solution* $\beta(\mathbf{y})$ *if and only if the series* (13) *satisfies*

$$\|\beta(y)\|^2 = \sum_{\ell=1}^{\infty} \frac{\langle u_\ell(x), f(x)\rangle^2}{\sigma_\ell^2} < \infty. \tag{14}$$

When the kernel is symmetric, $u_\ell = v_\ell$, and $\langle u_\ell(x), f(x)\rangle$ can be viewed as the generalized Fourier expansion coefficients of f with respect to the orthogonal system $\{u_\ell\}$ [19, p.160]. Moreover, the absolute and uniform convergence of (10) can be guaranteed by the well-known Mercer theorem [21]. From a historical perspective the SVE is a generalization of the eigenfunction/Fourier series expansion of a symmetric kernel [19, 20]. The singular value system $\{\sigma_\ell, u_\ell, v_\ell\}$ is the analogue of the Fourier series expansion in the sense that low frequency singular functions correspond to the large singular values and higher frequency singular functions correspond to smaller singular values [7, p.17]. For a given f, the faster decay of $|\langle u_\ell(x), f(x)\rangle|$, on average for some $\ell > 0$, as compared to σ_ℓ is indicative that the high frequency content of $f(x)$ is less than that of $K(x, y)$. Equivalently, the magnitudes of $|\langle u_\ell(x), f(x)\rangle|$ convey information on the *smoothness* of f with respect to the basis $\{u_\ell(x)\}$, while the decay of the singular values is closely related to the *smoothness* of the kernel function $K(x, y)$. In the equivalence of the expansions for $f(x)$ in (11) and (12) it is the mapping of u_ℓ to $\sigma_\ell v_\ell$ which helps to explain why higher frequencies (larger ℓ) in β are damped more in f than the lower frequencies, as observed from [7, p.11].

Remark 1. *For a given* $K(x, y)$ *the solution* $\beta(y)$ *is only square integrable when* $f(x)$ *is sufficiently smooth with respect to the singular system for* $K(x, y)$*. In contrast if* $f(x)$ *is given but* $K(x, y)$ *can be chosen, then an appropriate kernel is one with a singular system such that coefficients of* $f(x)$ *in the singular basis* $\{u_\ell\}$ *decay faster, on average for a given* $\ell > 0$*, than the singular values* σ_ℓ*.*

This interpretation is not standard when treating the integral equation case, in which $K(x, y)$ and $f(x)$ are prescribed by an underlying model and associated measurements, respectively. For the RBF problem, measurements $f(x)$ are provided, but the choice of kernel $K(x, y)$ depends on a number of factors including the context in which the approximation (1) is to be used. Specifically, the choice for $K(x, y)$ may be impacted by the underlying smoothness of $f(x)$ with respect to the basis for $K(x, y)$.

2.2. The Discrete Picard Condition

The SVD, [22], is employed to analyse the discrete problem (2). For a general matrix $\Phi \in \mathbb{R}^{M\times N}$ the SVD is given by $\Phi = U\Sigma V^T$, [22], where, $U \in \mathbb{R}^{M\times M}$, $V \in \mathbb{R}^{N\times N}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{M\times N}$ is the diagonal matrix of singular values, ordered from largest to smallest along the principal diagonal, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$. Given the SVD, and explicitly assuming that Φ has full column rank, $\sigma_N > 0$, the solution of (2) is represented by the dyadic sum

$$\boldsymbol{\alpha} = \sum_{\ell=1}^{N} \frac{\mathbf{u}_\ell^T \mathbf{f}}{\sigma_\ell} \mathbf{v}_\ell = V\Sigma^{-1}U^T\mathbf{f}, \tag{15}$$

where $\mathbf{u}_\ell$ and $\mathbf{v}_\ell$ are the ℓ^{th} columns of matrices U and V, respectively. This relation is the immediate analogue of the continuous expression (13) using the SVD as the discrete analogue of the continuous SVE. The solution is a weighted linear combination of the basis vectors $\mathbf{v}_\ell$ with weights $\mathbf{u}_\ell^T\mathbf{f}/\sigma_\ell$. Moreover

$$\|\boldsymbol{\alpha}\|_2^2 = \sum_{\ell=1}^{N} \frac{(\mathbf{u}_\ell^T\mathbf{f})^2}{\sigma_\ell^2}, \tag{16}$$

which is the equivalent of (14) for the square integrability of the discrete solution $\boldsymbol{\alpha}$ and leads immediately to the extension of the Picard condition for the discrete case.

Condition 2 (Discrete Picard Condition [23, p.9] [7, p.37])**.** *Suppose that there exist r and τ such that numerically $\sigma_\ell = O(\tau), \forall\, \ell > r$. Then the discrete Picard condition (DPC) is said to be satisfied if, for all singular values which are larger than τ, the corresponding coefficients $|\mathbf{u}_\ell^T\mathbf{f}|$, on average, decay faster than σ_ℓ.*

Observe first that while it cannot be assumed that satisfaction of the DPC implies satisfaction of the PC for the underlying continuous problem [23], it does have an equivalent interpretation. In particular, the DPC depends not only on the smoothness of the data $\mathbf{f}$ but also on the discrete singular system for the underlying kernel Φ which depends on the choice of RBF, ϕ. It is well-recognized that the solvability of a given interpolation problem depends on the condition of Φ, while the DPC makes it clear that the quality of the solution is also directly related to the frequency content of the data $\mathbf{f}$, measured with respect to the singular basis for Φ, even when no noise is present in $\mathbf{f}$. The decay rate of $|\mathbf{u}_\ell^T\mathbf{f}|$ should be no slower than the decay rate of σ_ℓ. In practice,

the DPC is examined through the Picard plot, which is the plot of $|\mathbf{u}_\ell^T\mathbf{f}|$, σ_ℓ and the ratios $|\mathbf{u}_\ell^T\mathbf{f}|/\sigma_\ell$ against index ℓ. If the DPC is not satisfied, the two norm of the solution $\boldsymbol{\alpha}$ will be large. Moreover, it is immediate from (16) that

$$\|\boldsymbol{\alpha}\|_2^2 = \sum_{\ell=1}^{N} \frac{|\mathbf{u}_\ell^T\mathbf{f}|^2}{\sigma_\ell^2} \leq \left(\frac{1}{\sigma_N^2}\right) \sum_{\ell=1}^{N} |\mathbf{u}_\ell^T\mathbf{f}|^2 = \|\Phi^{-1}\|_2^2\|\mathbf{f}\|_2^2\|. \tag{17}$$

In particular, $\|\boldsymbol{\alpha}\|_2$ large relative to $\|f(\mathbf{x})\|_2$ is indicative of ill-conditioning of Φ and sensitivity of the solution $\boldsymbol{\alpha}$ to errors in both $f(\mathbf{x})$ and to numerical errors accrued in evaluating $\boldsymbol{\alpha}$.

2.3. Error in the approximation due to ill-conditioning of Φ

As already detailed in Section 2.2, the boundedness of the solution $\boldsymbol{\alpha}$ is directly related to the spectrum of the operator Φ. Error in $\boldsymbol{\alpha}$, which depends on both the ill-conditioning of Φ and the accuracy of the measurements f, also impacts the quality of the approximation $s(\mathbf{y})$. Suppose $\tilde{f}_k$ are the actual measurements approximating the exact data f_k, such that $f_k = \tilde{f}_k + \eta_k$ for some noise vector $\boldsymbol{\eta}$ with components η_k and that $\tilde{s}(\mathbf{y})$ is the calculated approximation given $\tilde{\mathbf{f}}$. Practically, one may assume that calculating the solution of (2) in finite precision arithmetic yields $\tilde{\boldsymbol{\alpha}}$ which exactly solves the perturbed system

$$\tilde{\Phi}\tilde{\boldsymbol{\alpha}} = \tilde{\mathbf{f}}, \tag{18}$$

where $\tilde{\Phi}$ is a matrix near to Φ in the sense that $\tilde{\Phi} = U\tilde{\Sigma}V^T$, for U and V the orthogonal matrices from the SVD of Φ, and $\tilde{\Sigma}$ the matrix of perturbed singular values $\tilde{\sigma}_\ell$, with the standard ordering. Consequently, the obtained $\tilde{s}$ depends on $\tilde{\boldsymbol{\alpha}}$. The quality of the approximation may be measured through $E = \|\mathbf{f} - s(\mathbf{y})\|_2$ which is bounded as follows:

$$E = \|\mathbf{f} - s(\mathbf{y})\|_2 \leq \underbrace{\|\mathbf{f} - \tilde{\mathbf{f}}\|_2}_{E_0} + \underbrace{\|\tilde{\mathbf{f}} - \tilde{s}(\mathbf{y})\|_2}_{E_1} + \underbrace{\|\tilde{s}(\mathbf{y}) - s(\mathbf{y})\|_2}_{E_2}. \tag{19}$$

