Full-rank orthonormal bases for conditionally positive definite kernel based spaces

Maryam Mohammadi, Stefano De Marchi, Mohammad Karimnejad Esfahani

October 9, 2023

Abstract

In kernel-based approximation, it is well-known that the direct approach to interpolation is prone to ill conditioning of the interpolation matrix. One simple idea is to use other better-conditioned bases which span the same space of the translated kernels i.e. their associated *native* space. Pazouki and Schaback (2011) [1] tracked this issue by investigating different factorization of the interpolation matrix in order to build stable and orthonormal bases for the corresponding native space of the positive definite kernels. In this paper, we work with the reproducing kernel K for the native space $\mathcal{N}_\Phi$ corresponding to conditionally positive definite kernel Φ. We give a well-organized matrix formulation of the evaluation matrix $\mathbf{K}$ by constructing the matrices corresponding to cardinal bases from monomials. Then, we present two possible ways to find full-rank data-dependent orthonormal bases that are discretely ℓ_2 and $\mathcal{N}_\Phi$-orthonormal. The first approach is given by the factorization of the kernel matrix $\mathbf{K}$ and the next one is based on the eigenpairs approximation of linear operator associated with the reproducing kernel K given by Mercer's theorem. In the sequel, we employ the truncated singular value decomposition technique to find an optimal low-rank basis with the coefficient matrix whose rank is less than that of the original matrix. Special attention is also given to error analysis, duality, and stability. Some numerical experiments are also provided.

1 Introduction

One of the most basic problems in approximation theory is to construct an approximation of an unknown function f defined on a set $\Omega \subset \mathbb{R}^d$ from n specified distinct points $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\} \subset \Omega$. A simple approach consists of choosing n functions and then looking for the unique combination of these functions which effectively fits the data at the specified points X. To ensure the success of this procedure, the set of chosen functions must be linearly

independent over the set of interpolation points (also referred as data sites or centers) X. In this setting, the so-called kernel methods are of growing importance. The kernel can be Positive Definite (PD) or Conditionally Positive Definite (CPD).

Definition 1. *Let $\Phi : \Omega \times \Omega \to \mathbb{R}$ be a continuous symmetric kernel. It is said that Φ is a conditionally positive semi-definite kernel of order m on $\Omega \subset \mathbb{R}^d$ if, for all $n \in \mathbb{N}$, all pairwise distinct centers $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^d$, and all $\alpha \in \mathbb{R}^n$ satisfying*

$$\sum_{j=1}^{n} \alpha_j p(\mathbf{x}_j) = 0, \quad p \in \mathbb{P}_{m-1}^d,$$

the quadratic form

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j \Phi(\mathbf{x}_i, \mathbf{x}_j) \geq 0.$$

Moreover, Φ is said to be conditionally positive definite (CPD) of order m if equality holds only for $\alpha = 0$. Finally, when $m = 0$ the kernel is positive (semi)-definite, i.e., conditionally positive (semi)-definite kernels of order zero are positive (semi)-definite kernels.

Let Φ be a CPD kernel of order m and and $p_1, ..., p_q$ be a basis for the polynomial space $\mathbb{P}_{m-1}^d$. Then, for the $\mathbb{P}_{m-1}^d$-unisolvent set of data sites X and function values $f(\mathbf{x}_j) = f_j \in \mathbb{R}$, $1 \leq j \leq n$, the interpolant of unknown function f can be written as

$$s_f(\mathbf{x}) = \sum_{j=1}^{n} c_j \Phi(\mathbf{x}, \mathbf{x}_j) + \sum_{j=1}^{q} d_j p_j(\mathbf{x}), \ \forall\, \mathbf{x} \in \mathbb{R}^d. \tag{1}$$

In order to compute the coefficients c_j and d_j in (1), we may ask that s_f exactly reproduce the function values $\{f_j\}_{j=1}^n$. This leads to the linear system

$$\begin{bmatrix} A & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}, \tag{2}$$

with

$$A = [\Phi(\mathbf{x}_i, \mathbf{x}_j)]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}, \quad P = [p_j(\mathbf{x}_i)]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq q}}, \quad \mathbf{f} = [f(\mathbf{x}_j)]_{1 \leq j \leq n},$$

which is uniquely solvable (cf. e.g. [2, Chap. 8]) To investigate the linear system (2), consider the particular case $m = 0$ i.e., considering the *positive definite* kernels. In this case, the corresponding interpolation system would be

$$A\mathbf{c} = \mathbf{f} \tag{3}$$

Although the systems (3) are built to be well-posed for every data distribution, it is also well-known (see e.g [3]) that the interpolation based on translates of radial basis functions (RBFs) or non-radial kernels is numerically unstable due to the ill-conditioning of the *kernel matrix* A. Therefore, it is natural to devise strategies to prevent such instabilities by either preconditioning the system (see e.g [4]), or by finding a better basis for the approximation space we are using. The latter case gave rise to *stable algorithms* and has been introduced in [5] for the particular case of multiquadric kernels, and extended later to kernels on the sphere in [6]. Another approach for the construction of a better alternate basis for PD kernels has been introduced in [7] and was extended later in [1]. The main idea is to produce orthonormal data-dependent bases that span the associated *native space* $\mathcal{N}_\Phi$ by decomposing the kernel matrix A using different factorization techniques, such as SVD or Cholesky factorization. This has led to different bases with different properties. For these new bases stability issues, recursive compatibility, duality and orthogonality properties were investigated.