Here $E_0 = \|\boldsymbol{\eta}\|^2$ is the measurement error, E_1 is the conventional approximation error under the assumption that $\tilde{s}$ is obtained to approximate the measured data $\tilde{\mathbf{f}}$, while E_2 is the error in the approximation. Observe first that the error in approximating f at vector $\mathbf{y}$ is given by

$$s(\mathbf{y}) - \tilde{s}(\mathbf{y}) = \sum_{j=1}^{N} (\alpha_j - \tilde{\alpha}_j)\Phi(\mathbf{y}, \mathbf{x}_j) = \Phi(\mathbf{y}, \mathbf{x})(\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}) \quad \text{which yields} \tag{20}$$

$$\|s(\mathbf{y}) - \tilde{s}(\mathbf{y})\|_2 = \|\Phi(\mathbf{y}, \mathbf{x})(\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}})\|_2 \leq \|\Phi(\mathbf{y}, \mathbf{x})\|_2\|\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\|_2. \tag{21}$$

Now

$$\|\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}\|_2^2 = \|(\Sigma^{-1}\tilde{\Sigma} - I)\tilde{\boldsymbol{\alpha}} + \Sigma^{-1}U^T\boldsymbol{\eta}\|_2^2 \leq \|(\Sigma^{-1}\tilde{\Sigma} - I)\|_2^2\|\tilde{\boldsymbol{\alpha}}\|_2^2 + \|\Phi^{-1}(\mathbf{y},\mathbf{x})\|_2^2\|\boldsymbol{\eta}\|_2^2,$$

which then yields

$$\begin{aligned}\|s(\mathbf{y}) - \tilde{s}(\mathbf{y})\|_2^2 &\leq \mathrm{cond}^2(\Phi(\mathbf{y},\mathbf{x}))\|\boldsymbol{\eta}\|_2^2 + \|\Phi(\mathbf{y},\mathbf{x})\|_2^2\|(\Sigma^{-1}\tilde{\Sigma} - I)\|_2^2\|\tilde{\boldsymbol{\alpha}}\|_2^2 \\ &\leq \mathrm{cond}^2(\Phi(\mathbf{y},\mathbf{x}))\left(\|\boldsymbol{\eta}\|_2^2 + \sigma_1^2(\Delta)\|\tilde{\boldsymbol{\alpha}}\|_2^2\right), \quad \Delta = \Sigma - \tilde{\Sigma}, \end{aligned} \tag{22}$$

and where the last inequality follows by, for example, [22, Corollary 8.6.2]. The matrix Δ measures the error in the approximation of $\Phi(\mathbf{y},\mathbf{x})$ by the calculated $\tilde{\Phi}(\mathbf{y},\mathbf{x})$, which is, in general, unknown, but can certainly be assumed to be bounded below by the machine epsilon. Of course the bound in (22) can be replaced by a bound in terms of $\|\tilde{\mathbf{f}}\|_2^2$, but this is less useful and requires further information on the maximum ratio in the singular values $\{\sigma_\ell/\tilde{\sigma}_\ell\}$. On the other hand, from (22) one sees immediately the contamination of the approximation $s(\mathbf{y})$ due to the ill-conditioning of $\Phi(\mathbf{y},\mathbf{x})$ and the relation to the size of the calculated coefficients $\tilde{\boldsymbol{\alpha}}$. Moreover, ignoring the contribution due to the measurement error, the error in the approximation is bounded by a scaling of the size of the calculated weight vector, where the scaling is controlled if $\sigma_1(\Delta)$ is such that the product $\sigma_1(\Delta)\mathrm{cond}(\Phi(\mathbf{y},\mathbf{x}))$ is not too large. To reiterate:

Remark 2. *The size of the calculated* $\|\boldsymbol{\alpha}\|_2$ *relative to* $\|\mathbf{f}\|_2$ *reflects the smoothing effects of the underlying kernel matrix and can itself be viewed as an indication of the ill-conditioning of the underlying problem.*

3. Spectral Distribution of Kernel Matrices

From the results presented in Section 2 it is clear that an exponentially decaying spectrum for the interpolation matrix Φ negatively impacts the accuracy of the estimation of $\boldsymbol{\alpha}$ and, through this, also the quality of the resulting approximation $s(\mathbf{y})$. Indeed the stability of the solution of the RBF interpolation problem is closely related to the spectral distribution of the underlying kernel $K(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x},\mathbf{y})$ which is explicitly assumed to be symmetric for the the interpolation problem described by (2). In this case, with the standard ordering for the singular values $\sigma_\ell =$

$|\lambda_\ell|$ where λ_ℓ is an eigenvalue of Φ, it is therefore sufficient to analyse the eigenvalues of Φ. Moreover, when the continuous kernel $K(\mathbf{x}, \mathbf{y})$ is square integrable, the discrete singular values from the SVD for the finite operator are increasingly better approximations, with increasing N, to the continuous singular values of the SVE when calculated using a standard quadrature, [24]. Given these connections we therefore study the continuous eigenvalue problem

$$\int_\Omega K(\mathbf{x}, \mathbf{y})\alpha(\mathbf{y})d\mathbf{y} = \lambda\alpha(\mathbf{x}), \tag{23}$$

which has been studied by many authors including in Bôcher [25], Weyl [26], Hille and Tamarkin [27], Little and Reade [28], as well as in the texts of Smithies [20] and Cochran [29], in which also important early results of Fredholm and Hilbert are detailed.

Consistent with the standard ordering of the singular values it is assumed throughout that the eigenvalues are ordered by decreasing magnitude with increasing ℓ. Moreover, a kernel is symmetric if $K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{x})$, while it is translation invariant if $K(\mathbf{x}, \mathbf{y}) = k(\mathbf{x} - \mathbf{y})$ for some function $k(\mathbf{x})$. If the underlying function $k(\mathbf{x})$ is also periodic, then the kernel is said to be translation invariant and periodic. These definitions are relevant in the context of the translation invariant symmetric RBF $\phi_\epsilon(\|\mathbf{x} - \mathbf{y}\|_2)$. Generally, it is assumed here that the kernel is real and square integrable, and thus, asymptotically, the results may be viewed as providing results which are relevant in the discrete case.

3.1. Scalar kernels in $\mathbb{R}$

Theorem 1 (Weyl [26, p.449][16, p.123] [20])**.** *Suppose $K(x, y)$ is a symmetric kernel and its partial derivatives $\partial^q K/\partial^q x$, $q = 0, \ldots, \nu$ exist and are continuous, then $|\lambda_\ell| = O(\ell^{-\nu-1/2})$.*

While it can be shown that Weyl's result is sharp for the general case [30], there are cases in which sharper results can be obtained.

Theorem 2 ([31])**.** *Suppose $K(x, y)$ is a symmetric positive definite kernel which is periodic in both variables with the same period, and its partial derivatives $\partial^q K/\partial^q x$, $q = 0, \ldots, \nu$ exist and are continuous, then $\lambda_\ell = o(\ell^{-\nu-1})$.*

Theorem 3. *[32, Theorem 4] Suppose $K(x, y) = k(x - y)$ is a translation invariant kernel on* $[-1, 1]$.

1. *If $k(x) \in C^{\nu-1}$ and its ν^{th} derivative is of bounded variation, then $|\lambda_\ell| = O(\ell^{-\nu})$ as $\ell \to \infty$.*
2. *If $k(x) \in C^\infty$, then $|\lambda_\ell|$ decay faster than any $O(\ell^{-\nu})$ as $\ell \to \infty$ for every $\nu > 0$.*

When the kernel $K(x, y)$ is both analytic and symmetric, its eigenvalues decay much faster.

Theorem 4 (Little-Reade [28])**.** *Suppose $\mathcal{E}_R$ is an ellipse with foci at ± 1 and semi-axis sum $R > 1$. 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 $\mathcal{E}_R$ which is uniformly bounded in z, y in this range, then $|\lambda_\ell| = O(R^{-\ell})$.*

Remark 3. *For a translation invariant kernel $k(x)$, the smoother $k(x)$ the more quickly its eigenvalues decay to zero. This is a negative result with respect to finding stable solutions of the underlying integral equation* (6)*.*

The shape parameter, $\epsilon > 0$, of the kernel is also relevant. The explicit dependence of the eigenvalues on the shape parameter is noted by introducing the notation $\lambda_\ell(\epsilon)$.

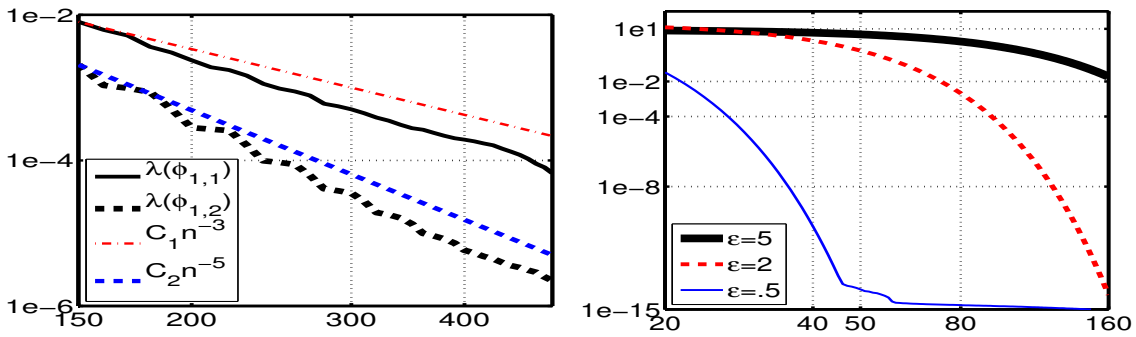
Theorem 5 ([32])**.** *Suppose that $K(x, y) = k(x - y)$ is a translation invariant kernel.*

1. *If $k(x) \in C^{\nu-1}$ and its ν^{th} derivative is of bounded variation, then $|\lambda_\ell(\epsilon)| = O((\epsilon/\ell)^\nu)$ as $\ell \to \infty$.*
2. *If $k(x) \in C^\infty$, then $|\lambda_\ell(\epsilon)|$ decay faster than $O((\epsilon/\ell)^\nu)$ as $\ell \to \infty$ for every $\nu > 0$.*

It is immediate from Theorem 5 that the eigenvalues decay faster for smaller values of ϵ and thus the problem becomes more ill-posed when the underlying RBF is referred to as *flat*, [33, 34].

Remark 4. *For RBF ϕ_ϵ, the flatter the basis function, the more significant the smoothing effect, in the limit of large N, of the associated kernel matrix $\Phi(\epsilon)$.*

Note that a function containing only low frequency content, such as $\sin(x)$ is said to be *smoother* than a function with high frequency content such as $\sin(1000x)$, even though both functions have the same number of continuous derivatives. Figure 1 illustrates this result. The *flatter* the basis function, the faster the spectrum decays to zero.



(i) Eigenvalues of Φ with different smoothness.

(ii) Eigenvalues of $\Phi(\epsilon)$ with different ϵ.

Figure 1: (i) illustrates the eigenvalue distribution of kernel matrices for Wendland functions $\phi_{1,1} = (1-x)_+^3(3x+1) \in C^2$ and $\phi_{1,2} = (1-x)_+^5(8x^2+5x+1) \in C^4$, where $a_+ = \max\{a, 0\}$. The straight lines are the predicted trends for the eigenvalue decay according to Theorem 2, where $C_1 = \lambda_{150}(\phi_{1,1})150^3$, $C_2 = \lambda_{150}(\phi_{1,2})150^5$. (ii) shows the spectral distribution of kernel matrices for Gaussian radial basis functions with different scales, $\phi(x, \varepsilon) = e^{-(\varepsilon x)^2}$, with scale parameters $\varepsilon = 5, 2, 0.5$. The matrices Φ are obtained using an equal sampling on the interval $[-10, 10]$ using 501 points. In (ii) the first 160 eigenvalues are shown.

3.2. *Separable kernels in* $\mathbb{R}^d$

For the two dimensional case $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, separability means that $K(\mathbf{x}, \mathbf{y}) = K_1(x_1, y_1)K_2(x_2, y_2)$. While this does limit the kernels that can be considered, an example is given by the Gaussian kernels: $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2) = \exp(-(x_1 - y_1)^2)\exp(-(x_2 - y_2)^2)$. 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 $K(\mathbf{x}, \mathbf{y})$ [20]. Suppose that each K_j satisfies the conditions of Theorem 4 and that the eigenvalues of $K_j(x_j, y_j)$ are ordered in magnitude from largest to smallest $O(1), O(R^{-1}), O(R^{-2}), \ldots$ relative to an underlying ellipse with semi axes summing to R. It is not hard to determine that the related separable kernel $K(\mathbf{x}, \mathbf{y})$ therefore has the following ordering of eigenvalues: 1 of order $O(1)$, 2 of order $O(R^{-1})$, 3 of order $O(R^{-2})$ and $\binom{\ell}{1}$ of order $O(R^{-\ell})$. See for example the ordering of eigenvalues for a two dimensional tensor product kernel in Figure 2. The result extends in a similar manner for the tensor product kernel of dimension d.

Theorem 6 ([32])**.** *Suppose that $K(\mathbf{x}, \mathbf{y})$ is a separable, tensor product-like analytic kernel function in $\mathbb{R}^d$, and that the ℓ^{th} eigenvalue of each kernel K_j is of order $O(R^{-\ell-1})$, then there are exactly $\binom{\ell+d-1}{d-1} = \frac{(\ell+d-1)!}{\ell!(d-1)!}$, $\ell = 1, \ldots$, eigenvalues of the same order, namely $O(R^{-\ell-1})$ for some $R > 1$.*

$$
\begin{array}{cccccccc}
 & 1 & R^{-1} & R^{-2} & R^{-3} & R^{-4} & R^{-5} & \dots \\
1 & 1 & R^{-1} & R^{-2} & R^{-3} & R^{-4} & R^{-5} & \dots \\
R^{-1} & R^{-1} & R^{-2} & R^{-3} & R^{-4} & R^{-5} & R^{-6} & \iddots \\
R^{-2} & R^{-2} & R^{-3} & R^{-4} & R^{-5} & R^{-6} & \iddots & \iddots \\
R^{-3} & R^{-3} & R^{-4} & R^{-5} & R^{-6} & \iddots & \iddots & \iddots \\
R^{-4} & R^{-4} & R^{-5} & R^{-6} & \iddots & \iddots & \iddots & \iddots \\
R^{-5} & R^{-5} & R^{-6} & \iddots & \iddots & \iddots & \iddots & \iddots \\
\vdots & \iddots & \iddots & \iddots & \iddots & \iddots & \iddots & \iddots
\end{array}
$$

Figure 2: Ordered eigenvalues for a tensor product kernel in $\mathbb{R}^2$.

The following result is immediate using the separability of the Gaussian kernel in $\mathbb{R}^d$ and is closely related to the approach for choosing the diagonal matrix for the QR-RBF algorithm [2].

Corollary 1. *For the Gaussian radial basis function in $\mathbb{R}^d$, there are exactly $\binom{\ell+d-1}{d-1} = \frac{(\ell+d-1)!}{\ell!(d-1)!}$, $\ell = 1, \ldots,$ eigenvalues of the same order, namely $O(R^{-\ell-1})$ for some $R > 1$.*

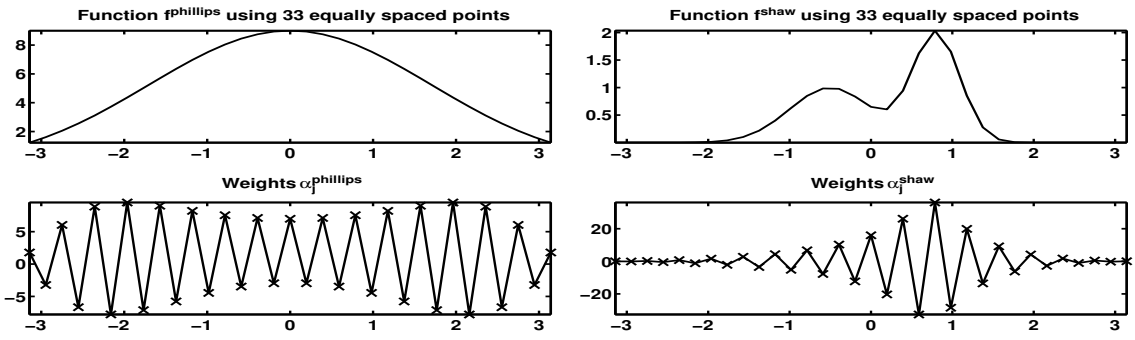
Remark 5. *Given a particular basis function $\phi(x)$ which defines a separable kernel in $\mathbb{R}^d$, the smoothing effect of the kernel matrix increases with dimension d.*

Being mindful of the limitations with regard to immediately extending results from the continuous to the discrete case, observe that it is precisely this smoothness of the basis function $\phi(x)$ which generates the difficulties with solving the discrete problems (3) and (6).

4. Numerical Verification

4.1. The Discrete Picard Condition

The discussion of Section 2 is illustrated in the context of the RBF interpolation problem (2). Figure 3 shows the observational data pairs $(\mathbf{x}_k, f(\mathbf{x}_k))$ and the recovered interpolation weights $\{\alpha_j\}$



(i) Problem `phillips` (ii) Problem `shaw`

Figure 3: Solving (2) with $N = 33$ for functions f^{phillips} and f^{shaw} without addition of any noise to obtain the weights $\{\alpha_j^{\text{phillips}}\}$ and $\{\alpha_j^{\text{shaw}}\}$ from samples at 33 equally spaced points on $[-\pi, \pi]$.