Following such an idea, in [8] a particular orthonormal basis built on a *weighted singular value decomposition* of the kernel matrix has been introduced. This basis is also related to a discretization of the compact integral operator T_Φ given by Mercer's theorem and provides a connection with the continuous basis that arises from an eigendecomposition of T_Φ.

Although effective, this basis is computationally expensive to compute, so in [9] the authors discussed methods related to *Krylov* subspaces to compute this basis in a fast way. Finally, in [10] the authors provide a new way to compute and evaluate *Gaussian* RBF interpolants in a stable way by using *Hilbert-Schmidt* series expansions of positive definite kernels.

Coming back to the CPD kernels, the linear system (2) may also suffer from ill-conditioning for some constellations of the interpolation points (see [11]). However, in contrast with the PD case, the literature contains very

few contributions that address finding more stable basis for CPD kernels. An exception is [12] in which the authors tried to extend the previous work in [1] to the CPD case. But in their idea, it is impossible to have a full orthonormal basis of n functions if $q > 0$. Explicitly, it is shown that one can not simply use factorization techniques due to the augmented polynomial space, and therefore some care needs to be taken. Another approach is taken in [13] to find bases that are in a certain sense *homogeneous*, meaning that they are not sensitive to poorly scaled problems. Some numerical results regarding these *homogeneous* bases are also reported in [14, Chap. 34].

In this paper, we present some possible ways to find *full-rank data-dependent orthonormal bases* that are discretely ℓ_2 and $\mathcal{N}_\Phi$-orthornormal. The paper is organized as follows. In section (2) a brief review of the native space regarding CPD kernels and their formulation in matrix form is provided. Section (3) contains an analysis of different approaches to obtain more stable bases that are full-rank orthonormal and span the same native space. Subsequently, in section (4) we investigate the interpolant representation with respect to the new bases and its error bounds. Section 5 is dedicated to analyzing dual of the new bases and their relation with the corresponding evaluation matrices. Finally, in section (6) some numerical tests are presented.

2 Preliminaries

In this section, we briefly review some basic notions regarding reproducing kernel which is related to the CPD kernels. A vast discussion can be found in [2, Chap. 10].

2.1 CPD kernels and associated native space

As mentioned before, the linear system arising in (2) becomes very ill-conditioned as the number of the data sites X is increased [14]. Therefore it is natural to devise strategies to prevent such instabilities by finding a more stable basis for the approximation space. The process of finding a "better" basis for CPD kernels is closely connected to finding the reproducing kernel of the associated

"native" space. To begin with, given $\alpha \in \mathbb{R}^n$, $\mathbf{x}_j \in \Omega$ such that

$$\sum_{j=1}^{n} \alpha_j p(\mathbf{x}_j) = 0 \text{ for all } p \in \mathbb{P}_{m-1}^d,$$

then the space

$$H_\Phi(\Omega) = \left\{ f : f = \sum_{j=1}^{n} \alpha_j \Phi(\cdot, \mathbf{x}_j) \right\},$$

is a pre-Hilbert space equipped with the inner product

$$\langle f, g \rangle_\Phi = \left\langle \sum_{j=1}^{n} \alpha_j \Phi(\cdot, \mathbf{x}_j), \sum_{k=1}^{m} \beta_k \Phi(\cdot, \mathbf{y}_k) \right\rangle_\Phi = \sum_{j=1}^{n} \sum_{k=1}^{m} \alpha_j \beta_k \Phi(\mathbf{x}_j, \mathbf{y}_k),$$

and the corresponding Hilbert-space by completion $\mathcal{H}_\Phi(\Omega) = \overline{H_\Phi(\Omega)}$. Now we define the mapping

$$\begin{aligned} &R : \mathcal{H}_\Phi(\Omega) \to C(\Omega), \\ &R(f(\mathbf{x})) = f(\mathbf{x}) - \Pi f(\mathbf{x}) = f(\mathbf{x}) - \sum_{k=1}^{q} f(\boldsymbol{\xi}_k) l_k(\mathbf{x}), \end{aligned}$$

where l_k, $1 \le k \le q$, are the Lagrange basis of $\mathbb{P}_{m-1}^d$ for the points $\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_q\}$ which is assumed to be a $\mathbb{P}_{m-1}^d$-unisolvent subset of X. Notice that $q = \dim(\mathbb{P}_{m-1}^d)$.

Definition 2. *The native space corresponding to a symmetric kernel Φ that is CPD of order m on Ω is defined by*

$$\mathcal{N}_\Phi(\Omega) = R(\mathcal{H}_\Phi(\Omega)) \oplus \mathbb{P}_{m-1}^d,$$

equipped with the inner product

$$\langle f, g \rangle_{\mathcal{N}_\Phi} = \langle f, g \rangle + \sum_{k=1}^{q} f(\boldsymbol{\xi}_k) g(\boldsymbol{\xi}_k),$$

where

$$\langle f, g \rangle = \left\langle R^{-1}(f - \Pi f), R^{-1}(g - \Pi g) \right\rangle_\Phi .$$

With this inner product, $\mathcal{N}_\Phi(\Omega)$ becomes a reproducing-kernel Hilbert space with the kernel

$$\begin{aligned} K(\mathbf{x},\mathbf{y}) &= \Phi(\mathbf{x},\mathbf{y}) - \sum_{k=1}^{q} l_k(\mathbf{x})\Phi(\boldsymbol{\xi}_k,\mathbf{y}) - \sum_{r=1}^{q} l_r(\mathbf{y})\Phi(\mathbf{x},\boldsymbol{\xi}_r) \\ &\quad + \sum_{k=1}^{q}\sum_{r=1}^{q} l_k(\mathbf{x}) l_r(\mathbf{y})\Phi(\boldsymbol{\xi}_k,\boldsymbol{\xi}_r) + \sum_{k=1}^{q} l_k(\mathbf{x}) l_k(\mathbf{y}). \end{aligned} \tag{4}$$

An advantage of having found the reproducing kernel K is that we can express the kernel-based interpolant of some function f at a given data set X as

$$s_f(\mathbf{x}) = \sum_{j=1}^{n} \alpha_j K(\mathbf{x},\mathbf{x}_j), \quad \mathbf{x} \in \mathbb{R}^d.$$

Note that the kernel K used here is a PD kernel (since it is a reproducing kernel) with built-in polynomial precision. The coefficients α_j are then determined by the interpolation conditions

$$s_f(\mathbf{x}_i) = f(\mathbf{x}_i), \quad i = 1,\ldots,n.$$

In order to obtain a better conditioned interpolation system, we aim to find different bases for the approximation space.

2.2 Matrix formulation

We start by providing an explicit representation for the matrix $\mathbf{K} = [K(\mathbf{x}_i,\mathbf{x}_j)]_{\substack{1\le i\le n\\ 1\le j\le n}}$ by considering a cardinal basis of the polynomial-based space. In fact, the Lagrange basis $l = [l_1,\ldots,l_q]$, can be expressed by the standard monomials $\widetilde{m} = [\widetilde{m}_1,\ldots,\widetilde{m}_q]$ (cf. e.g. [15]) via

$$l(\mathbf{x}) = \widetilde{m}(\mathbf{x}) \cdot C_l, \quad \forall \mathbf{x} \in \Omega \tag{5}$$

where C_l is known as the *construction* matrix. Letting

$$[l_j(\boldsymbol{\xi}_i)]_{\substack{1\le i\le q\\ 1\le j\le q}} = I,$$

and the Vandermonde matrix

$$V = [\widetilde{m}_j(\boldsymbol{\xi}_i)]_{\substack{1 \le i \le q \\ 1 \le j \le q}}$$

then by (5)

$$C_l = V^{-1}.$$

If the Lagrange basis is needed at another set of evaluation points, say $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_s\}$, by equation (5) we get

$$V^T L_Y^T = V_Y^T,$$

where

$$L_Y = [l_j(\mathbf{y}_i)]_{\substack{1 \le i \le s \\ 1 \le j \le q}} \quad , \quad V_Y = [\widetilde{m}_j(\mathbf{y}_i)]_{\substack{1 \le i \le s \\ 1 \le j \le q}} \quad .$$

Hence, for the kernel matrix $\mathbf{K} = K(\mathbf{x}_i, \mathbf{x}_j)$ with

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) \;=\; & \Phi(\mathbf{x}_i, \mathbf{x}_j) - \sum_{k=1}^{q} l_k(\mathbf{x}_i)\Phi(\boldsymbol{\xi}_k, \mathbf{x}_j) - \sum_{r=1}^{q} l_r(\mathbf{x}_j)\Phi(\mathbf{x}_i, \boldsymbol{\xi}_r) \\ & + \sum_{k=1}^{q}\sum_{r=1}^{q} l_k(\mathbf{x}_i) l_r(\mathbf{x}_j)\Phi(\boldsymbol{\xi}_k, \boldsymbol{\xi}_r) + \sum_{k=1}^{q} l_k(\mathbf{x}_i) l_k(\mathbf{x}_j), \quad i, j = 1, \ldots, n. \end{aligned} \tag{6}$$

we get

$$\mathbf{K} = A - L_1 \cdot A_1 - A_2 \cdot L_1^T + L_1 \cdot A_3 \cdot L_1^T + L_1 \cdot L_1^T,$$

where

$$\begin{aligned} L_1 &= [l_k(\mathbf{x}_i)]_{\substack{1 \le i \le n \\ 1 \le k \le q}} \quad , \qquad A_1 = [\Phi(\boldsymbol{\xi}_k, \mathbf{x}_j)]_{\substack{1 \le k \le q \\ 1 \le j \le n}} \quad , \\ A_2 &= [\Phi(\mathbf{x}_i, \boldsymbol{\xi}_r)]_{\substack{1 \le i \le n \\ 1 \le r \le q}} \quad , \quad A_3 = [\Phi(\boldsymbol{\xi}_k, \boldsymbol{\xi}_r)]_{\substack{1 \le k \le q \\ 1 \le r \le q}} \quad . \end{aligned}$$

To end this section, we list the most commonly used global CPD RBFs $\phi(r)$ in Table 1, where β is the RBF parameter and ε is the *shape parameter* that decide the flatness of the basis and can be found numerically for getting accurate numerical solutions and conditioning of the collocation matrix.