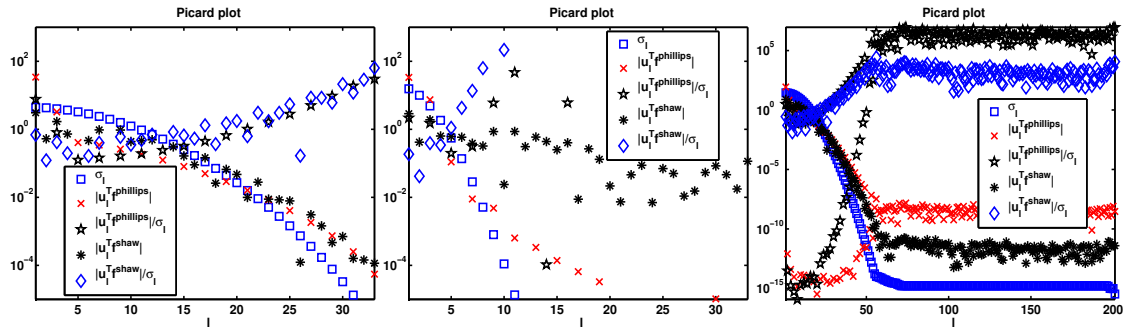
for two one dimensional examples based on the following two functions

$$f^{\text{phillips}}(x) = (6 - |x|)\,(1 + .5\cos(\pi x/3)) + \frac{4.5}{\pi}\sin(\pi|x|/3) \quad \text{and} \tag{24}$$

$$f^{\text{shaw}}(x) = 2\exp(-6(x - .8)^2) + \exp(-2(x + .5)^2). \tag{25}$$

These are the right hand sides for the `phillips` and `shaw` examples used in the Regularization toolbox of Hansen [35]. The RBF interpolation equation (2) is solved for the RBF $\phi_\epsilon(x) = \exp(-(\epsilon x)^2)$, $\epsilon = 2$, using $N = 33$, with equally spaced $\mathbf{x}_k = -\pi + (k-1)\Delta\mathbf{x}$, $k = 1, \ldots, N$, $\Delta\mathbf{x} = 2\pi/(N-1)$. The problem is moderately ill-conditioned, cond(Φ)= 2.45e6. The presented solutions $\{\alpha_j^{\text{phillips}}\}$ and $\{\alpha_j^{\text{shaw}}\}$ found by the naïve inversion of (2), using the dyadic sum (15), clearly illustrate the higher frequency content of the interpolating functions than that of the target functions, i.e. functions f^{phillips}, f^{shaw} are much smoother than the solutions $\alpha^{\text{phillips}}(\mathbf{y})$, $\alpha^{\text{shaw}}(\mathbf{y})$. On the other hand, given that the underlying system matrix $\Phi(\epsilon)$ is not well-conditioned, one cannot immediately assume that the given results are actually reasonable approximations to the true infinite precision solutions for $\boldsymbol{\alpha}$. Thus the conclusion concerning the frequency content may be an artifact of the sensitivity of the solutions due to the spectral system for $\Phi(\epsilon)$ and, as noted in Section 2.2, it is important to consider the DPC for the system.

Figure 4 shows the Picard plot, [23], as discussed in Section 2.2, the plot of $|\mathbf{u}_\ell^T\mathbf{f}|$, σ_ℓ and the ratios $|\mathbf{u}_\ell^T\mathbf{f}|/\sigma_\ell$ against index ℓ, for the problems illustrated in Figure 3. As ℓ increases the singular values σ_ℓ decrease to zero more quickly than the numerators $|\mathbf{u}_\ell^T\mathbf{f}|$. Although the matrix Φ has full



(i) $\epsilon = 2, N = 33$ (ii) $\epsilon = .5, N = 33$ (iii) $\epsilon = 2, N = 201$

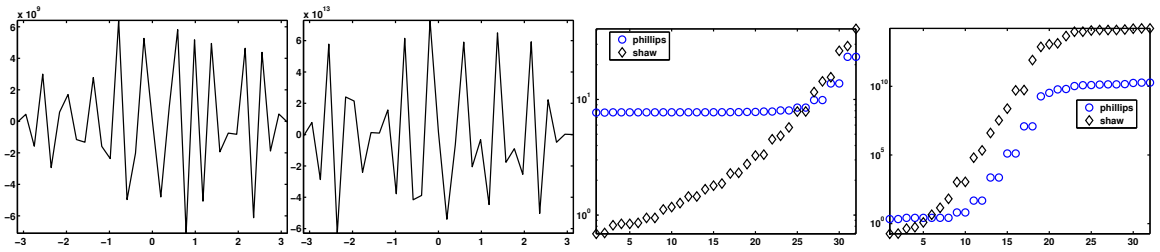
Figure 4: The Picard plot for problems f^{phillips} and f^{shaw} illustrated in Figure 3, (i) for $\epsilon = 2$, in (ii) for $\epsilon = .5$ and in (iii) the same problems but using $N = 201, \epsilon = 2$. For (ii) cond(Φ)= $7.3e17$ and for (iii) it is $6.0e18$. Note that in order to emphasize the growth in the coefficients the plots in (i)-(ii) are given for a limited range on the $y-$axis.

column rank the singular values $\sigma_\ell \to 0$ in such a way that $\Phi(\epsilon)$ is ill-conditioned, particularly when the size of the problem is increased as seen in Figure 4(iii), or when ϕ_ϵ becomes *flatter* as in Figure 4(ii). The observations are consistent with the discussion in Section 3. The impact of the shape parameter is clearly seen in the solutions illustrated in Figure 5 which have $\epsilon = .5$; notice in particular the scale of the $y-$axis in each case, the amplification in $\boldsymbol{\alpha}$ for shape parameter $\epsilon = .5$ is significant.

To highlight the problem with these naïve solutions obtained from direct inversion of (2) it is relevant to calculate *approximate* solutions using the truncated SVD

$$\boldsymbol{\alpha}_\ell = \sum_{\ell=1}^{N_\ell} \frac{\mathbf{u}_\ell^T \mathbf{f}}{\sigma_\ell} \mathbf{v}_\ell, \tag{26}$$

for truncation parameters $1 \le N_\ell \le N$. The dependence of $\|\boldsymbol{\alpha}_\ell\|^2$ on ℓ for the examples used in Figures 3 and 5(i)-5(ii), respectively, namely for the two different shape parameters, is illustrated in Figure 5. It is clear that as N_ℓ increases the solutions violate the DPC. Indeed the parameter N_ℓ is commonly regarded as a *regularization* parameter; in calculating the solution $\boldsymbol{\alpha}$ it is necessary to either use truncation to obtain a solution without blow-up or to introduce some alternative regularization. The choice of regularization parameter is extensively studied in the inverse problems literature, in which the dependence of the solution on the conditioning of the problem and the noise in the data is well-described, [23]. The focus here is on results using truncation, for which



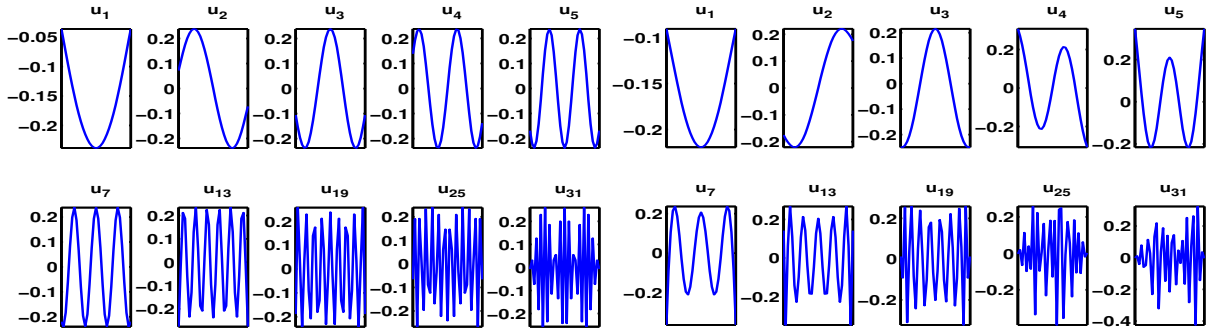
(i) Problem `phillips` (ii) Problem `shaw` (iii) $\epsilon = 2$ (iv) $\epsilon = .5$

Figure 5: (i)-(ii) show the solution of (2) with $N = 33$ for functions f^{phillips} and f^{shaw}. The weights $\{\alpha_j^{\text{phillips}}\}$ and $\{\alpha_j^{\text{shaw}}\}$ are obtained from samples at 33 equally spaced points on $[-\pi, \pi]$ using a Gaussian kernel with $\epsilon = .5$. The solutions are significantly amplified as compared to the solutions with $\epsilon = 2$ shown in Figure 3. In (iii)-(iv) the discrete Picard condition, Condition 2, for functions f^{phillips} and f^{shaw} for the truncated solutions (26) $N_\ell \leq N$. $\|\boldsymbol{\alpha}_\ell\|^2$ is plotted against index N_ℓ. The data is sampled at 33 equally spaced points on $[-\pi, \pi]$ using a Gaussian kernel with the noted shape parameter ϵ.