Table 1: Some global CPD RBFs

Name	$\phi(r)$	Condition	Order (m)
Generalized Multiquadrics (MQ)	$(-1)^{\lceil\beta\rceil}\left(1+(\epsilon r)^2\right)^{\beta}$	$0<\beta\notin\mathbb{N}$	$\lceil\beta\rceil$
Radial powers	$(-1)^{\lceil\frac{\beta}{2}\rceil}r^{\beta}$	$0<\beta\notin 2\mathbb{N}$	$\lceil\frac{\beta}{2}\rceil$
Thin-plate splines (TPS)	$(-1)^{\beta+1}r^{2\beta}\log(r)$	$\beta\in\mathbb{N}$	$\beta+1$

3 Full-rank orthonormal bases

In what follows, we investigate suitable bases for subspaces of $\mathcal{N}_\Phi(\Omega)$ when Φ is CPD. Hence, let Φ be a fixed CPD kernel with corresponding reproducing kernel K in (6), $X=\{\mathbf{x}_1,...,\mathbf{x}_n\}$ a fixed set of centers, and $U=[u_1,\ldots,u_n]$ a general data-dependent basis such that

$$\mathcal{N}_\Phi(X)=\operatorname{span}\{K(\cdot,\mathbf{x}_j):\mathbf{x}_j\in X\}=\operatorname{span}\{u_1,...,u_n\}\subset\mathcal{N}_\Phi(\Omega).$$

Following [1], any element of the basis U can be written as a linear combinations of the translates $K(\cdot,\mathbf{x}_j),\ j=1,...,n$ via the construction matrix C

$$u_i=\sum_{j=1}^{n}K(\cdot,\mathbf{x}_j)c_{ji},\quad 1\le i\le n, \tag{7}$$

or in the matrix form

$$E=\mathbf{K}C, \tag{8}$$

where $E=[u_j(\mathbf{x}_i))]_{1\le i,j\le n}$. In what follows, instead of $\mathcal{N}_\Phi(X)$ we use $\mathcal{N}_\Phi$.

Theorem 1. *The $\mathcal{N}_\Phi$ and ℓ_2 Gramian matrices associated with the general basis U are symmetric and positive definite with full-rank n.*

Proof. The Gramian matrices associated with the basis U corresponding to

the $\mathcal{N}_\Phi$ and ℓ_2 are

$$\begin{aligned} G_{\mathcal{N}_\Phi} &= [\langle u_i, u_j\rangle_{\mathcal{N}_\Phi}]_{\substack{1\le i\le n\\ 1\le j\le n}} = C^T\mathbf{K}C, \\ G_{\ell_2} &= [\langle u_i, u_j\rangle_{\ell_2}]_{\substack{1\le i\le n\\ 1\le j\le n}} = \left(\sum_{k=1}^{n} u_i(\mathbf{x}_k)u_j(\mathbf{x}_k)\right)_{\substack{1\le i\le n\\ 1\le j\le n}} = E^TE = C^T\mathbf{K}^2C. \end{aligned}$$

The evaluation matrix E is necessarily full-rank because the basis must allow unique interpolation on X. Since $C = \mathbf{K}^{-1}E$ the construction matrix C is also full-rank, it results the same for $G_{\mathcal{N}_\Phi}$ and G_{ℓ_2}.The matrices $G_{\mathcal{N}_\Phi}$ and G_{ℓ_2} are clearly symmetric. Now, since $\mathbf{K}$ is a positive definite matrix then, for all nonzero vectors $\mathbf{z} \in \mathbb{R}^n$, we have

$$\mathbf{z}^T \cdot G_{\mathcal{N}_\Phi} \cdot \mathbf{z} = (C\mathbf{z})^T\mathbf{K}(C\mathbf{z}) > 0.$$

and since E is a full-rank matrix similarly

$$\mathbf{z}^T \cdot G_{\ell_2} \cdot \mathbf{z} = \langle E\mathbf{z}, E\mathbf{z}\rangle = \|E\mathbf{z}\|_2 > 0.$$

We have then show that $G_{\mathcal{N}_\Phi}$ and G_{ℓ_2} are positive definite. □

Remark 1. *Suppose that we construct the evaluation matrix E through the augmented system*

$$\begin{bmatrix} E_{n\times n} \\ 0_{q\times n} \end{bmatrix} = \begin{bmatrix} A_{n\times n} & P_{n\times q} \\ P^T_{q\times n} & 0_{q\times q} \end{bmatrix} \begin{bmatrix} \widetilde{C}_{n\times n} \\ \widetilde{D}_{q\times n} \end{bmatrix}.$$

The moment conditions $P^T\widetilde{C} = 0$, reveals that the $n \times n$ matrix C has rank $n-q$ and the evaluation matrix E is necessarily rank n, since the basis must allow unique interpolation on X. Then the Gramian matrix $G_{\mathcal{N}_\Phi} = \widetilde{C}^TE$ is symmetric and positive semi-definite with rank $n-q$. So it is impossible to have a full orthonormal bases if $q > 0$.

The above remark highlights why we prefer to use the reproducing kernel rather than the standard bases with augmented polynomials. Besides, it must be noted that we assumed that the space $\mathcal{N}_\Phi(X)$ is fixed, meaning that the data sites X have been specified once. This clarifies why the new bases U are data-dependent.

In the following, we address two possible ways to find data-dependent orthonormal bases corresponding to the CPD kernel Φ on a domain $\Omega \subset \mathbb{R}^d$.

3.1 Matrix decomposition approach

According to [1] equation (8) reveals that one can find data-dependent basis U from the decomposition of the symmetric and positive definite matrix $\mathbf{K}$ corresponding to the CPD kernel Φ since

$$\mathbf{K} = EC^{-1}.$$

We can characterize $\mathcal{N}_\Phi$ and discretely ℓ_2 orthonormal bases based on the Gramian matrices as follows:

1) For $\mathcal{N}_\Phi-$orthonormal bases, we have

$$G_{\mathcal{N}_\Phi} = I \Longleftrightarrow C^T\mathbf{K}C = I \Longleftrightarrow \mathbf{K} = (C^{-1})^T C^{-1} \Longleftrightarrow E = (C^{-1})^T.$$

Then, there are two important cases.

i) The Choleskey decomposition $\mathbf{K} = LL^T$ with a nonsingular lower triangular matrix L which leads to the Newton basis with a different normalization [7]. In this case $E = L$ and $C = (L^T)^{-1}$.

ii) The singular value decomposition (SVD) decomposition of the form $\mathbf{K} = QDQ^T$ with an orthogonal matrix Q and a diagonal matrix D having the eigenvalues of $\mathbf{K}$ on its diagonal. In this case $E = Q\sqrt{D}$ and $C = Q\left(\sqrt{D}\right)^{-1}$.

2) For ℓ_2-orthonormal bases, we have

$$G_{\ell_2} = I \Longleftrightarrow C^T\mathbf{K}^T\mathbf{K}C = I \Longleftrightarrow \mathbf{K}C = Q \Longleftrightarrow \mathbf{K} = QC^{-1} \Longleftrightarrow E = Q.$$

Also here there are two important special cases.

i) The standard QR decomposition $\mathbf{K} = QR$ into an orthogonal matrix Q and an upper triangular matrix R will lead to a basis with $E = Q$ and $C = R^{-1}$.

ii) The SVD of $\mathbf{K} = QDQ^T$ which leads to $E = Q$ and $C = Q(D)^{-1}$.

3.2 Eigenpairs approximation approach

We discuss another family of orthonormal bases based on the eigenvalues and eigenfunctions of *Hilbert-Schmidt* operator [16, Chap.2] associated with the reproducing kernel K given in (4). Mercer's theorem expresses the connection of such a linear operator with the infinite series representation of a positive definite kernel.

Theorem 2. ***(Mercer's theorem)*** *Let K be a continuous positive definite kernel that satisfies*

$$\int_{\Omega} K(\mathbf{x},\mathbf{y})v(\mathbf{x})v(\mathbf{y})d\mathbf{x}d\mathbf{y} \geq 0, \quad \forall v \in L_2(\Omega),\ \mathbf{x},\mathbf{y} \in \Omega.$$

Then K can be represented by

$$K(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{\infty} \lambda_j \widetilde{u}_j(\mathbf{x})\widetilde{u}_j(\mathbf{y}), \tag{9}$$

where λ_j are the eigenvalues and $\widetilde{u}_j$ are the L_2-orthonormal eigenfunctions of the operator $T_K : L_2(\Omega) \to L_2(\Omega)$ given by

$$T_K(v)(\mathbf{x}) = \int_{\Omega} K(\mathbf{x},\mathbf{y})v(\mathbf{y})d\mathbf{y}, \quad v \in L_2(\Omega),\ \mathbf{x} \in \Omega.$$

Moreover, this representation is absolutely and uniformly convergent.

Theorem 2 can lead to another characterization of the Native space $\mathcal{N}_\Phi(\Omega)$ as

$$\mathcal{N}_\Phi(\Omega) = \left\{ f : f = \sum_{j=1}^{\infty} c_j \widetilde{u}_j \right\},$$

where the kernel K itself is in $\mathcal{N}_\Phi(\Omega)$ because of the eigenfunction expansion (9). The reproducing property of the kernel K should be checked by the following equation

$$\langle f(\cdot), K(\cdot,\mathbf{x})\rangle_{\mathcal{N}_\Phi} = \left\langle \sum_{j=1}^{\infty} c_j \widetilde{u}_j(\cdot), \sum_{j=1}^{\infty} \lambda_j u_j(\cdot)\widetilde{u}_j(\mathbf{x}) \right\rangle_{\mathcal{N}_\Phi} = \sum_{j=1}^{\infty} c_j \widetilde{u}_j(\mathbf{x}) = f(\mathbf{x}),$$

which leads to the $\mathcal{N}_\Phi(\Omega)$-orthogonality of the eigenfunctions

$$\langle \widetilde{u}_i, \widetilde{u}_j \rangle_{\mathcal{N}_\Phi} = \frac{\delta_{ij}}{\sqrt{\lambda_i}\sqrt{\lambda_j}}. \tag{10}$$