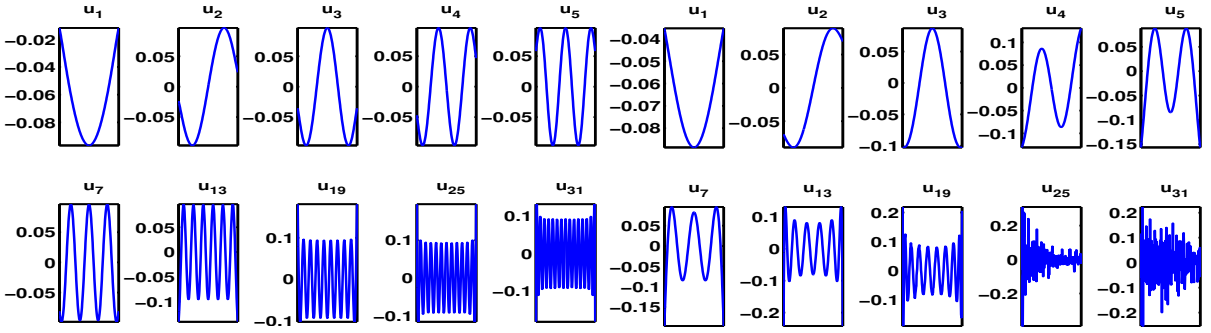
Figure 4(i) suggests that when $\epsilon = 2$ one should expect to truncate at about $N_\ell = 16$ in order to avoid the coefficients which start to increase with index ℓ, although strictly speaking for this choice the singular values have not reached the machine tolerance, τ, such that $\sigma_\ell = O(\tau)$, rather $\sigma_\ell > O(\tau)$ for $\ell > N_\ell$. In general, this choice depends not only on these singular values but also on the noise level in the measured data f.

4.2. Basis Vectors for the Solution

It is clear that the spectrum of the kernel matrix impacts the determination of feasible solutions of the RBF interpolation problem. The more significant the smoothing effect of the kernel matrix, the more difficult it is to find a satisfactory solution of (2). While this discussion has so far emphasized the effect of the poor conditioning of Φ due to the exponentially decaying spectrum, it should also be apparent that the existence of the small smoothing singular values will also lead to finite precision errors in the calculations of the basis vectors. High frequency basis vectors carry information corresponding to the extremely small smoothing eigen/singular values and are thus more likely to be contaminated by the effects of finite precision arithmetic. Incorrect estimates of these basis vectors may, therefore, contaminate the solutions and impact the quality of the RBF approximation.



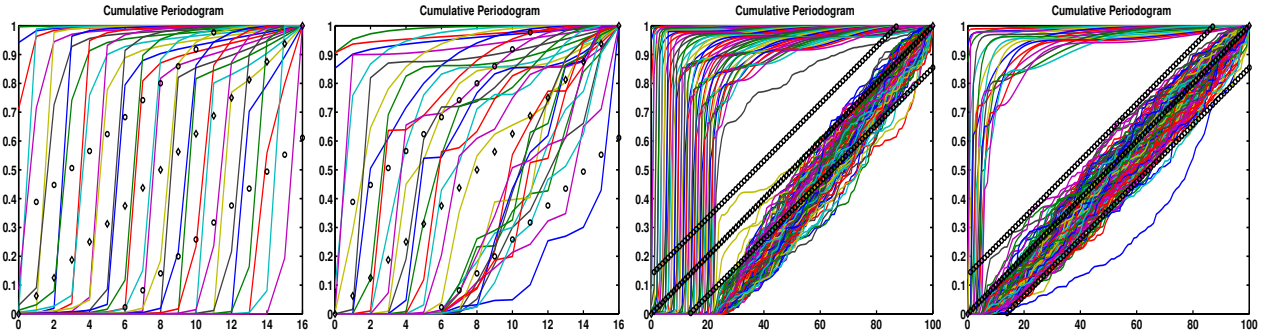
(i) $N = 33$, $\epsilon = 2$ (ii) $N = 33$, $\epsilon = .5$



(iii) $N = 201$, $\epsilon = 2$ (iv) $N = 201$, $\epsilon = .5$

Figure 6: The basis vectors $\mathbf{u}_\ell$ for the the same system matrices used to generate the Picard plots in Figure 4, for $N = 33$ and $N = 201$. Note that these vectors are for the Gaussian radial basis function with $\epsilon = 2$ and $\epsilon = .5$.

Figure 6 illustrates the basis vectors $\mathbf{u}_\ell$ for a selection of ℓ for both cases $N = 33$ and $N = 201$, for the examples in Figure 4. The $y-$axis is not shown because the vectors are normalized to norm 1 and thus the scales in each case are comparable, with definitely no blow up in any case. The dependence of the basis vectors on the shape parameter ϵ is shown in Figures 6(ii) and 6(iv). There is an increasing number of oscillations, corresponding to increasing frequency content of the vectors, with increasing ℓ. This confirms that the solutions for $\boldsymbol{\alpha}$ will contain broad frequency content according to (26) dependent on the choice for N_ℓ. Still it is not possible to examine the frequency content of the basis vectors and then immediately determine an appropriate choice for N_ℓ. Indeed there is apparent mixing of frequency content for larger ℓ, when considered for the broader kernel with $\epsilon = .5$. The frequency content of a given vector $\mathbf{u}_\ell$ can be examined using the normalized cumulative periodogram (NCP), which is a standard time series analysis technique, e.g. [5]. A brief description of the approach follows.



(i) $N = 33, \epsilon = 2$ (ii) $N = 33, \epsilon = .5$ (iii) $N = 201, \epsilon = 2$ (iv) $N = 201, \epsilon = .5$

Figure 7: The NCP for vectors $\mathbf{u}_\ell$ for $N = 33$ and $N = 201$, with $\epsilon = 2$ and $\epsilon = .5$, using equally spaced sampling. The black markers indicate the upper and lower KS-5% bounds, and the diagonal white noise line given by $y = \ell/\tilde{N}$.

The power spectrum of a given vector $\mathbf{b}$ of length N is determined via the discrete Fourier transform (dft) of the vector treated as a time series. For given $\mathbf{b}$ define $\mathbf{c}_\ell = |(\text{dft}(\mathbf{b}))_\ell|^2$, for $\ell = 1, \ldots, \tilde{N} = \lfloor N/2 \rfloor + 1$. Then the NCP is generated as the ratio of the cumulative sum of entries

$$\mathbf{z}_j = \frac{\sum_{\ell=1}^{j} \mathbf{c}_\ell}{\sum_{\ell=1}^{\tilde{N}} \mathbf{c}_\ell}, \quad j = 1, \ldots, \tilde{N}. \tag{27}$$

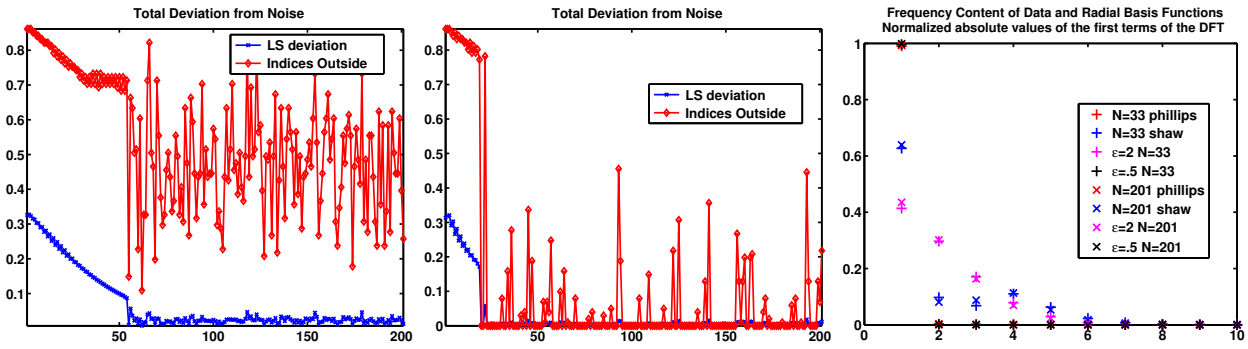
This definition includes the bias term (or DC term) $\mathbf{c}_0$. It is then well-known that for a white noise time series, the plot of $\mathbf{z}_j$ against frequency j normalized between 0 and 1 should lie within the Kolmogorov-Smirnoff (KS) limits of a straight line, which has slope 1 and length $\sqrt{5}/2$ [5, p. 363]. Moreover, vectors with lower frequency content lie primarily above the KS limits, while those with high frequency lie distinctly below the KS limits. In all cases the normalization reaches 1 for the last index; see Figures 7 for the basis vectors $\ell = 1 : N$ for the same cases as Figure 6, and observe not only the dependence on both N and ϵ, but also the finite precision impact of the calculations. In particular for large N the lack of accuracy in the calculation of the singular values, translates to inaccurate estimates of the underlying singular vectors, as exhibited by the large number of vectors within the white noise KS bounds for the case with $N = 201$, i.e. these vectors degenerate to becoming white noise for large ℓ [36]. On the other hand, the lower frequency content for the *smoother* kernel with $\epsilon = .5$ translates to a larger number of vectors with low frequency content; more vectors above the upper KS line.

It should be apparent from the NCP that it is important to include in any given basis for a

solution, e.g. in the truncated dyadic sum (26), only those basis vectors which represent *true* frequency content. Thus one needs a test to detect for the contamination of a vector by white noise, with respect to some tolerance, so as to estimate an appropriate choice for $N_\ell \leq N$. Observe that this determination is totally independent of the measured data $\mathbf{f}$. A number of suggestions for automatically determining whether a residual vector for the data fit term, $\mathbf{r} = \Phi\boldsymbol{\alpha} - \mathbf{f}$, represents white noise have been presented in the literature, [37–39], all of which in some manner measure the deviation of the NCP from the diagonal white noise line $y = \ell/\tilde{N}$. In this case it is not a residual which is of interest, rather given a basis of orthogonal vectors, here the set $\{\mathbf{u}_\ell\}$, what is the largest N_ℓ such that all vectors with $\ell > N_\ell$ are effectively, with respect to some tolerance, contaminated by white noise?

Two standard approaches for measuring this deviation from white noise are considered, [37–39]; the first is the total least squares deviation from the diagonal white noise line summed over all entries for each vector, and the second is the percentage count of vector entries outside the KS bounds for each vector. Given these measures it remains to determine an appropriate tolerance for accepting the hypothesis that the vector is effectively a white-noise vector, relative to the chosen KS confidence level. Here the approach of [36] is followed: for the Kolmogorov-Smirnov test with a 5% confidence interval set $N_\ell = r$ which is the least index r such that for vector $\mathbf{u}_{r+1}$ more than 50% of the NCP lies within the KS-5% bounds. The 50% choice is in some sense arbitrary, as is the choice of any regularization tolerance, but in this case one can see that there is a distinct cut-off for these basis vectors. This is illustrated in Figure 8(i)-8(ii) for the case with $N = 201$ with shape parameters $\epsilon = 2$ and $\epsilon = .5$, respectively. With respect to both measures – the red and blue curves – the sudden drop for a specific r is preserved. It is also clear from Figure 7 that the noise is much more significant for $N = 201$ than for $N = 33$ and the number of reliable vectors when $\epsilon = .5$ is much fewer than when $\epsilon = 2$.

Figure 8(iii) illustrates the underlying frequency content of the right hand sides for `phillips` and `shaw` with $N = 33$ and $N = 201$ equally spaced samples, and the Gaussian radial basis functions sampled on the same grids, with $\epsilon = 2$ and $\epsilon = .5$, with the total frequency normalized to 1. The dft of each vector, normalized to 1 is plotted, rather than the cumulative vector which is not useful due to the limited frequency content of these examples, Figure 8(iii). One can see that



(i) $N = 201$, $\epsilon = 2$ (ii) $N = 201$, $\epsilon = .5$ (iii)

Figure 8: (i)-(ii) the deviation of vectors $\mathbf{u}_\ell$ from white noise for $N = 201$, with $\epsilon = 2$ and $\epsilon = .5$, and equal sampling. In (iii) the dft for f^{phillips} and f^{shaw} with $N = 33$ and $N = 201$ and for the Gaussian radial basis functions with $\epsilon = 2$ and $\epsilon = .5$.

the analysis is virtually independent of the sampling. Moreover, `phillips` is far smoother than `shaw`, and the RBF with shape parameter $\epsilon = .5$ is smoother than that with $\epsilon = 2$.

In summary, the DPC provides significant information on the impact of both noise in the measurements f and the accuracy of the smaller spectral values for estimating the number of potentially reliable terms in the TSVD estimate of the solution. The NCP, on the other hand, is measurement independent and may more precisely show the first deterioration in the basis and so more effectively indicate a good truncation parameter for the TSVD solution.

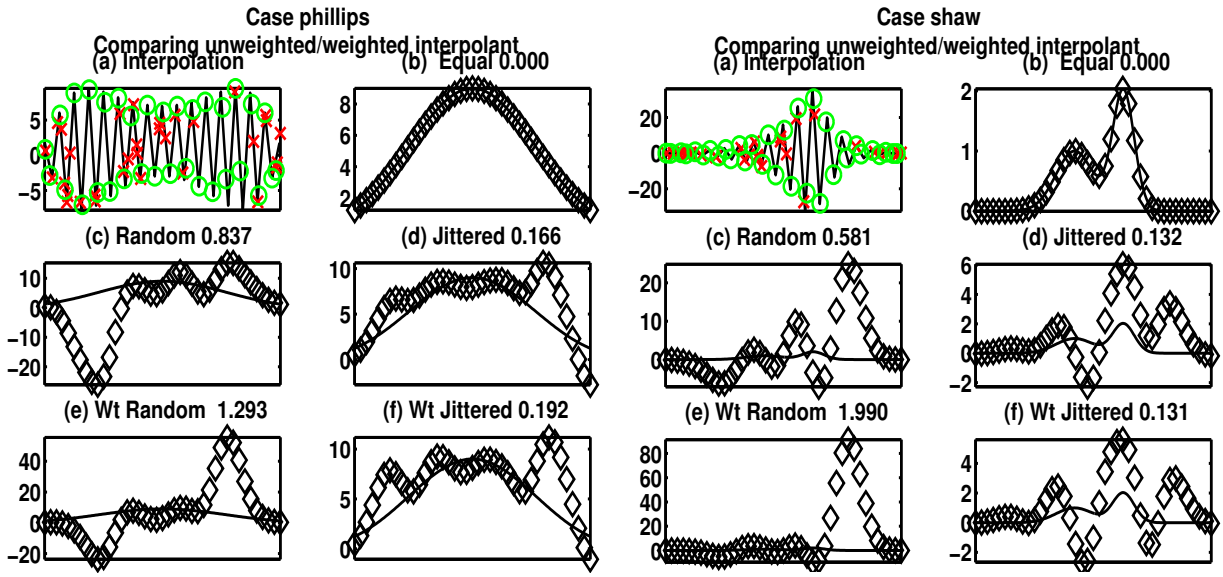
5. Applying Results to find interpolants

5.1. Impact on interpolation

The relationship between the integral equation (6) and the quadrature (7) leads to the following observation.

Remark 6. *The weights* $\{\alpha_j\}$ *in* (1) *may be regarded as the sampling of a function* $\alpha(\mathbf{y})$.

The impact of this remark is now discussed with respect to the use of both the naïve and the truncated solutions with $N_\ell = 16$ when $\epsilon = 2$ for interpolating the functions f^{phillips} and f^{shaw}. Results are shown first for the naïve solutions in Figure 9, and then for the truncated solutions in Figure 10. As before, note that the choice $N_\ell = 16$ is a regularization parameter choice, here based



(i) Problem `phillips` (ii) Problem `shaw`

Figure 9: Using the direct solution $\boldsymbol{\alpha}$ from (2) to obtain approximants to f^{phillips} and f^{shaw}, the left and right blocks (i)-(ii) for each problem, respectively. In panels (a) the solutions α are given as the solid black lines. This data is interpolated as described in the text to find estimates for $\boldsymbol{\alpha}(\mathbf{y})$ for random and jittered sampling in $\mathbf{y}$, giving weights α_{R} and α_{J}, red x and green o symbols respectively. The resulting weights are used to generate approximations to f^{phillips} and f^{shaw} at 201 equally spaced points, using the original α in panel (b), α_{R} in panel (c) and α_{J} in panel (d). In panels (c)-(d) no weighting to account for the quadrature is performed, whereas in panels (e)-(f) weighted quadrature based on (7), assuming trapezoidal approximation, is used. The titles show the respective root mean square errors calculated using the approximated and exact data. The shape parameter is $\epsilon = 2$ as in Figure 3.

on the results of the NCP analysis for the case with no noise in the measurements and chosen only as an illustration of the use of $\boldsymbol{\alpha}(\mathbf{y})$ as an interpolating function to the weights $\{\alpha_j\}$.

Based on Remark 6 one should be able to resample $\boldsymbol{\alpha}$ at a set $\{\mathbf{y}_j\}$ to obtain (1), provided that the resampled points $\{\boldsymbol{\alpha}(\mathbf{y}_j)\}$ are appropriately weighted in accordance with (7). Consider now resampling for two additional data sets, $\mathbf{y}_{\text{R}}$ and $\mathbf{y}_{\text{J}}$, each at 33 points on the interval $[-\pi, \pi]$, where $\mathbf{y}_{\text{R}}$ is randomly sampled on the interval, and $\mathbf{y}_{\text{J}}$ represents a random jitter of the points away from equally spaced points by $.2\zeta_j(2\pi)/33$ for ζ_j, $j = 1, \ldots, 33$ sampled from a normal distribution with mean 0 and variance 1. The interpolation from equally spaced points to the random or jittered points uses, here for purposes of illustration only, a cubic spline interpolant and

provides two new sets of weights $\{\alpha_j^{\text{random}}\}$ and $\{\alpha_j^{\text{jitter}}\}$ illustrated as the red x and green o symbols in panels (a) in Figures 9. Given a set of weights one can assess their accuracy in approximating the underlying target function f for a different set of sampled points by using (2). Taking 201 equally spaced points on the given interval and forming the appropriate matrices Φ of size 201×33 yields then approximations for f^{phillips}, f^{shaw} at these 201 equally spaced points. The *root mean square error* calculated using the exact values for f^{phillips} and f^{shaw} given by (24) and (25) can then be measured. The results using $\{\alpha_j^{\text{equispaced}}\}$, $\{\alpha_j^{\text{random}}\}$ and $\{\alpha_j^{\text{jitter}}\}$ are illustrated in Figure 9 panels (b)-(d), respectively, for each problem. In these plots the interpolants are shown as the black diamonds as compared to the underlying exact solutions given by the continuous black lines. The diamonds are plotted for indices $1 : 4 : 201$ so as to permit visualization of the differences between exact and interpolated solutions. The title in each case gives the root mean square of the approximation. While the interpolation works well for $\{\alpha_j^{\text{equispaced}}\}$, it is immediate that the errors are large for the other two samplings, apparently contradicting Remark 6. But these results are obtained without appropriately reweighting the sampled $\{\alpha_j\}$ to be consistent with the integral equation (7). Results using the appropriately weighted samples using weights w_j in (7) chosen for a trapezoidal approximation to the integrals are illustrated in panels (e)-(f), yielding no improvement and further suggesting a problem with Remark 6, unless the solutions $\{\alpha_j^{\text{equispaced}}\}$ are not good approximations to the underlying infinite precision solutions.