The inner product for $\mathcal{N}_\Phi(\Omega)$ is then given by

$$\langle f, g \rangle_{\mathcal{N}_\Phi} = \left\langle \sum_{j=1}^{\infty} c_j \widetilde{u}_j, \sum_{i=1}^{\infty} d_i \widetilde{u}_i \right\rangle_{\mathcal{N}_\Phi} = \sum_{j=1}^{\infty} \frac{c_j d_j}{\lambda_j}.$$

Equation (10) reveals two important cases for the basis functions.

i) Basis functions

$$\{u_j\}_{j=1}^{\infty} = \left\{ \sqrt{\lambda_j} \widetilde{u}_j \right\}_{j=1}^{\infty}, \quad \|u_j\|_{\mathcal{N}_\Phi}^2 = 1, \quad \|u_j\|_{L_2}^2 = \lambda_j \tag{11}$$

which is orthonormal in $\mathcal{N}_\Phi(\Omega)$ and orthogonal in $L_2(\Omega)$.

ii) Basis functions

$$\{v_j\}_{j=1}^{\infty} = \{\widetilde{u}_j\}_{j=1}^{\infty}, \quad \|v_j\|_{\mathcal{N}_\Phi}^2 = \frac{1}{\lambda_j}, \quad \|v_j\|_{L_2}^2 = 1 \tag{12}$$

which is orthogonal in $\mathcal{N}_\Phi(\Omega)$ and orthonormal in $L_2(\Omega)$.

Unfortunately, in most cases, eigenpairs of the operator T_K are not known analytically. The exception is the Gaussian kernel which is a PD kernel by definition. On the other hand, to our knowledge, no research has been conducted on investigating the analytical form of the eigenpairs related to CPD kernels in (4). Thus it will be required to approximate them using numerical schemes. This leads to the following eigenvalue problem on X

$$\int_\Omega K(\mathbf{x}_i, \mathbf{y}) \widetilde{u}_j(\mathbf{y}) d\mathbf{y} = \lambda_j \widetilde{u}_j(\mathbf{x}_i), \quad i = 1, \ldots, n, \forall j > 0,$$

which can be discretized by using the *symmetric Nyström method* [17], giving

$$\sum_{r=1}^{n} K(\mathbf{x}_i, \mathbf{x}_r) \widetilde{u}_j(\mathbf{x}_r) w_r \approx \lambda_j \widetilde{u}_j(\mathbf{x}_i), \quad i, j = 1, \ldots, n, \tag{13}$$

with a set of positive weights $\{w_r\}_{r=1}^n$. Equation (13) can be re-written in matrix form

$$(\mathbf{K}W)\widetilde{\mathbf{e}}^{(j)} = \lambda_j \widetilde{\mathbf{e}}^{(j)}, \quad j = 1, \ldots, n,$$

with

$$\begin{aligned} W &= \mathrm{diag}(w_r), \\ \widetilde{\mathbf{e}}^{(j)} &= [\widetilde{u}_j(\mathbf{x}_i))]_{1 \le i \le n} \ . \end{aligned} \tag{14}$$

Then, the continuum eigenvalue problem reduces the solution of an unsymmetric eigenvalue problem

$$(\mathbf{K}W)\widetilde{\mathbf{e}} = \lambda \widetilde{\mathbf{e}}, \tag{15}$$

for the positive definite matrix $\mathbf{K}W$. One possible way to deal with (15) would be to make some manipulation to convert the unsymmetric problem of (15) to the following symmetric one

$$(\sqrt{W}\mathbf{K}\sqrt{W})(\sqrt{W} \cdot \widetilde{\mathbf{e}}) = \lambda(\sqrt{W} \cdot \widetilde{\mathbf{e}}).$$

Now, the SVD decomposition for the symmetric matrices, which is nothing that a unitary diagonalization, leads to

$$\sqrt{W}\mathbf{K}\sqrt{W} = QDQ^T, \tag{16}$$

where $D = \mathrm{diag}(\lambda_j)$,

$$Q = \left[\sqrt{W}\widetilde{\mathbf{e}}^{(1)}, \ldots, \sqrt{W}\widetilde{\mathbf{e}}^{(n)}\right],$$

is an orthogonal matrix w.r.t the Euclidean norm. Equations (11), (12), and (14) lead to the evaluation matrices

i) $E_1 = [u_j(\mathbf{x}_i))]_{\substack{1 \le i \le n \\ 1 \le j \le n}} = \left[\sqrt{\lambda_j}\widetilde{u}_j(\mathbf{x}_i))\right]_{\substack{1 \le i \le n \\ 1 \le j \le n}} = \left[\sqrt{\lambda_1}\widetilde{\mathbf{e}}^{(1)}, \ldots, \sqrt{\lambda_n}\widetilde{\mathbf{e}}^{(n)}\right] = \left(\sqrt{W}\right)^{-1} Q\sqrt{D}.$

ii) $E_2 = [v_j(\mathbf{x}_i))]_{\substack{1 \le i \le n \\ 1 \le j \le n}} = [\widetilde{u}_j(\mathbf{x}_i))]_{\substack{1 \le i \le n \\ 1 \le j \le n}} = \left[\widetilde{\mathbf{e}}^{(1)}, \ldots, \widetilde{\mathbf{e}}^{(n)}\right] = \left(\sqrt{W}\right)^{-1} Q.$