Figure 10 illustrates the same experiments as in Figure 9 but using the truncated solutions. First from the solutions in panels (a) of Figure 10 it is clear that $\alpha^{\text{phillips}}(\mathbf{y})$, $\alpha^{\text{shaw}}(\mathbf{y})$ are far smoother than those illustrated in Figure 3. Carrying out the resampling process for the same sets of jittered and random samplings, yields the results in Figure 10. Now the obtained results are much improved and support the hypothesis of Remark 6, while also indicating that the frequency content of the underlying functions $\boldsymbol{\alpha}(\mathbf{y})$ need not be as severe as originally suggested. These results emphasize that weights $\{\alpha_j\}$ may indeed be considered as sampling of a function $\alpha(\mathbf{y}) = w(\mathbf{y})\beta(\mathbf{y})$. Moreover, it is relevant to use techniques for the solution of the ill-posed integral equation (6), when finding *good* approximations to the interpolating function $\alpha(\mathbf{y})$.

To summarize: the presented results illustrate that the solution of the RBF approximation problem is closely connected to the solution of ill-posed integral equations and may therefore be

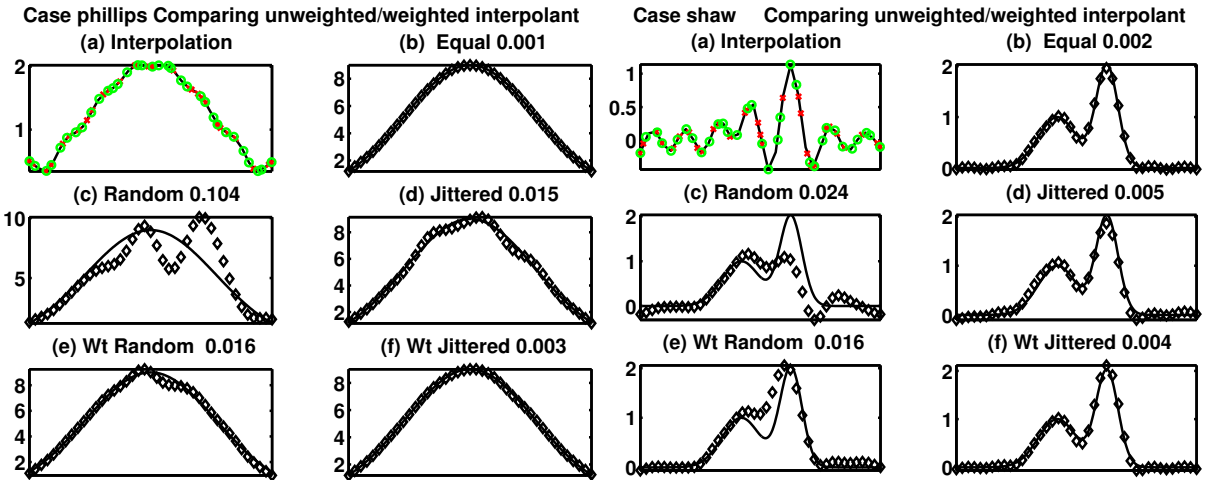


Figure 10: Using the truncated solution (26) with $N_\ell = 16$ to obtain approximants to f^{phillips} and f^{shaw}, the left and right blocks for each problem, respectively. Illustrated are then the equivalent results as shown in Figure 9 but obtained from the truncated solutions. The shape parameter is $\epsilon = 2$ as in Figure 3.

analysed using relevant techniques from the inverse problem literature.

6. Conclusions

Previous investigations of the properties of RBF interpolation/approximation seldom considered the close connection between the solution of the interpolation problem and the inverse problem for linear integral equations. Carefully examining this relationship shows that the system matrices of RBF interpolation may exhibit the same exponentially decaying spectrum that is commonly seen when solving ill-posed inverse problems. In this case the spectral properties depend directly on the shape parameters of the RBFs, which determine whether a given RBF is *flat*. The *flatter* (more smooth) the RBF ϕ_ϵ, the faster the exponential decay of the spectrum, due to the lack of high frequency content in ϕ_ϵ, and the consequently greater difficulty in solving the underlying system of equations. Many prior results have focused on the solvability and stability of the RBF system, stressing the importance of the condition number of the system, [10–13, 40–44]. This paper, in contrast, demonstrates that it is the exponential decay of the spectrum, related directly to the smoothing effects of the operator, which better explains the numerical difficulties of solving the underlying problems.

The Picard condition is used in standard investigations of ill-posed problems to determine a meaningful solution of a given set of equations, by revealing the number of reliable spectral coefficients in the SVD expansion of the kernel. For the RBF problem, the Picard condition reveals that the solution depends on both the smoothness of the underlying function to be approximated (the target function) and the *smoothness / flatness* of ϕ_ϵ. In contrast to the situation when seeking the solution of a given integral equation, in which the kernel function is prescribed by an underlying model of the problem, in the RBF case ϕ_ϵ can be chosen based on required criteria for the given problem. For example, one may deduce that for a function which is itself relatively *flat* that ϕ_ϵ should also be relatively *flat*. On the other hand, the Picard condition analysis demonstrates that it is necessary to balance the smoothness of the target function and the SVE of the kernel function, in order to find a solution which is meaningful.

In analyzing the stability of the RBF interpolation problem it is important to look at the complete spectral decomposition, namely the singular vectors which provide a basis for the solution, as well as the singular values. An approach for determining the number of reliable, *noise free* basis vectors has been presented, which goes beyond the standard Picard condition analysis, and is independent of the functional data. When the norm of the resulting weight vector is relatively large in contrast to the norm of the data, one may deduce that this is indicative of ill-conditioning of the problem, and that one many need to look for a more stable kernel, or to apply some kind of regularization. Here regularization based on using the truncated singular value expansion of the solution has been shown to provide a meaningful solution for the case of noise-free data. The truncation parameter can be estimated by examining the noise properties of the basis vectors.

The presented numerical results demonstrate that the obtained solutions are reliable in the context of quadrature applied to the integral equation; the weights for the RBF interpolant are samplings of a function $\boldsymbol{\alpha}(\mathbf{y})$ weighted by relevant quadrature weights dependent on the sampling. As such, the functional form of the weights is amenable to resampling so as to obtain weights for an alternative set of centers. Sampling from the functional form which is obtained from the truncated expansion yields reliable approximations, while resampling from the solution obtained without truncation is not meaningful. The idea of treating the weights as samplings from an underlying continuous function promises to provide an approach to adaptively move the centers of

the RBF kernels so as to either yield more efficient approximation evaluation, or, for example, to calculate derivative approximations at a different set of points which may be useful in improving the conditioning of the obtained operators for partial differential equation solvers.

In general there is significant scope to apply analysis techniques from the field of ill-posed inverse problems to the solution of the RBF approximation problem, with respect to areas such as investigation of the impact of error in measured data, determining feasible regularization techniques, and understanding the impact of iterative techniques in solving the resulting systems of equations. Likewise, developments from the RBF literature, most notably the QR-RBF algorithm [2], may provide useful insights for finding solutions of ill-posed integral equations. These are topics, amongst others, for future research relating to the use of the inverse theory for RBFs.

Acknowledgement

We'd like to thanks Prof P. C. Hansen for pointing out original reference of [16, 20] and Professor A. Wathen for comments on an early version of this manuscript.

Reference

[1] N. Flyer, B. Fornberg, Radial basis functions: Developments and applications to planetary scale flows, Computers and Fluids 46 (1) (2011) 23 – 32, ¡ce:title¿10th ICFD Conference Series on Numerical Methods for Fluid Dynamics (ICFD 2010)¡/ce:title¿. doi:10.1016/j.compfluid.2010.08.005.
URL http://www.sciencedirect.com/science/article/pii/S0045793010001982

[2] B. Fornberg, E. Larsson, N. Flyer, Stable computations with Gaussian radial basis functions, SIAM J. Sci. Comput. 33 (2) (2011) 869–892. doi:10.1137/09076756X.
URL http://dx.doi.org/10.1137/09076756X

[3] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting 20 (1) (2004) 5–10.

[4] P. R. Winters, Forecasting sales by exponentially weighted moving averages, Management Science 6 (3) (1960) pp. 324–342.
URL http://www.jstor.org/stable/2627346

[5] W. A. Fuller, Introduction to statistical time series / Wayne H. Fuller, Wiley, New York :, 1976.
URL http://www.loc.gov/catdir/enhancements/fy0607/76006954-t.html

[6] R. S. Varga, Matrix iterative analysis, expanded Edition, Vol. 27 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2000. doi:10.1007/978-3-642-05156-2.
URL http://dx.doi.org/10.1007/978-3-642-05156-2

[7] P. C. Hansen, Discrete inverse problems:Insight and algorithms, Vol. 7 of Fundamentals of Algorithms, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010.

[8] R. Kress, Linear integral equations, 2nd Edition, Vol. 82 of Applied Mathematical Sciences, Springer-Verlag, New York, 1999.

[9] C. A. Micchelli, Interpolation of scattered data: distance matrices and conditionally positive definite functions, Constr. Approx. 2 (1) (1986) 11–22. doi:10.1007/BF01893414.
URL http://dx.doi.org/10.1007/BF01893414

[10] F. Narcowich, J. Ward, Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices, J. Approx. Theory 69 (1) (1992) 84–109. doi:10.1016/0021-9045(92)90050-X.
URL http://dx.doi.org/10.1016/0021-9045(92)90050-X

[11] F. Narcowich, N. Sivakumar, J. Ward, On condition numbers associated with radial-function interpolation, J. Math. Anal. Appl. 186 (2) (1994) 457–485. doi:10.1006/jmaa.1994.1311.
URL http://dx.doi.org/10.1006/jmaa.1994.1311

[12] R. Schaback, Lower bounds for norms of inverses of interpolation matrices for radial basis functions, J. Approx. Theory 79 (2) (1994) 287–306. doi:10.1006/jath.1994.1130.
URL http://dx.doi.org/10.1006/jath.1994.1130

[13] R. Schaback, Error estimates and condition numbers for radial basis function interpolation, Adv. Comput. Math. 3 (3) (1995) 251–264. doi:10.1007/BF02432002.
URL http://dx.doi.org/10.1007/BF02432002

[14] P. C. Hansen, Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.

[15] H. Wendland, Scattered data approximation, Vol. 17 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.

[16] F. R. de Hoog, Review of Fredholm equations of the first kind, in: Application and numerical solution of integral equations (Proc. Sem., Australian Nat. Univ., Canberra, 1978), Vol. 6 of Monographs Textbooks Mech. Solids Fluids: Mech. Anal., Sijthoff & Noordhoff, Alphen aan den Rijn, 1980, pp. 119–134.

[17] J. M. Varah, A practical examination of some numerical methods for linear discrete ill-posed problems, SIAM Rev. 21 (1) (1979) 100–111. doi:10.1137/1021007.
URL http://dx.doi.org/10.1137/1021007

[18] J. M. Varah, Pitfalls in the numerical solution of linear ill-posed problems, SIAM J. Sci. Statist. Comput. 4 (2) (1983) 164–176. doi:10.1137/0904012.

URL http://dx.doi.org/10.1137/0904012

[19] R. Courant, D. Hilbert, Methods of mathematical physics. Vol. I, Interscience Publishers, Inc., New York, N.Y., 1953.

[20] F. Smithies, The eigenvalue and singular values of integral equations, Proc. London Math. Soc. 43 (1937) 255–279.

[21] J. Mercer, Functions of positive and negative type, and their connection with the theory of integral equations, Philosophical Transactions of the Royal Society of London. Series A, 209 (1909) pp. 415–446.
URL http://www.jstor.org/stable/91043

[22] G. H. Golub, C. F. Van Loan, Matrix computations, 3rd Edition, Johns Hopkins University Press, Baltimore, MD, USA, 1996.
URL http://portal.acm.org/citation.cfm?id=248979

[23] P. C. Hansen, The discrete Picard condition for discrete ill-posed problems, BIT Numerical Mathematics 30 (1990) 658–672, 10.1007/BF01933214.
URL http://dx.doi.org/10.1007/BF01933214

[24] P. Hansen, Computation of the singular value expansion, Computing 40 (1988) 185–199, 10.1007/BF02251248.
URL http://dx.doi.org/10.1007/BF02251248

[25] M. Bôcher, An Introduction to the Study of Integral Equations, Cambridge tracts in mathematics and mathematical physics, Cambridge University Press, 1909.
URL http://archive.org/details/introductiontost00bcuoft

[26] H. Weyl, Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung), Mathematische Annalen 71 (1912) 441–479, 10.1007/BF01456804.
URL http://dx.doi.org/10.1007/BF01456804

[27] E. Hille, J. D. Tamarkin, On the characteristic values of linear integral equations, Acta Math. 57 (1) (1931) 1–76. doi:10.1007/BF02403043.
URL http://dx.doi.org/10.1007/BF02403043

[28] G. Little, J. B. Reade, Eigenvalues of analytic kernels, SIAM J. Math. Anal. 15 (1) (1984) 133–136. doi:10.1137/0515009.
URL http://dx.doi.org/10.1137/0515009

[29] J. A. Cochran, The analysis of linear integral equations, McGraw-Hill Book Co., New York, 1972, mcGraw-Hill Series in Modern Applied Mathematics.

[30] J. B. Reade, On the sharpness of Weyl's estimate for eigenvalues of smooth kernels, SIAM J. Math. Anal. 16 (3) (1985) 548–550. doi:10.1137/0516040.
URL http://dx.doi.org/10.1137/0516040

[31] J. Reade, Eigenvalues of positive definite kernels II, SIAM Journal on Mathematical Analysis 15 (1) (1984) 137–142. arXiv:http://epubs.siam.org/doi/pdf/10.1137/0515010, doi:10.1137/0515010.
URL http://epubs.siam.org/doi/abs/10.1137/0515010

[32] A. Wathen, S. Zhu, On spectral distribution of kernel matrices related to radial basis functions, submitted (2012).

[33] T. A. Driscoll, B. Fornberg, Interpolation in the limit of increasingly flat radial basis functions, Comput. Math. Appl. 43 (3-5) (2002) 413–422, radial basis functions and partial differential equations. doi:10.1016/S0898-1221(01)00295-4.
URL http://dx.doi.org/10.1016/S0898-1221(01)00295-4

[34] B. Fornberg, G. Wright, E. Larsson, Some observations regarding interpolants in the limit of flat radial basis functions, Comput. Math. Appl. 47 (1) (2004) 37–55. doi:10.1016/S0898-1221(04)90004-1.
URL http://dx.doi.org/10.1016/S0898-1221(04)90004-1

[35] P. C. Hansen, Regularization tools version 4.0 for matlab 7.3, Numerical Algorithms 46 (2007) 189–194.

[36] R. A. Renaut, Examining noise in the basis vectors of the singular value decomposition, paper in preparation (2012).

[37] P. C. Hansen, M. E. Kilmer, R. H. Kjeldsen, Exploiting residual information in the parameter choice for discrete ill-posed problems, BIT 46 (1) (2006) 41–59. doi:10.1007/s10543-006-0042-7.
URL http://dx.doi.org/10.1007/s10543-006-0042-7

[38] B. W. Rust, D. P. O'Leary, Residual periodograms for choosing regularization parameters for ill-posed problems, Inverse Problems 24 (3) (2008) 034005, 30. doi:10.1088/0266-5611/24/3/034005.
URL http://dx.doi.org/10.1088/0266-5611/24/3/034005

[39] B. W. Rust, Truncating the singular value decomposition for ill-posed problems, Tech. Rep. NISTIR 6131, Technical Report, National Institute of Standards and Technology (1998).
URL http://math.nist.gov/ BRust/pubs/TruncSVD/MS-TruncSVD.ps

[40] K. Ball, Eigenvalues of Euclidean distance matrices, J. Approx. Theory 68 (1) (1992) 74–82. doi:10.1016/0021-9045(92)90101-S.
URL http://dx.doi.org/10.1016/0021-9045(92)90101-S

[41] K. Ball, N. Sivakumar, J. D. Ward, On the sensitivity of radial basis interpolation to minimal data separation distance, Constr. Approx. 8 (4) (1992) 401–426. doi:10.1007/BF01203461.
URL http://dx.doi.org/10.1007/BF01203461

[42] B. J. C. Baxter, Norm estimates for inverses of Toeplitz distance matrices, J. Approx. Theory 79 (2) (1994) 222–242. doi:10.1006/jath.1994.1126.
URL http://dx.doi.org/10.1006/jath.1994.1126

[43] X.-P. Sun, Norm estimates for inverses of Euclidean distance matrices, J. Approx. Theory 70 (3) (1992) 339–347. doi:10.1016/0021-9045(92)90064-U.

URL http://dx.doi.org/10.1016/0021-9045(92)90064-U

[44] F. J. Narcowich, J. D. Ward, Norms of inverses for matrices associated with scattered data, in: Curves and surfaces (Chamonix-Mont-Blanc, 1990), Academic Press, Boston, MA, 1991, pp. 341–348.