According to (8) and (16), the corresponding construction matrices can be derived

$$i)\ C_U = \mathbf{K}^{-1}E_1 = \sqrt{W}QD^{-1}Q^T\sqrt{W}\left(\sqrt{W}\right)^{-1}Q\sqrt{D} = \sqrt{W}Q\left(\sqrt{D}\right)^{-1}$$

$$ii)\ C_V = \mathbf{K}^{-1}E_2 = \sqrt{W}QD^{-1}Q^T\sqrt{W}\left(\sqrt{W}\right)^{-1}Q = \sqrt{W}QD^{-1}.$$

By considering the discretized scaled inner product

$$\langle f, g\rangle^2_{\ell_{2,w}} = \sum_{j=1}^{n} w_j f(\mathbf{x}_j)g(\mathbf{x}_j) \approx \langle f, g\rangle^2_{L_2(\Omega)} = \int_\Omega f(\mathbf{x})g(\mathbf{x})d\mathbf{x},$$

we have $\mathcal{N}_\Phi-$orthonormal and $\ell_{2,w}-$orthogonal basis functions

$$\begin{aligned} u_j(\mathbf{x}) &= \sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)c_{ij} = \sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)\frac{\sqrt{w_i}}{\sqrt{\lambda_j}}Q(i,j) \\ &= \sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)\frac{\sqrt{w_i}}{\sqrt{\lambda_j}}\frac{\sqrt{w_i}}{\sqrt{\lambda_j}}E_1(i,j) = \frac{1}{\lambda_j}\sum_{i=1}^{n} w_i K(\mathbf{x},\mathbf{x}_i)u_j(\mathbf{x}_i), \end{aligned}$$

with $\|u_j\|^2_{\ell_{2,w}} = \lambda_j$, and $\mathcal{N}_\Phi-$orthogonal and $\ell_{2,w}-$orthonormal basis functions

$$\begin{aligned} v_j(\mathbf{x}) &= \sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)c_{ij} = \sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)\frac{\sqrt{w_i}}{\lambda_j}Q(i,j) \\ &= \sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)\frac{\sqrt{w_i}}{\lambda_j}\sqrt{w_i}E_2(i,j) = \frac{1}{\lambda_j}\sum_{i=1}^{n} w_i K(\mathbf{x},\mathbf{x}_i)v_j(\mathbf{x}_i), \end{aligned}$$

with $\|v_j\|^2_{\mathcal{N}_\Phi} = \frac{1}{\lambda_j}$.

Remark 2. *According to this theory, we can deduce that all data-dependent bases, which are both discretely and $\mathcal{N}_\phi$-orthogonal, are scaled SVD bases derived from the eigenpairs approximation approach. It is also clear that the SVD bases given in 3.1 are special cases of the general bases given in this section.*

Remark 3. *The reader should note that the expansion series in* (9) *is valid only for the PD kernels. So, working with the associated reproducing kernel of a CPD kernel rather than the standard basis itself, enables us to use Mercer's theorem and find two additional classes of bases.*

3.3 Low-rank approximation

As stated before, one of the defining features of a reproducing kernel (invertibility of the kernel matrix) was lost in the translation between theory and implementation since the *numerical rank* of K is often much lower than n and so the matrix $\mathbf{K}$ is ill-conditioned. In other words, for many kernels, the eigenvalues in (9) decrease very rapidly toward zero, and this implies that there is a very good low-rank approximation to the kernel. Notice that vectors involved in kernel representation in (9), are of infinite size and so need to be truncated at some finite length M possibly mush smaller than n. Accordingly, we have the following theorem from [18].

Theorem 3. *Let $K : \Omega \times \Omega \to \mathbb{R}$ be a PD kernel with Mercer series* (9)*. Then, M-term truncation*

$$K_M(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{M} \lambda_n \widetilde{u}_j(\mathbf{x}) \widetilde{u}_j(\mathbf{y}), \tag{17}$$

for a fixed $\mathbf{x}$ provides the best M-term least squares approximation of $K(\mathbf{x}, \mathbf{y})$ from $L_2(\Omega)$.

The summation (17) yields the best $M-$term approximation of each kernel matrix in $L_2(\Omega)$ norm, but this in not necessarily the best low-rank approximation in the 2-norm sense. Therefore we consider SVD low-rank representation (truncated SVD) of the kernel matrix $\mathbf{K}$ which is obtained by discarding all but the k largest eigenvalues and the corresponding eigenvectors and is represented as

$$\mathbf{K}_k = Q_k D_k Q_k^T, \tag{18}$$

where $Q_k \in \mathbb{R}^{n \times k}$ and $D_k \in \mathbb{R}^{k \times k}$. It means that $\mathbf{K}_k$ is the projection of the $\mathbf{K}$ onto the space spanned by the top k eigenvectors of $\mathbf{K}$. The followings state that the above approximation is the best rank-k approximation in both *Frobenius* and *spectral* norm.

Theorem 4. *(**Eckart-Young** [19]) Let A_k be the rank-k approximation of A achieved by truncated SVD. Then A_k is the closest rank-k matrix to A, i.e.*

$$\min_{\text{rank}(G)=k} \|A - G\|_F = \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2},$$

where σ_i's denote singular values of A and G is an arbitrary rank-k matrix.

Remark 4. *SVD also gives the best low rank approximation in spectral norm i.e.*

$$\min_{\text{rank}(G)=k} \|E - G\|_2 = \|E - E_k\|_2 = \sigma_{k+1}.$$

So the rank-reduced system will be very close to the exact system but with a more well-behaved linear system with better-conditioned value matrix [20]. Accordingly, the evaluation matrix of the new bases can be represented as

$$\begin{aligned} E_{1k} &= Q_k\sqrt{D_k}, \\ E_{2k} &= Q_k, \end{aligned} \tag{19}$$

such that $E_{1k}, E_{2k} \in \mathbb{R}^{n\times k}$.

Remark 5. *One should notice that in order to consider the truncated SVD, it requires that there exists well-defined gap in the singular values, i.e., $\frac{\sigma_{k+1}}{\sigma_k}$ must be large enough. Otherwise determination of the optimal rank k would be complicated and low rank approximation of matrix $\mathbf{K}$ is meaningless. A detailed discussion is available in [20, Chap 12.2].*

4 Application to interpolation

4.1 General interpolant

Having derived different types of data-dependent bases U, the interpolant $s_f \in \mathcal{N}_\Phi$ to vector values $\mathbf{f}$ of some function f, can be represented as

$$s_f(\mathbf{x}) = \sum_{j=1}^{n} \alpha_j u_j(\mathbf{x}),$$

where the coefficients α_j are determined by solving the linear system

$$E\alpha = \mathbf{f}, \tag{20}$$

where $\alpha = [\alpha_j]_{1\le j\le n}$ and $E = [u_j(\mathbf{x}_i))]_{\substack{1\le i\le n\\ 1\le j\le n}}$ can be one of the evaluation matrices obtained in Subsections 3.1 or 3.2. Once the coefficient vector α

is calculated through (20), one can obtain the approximate function values $\mathcal{F}_Y \approx \mathbf{f}_Y = [f(\mathbf{y}_i)]_{1 \le i \le s}$ at the set of test points $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_s\}$ by

$$\mathcal{F}_Y = E_Y \cdot \alpha, \tag{21}$$

where $E_Y = [u_j(\mathbf{y}_i)]_{\substack{1 \le i \le s \\ 1 \le j \le n}}$ is obtained by (8) as

$$E_Y = \mathbf{K}_Y C,$$

where C is the corresponding construction matrices and $\mathbf{K}_Y = [K(\mathbf{y}_i, \mathbf{x}_j)]_{\substack{1 \le i \le s \\ 1 \le j \le n}}$ can be computed via the same procedure explained at the end of Section 2.

Theorem 5. *The evaluation matrices of the $\mathcal{N}_\Phi$ and $\ell_{2,w}-$orthonormal basis functions are better conditioned than the kernel matrix $\mathbf{K}$.*

Proof. If E is the evaluation matrix corresponding to a $\ell_{2,w}-$orthonormal basis then it is an orthogonal matrix and so $\mathrm{cond}_{2,w}(E) = 1$. Now let E be the evaluation matrix corresponding to a $\mathcal{N}_\Phi-$orthonormal basis derived from the general scaled SVD bases, then

$$E = \left(\sqrt{W}\right)^{-1} Q\sqrt{D}.$$

Moreover according to (16), we have

$$\mathbf{K} = (\sqrt{W})^{-1} Q\sqrt{D}\sqrt{D}Q^T(\sqrt{W})^{-1} = EE^T,$$

and

$$(\sqrt{D})^{-1} Q^T\sqrt{W}EE^T\sqrt{W}Q(\sqrt{D})^{-1} = I.$$

Therefore

$$\widehat{Q} = (\sqrt{D})^{-1} Q^T\sqrt{W}E,$$

is an orthogonal matrix w.r.t the norm $\|\cdot\|_{\ell_{2,w}}$, which in turn gives

$$E = \left(\sqrt{W}\right)^{-1} Q\sqrt{D}\widehat{Q},$$

that is nothing with the SVD of the matrix E. Therefore the spectral condition number of E is the square root of the spectral condition number of $\mathbf{K}$. The same theory can be used for the Newton basis functions given in 3.1. $\square$

Remark 6. *If linear maps $\mathcal{L}$ like derivatives have to be evaluated, we use the system*

$$\mathcal{L}E_Y = \mathcal{L}\mathbf{K}_Y C,$$

where $\mathcal{L}E_Y = [\mathcal{L}u_j(\mathbf{y}_i)]_{\substack{1 \le i \le s \\ 1 \le j \le n}}$ *and* $\mathcal{L}\mathbf{K}_Y = [\mathcal{L}\mathbf{K}(\mathbf{y}_i, \mathbf{x}_j)]_{\substack{1 \le i \le s \\ 1 \le j \le n}}$ *is given by applying the operator $\mathcal{L}$ to the reproducing kernel (6) and doing the same procedure explained at the end of Section 2.*

4.2 Error bound

In this section we provide the error estimate for the approximation given in (21). First the following stability issue is proved.

Theorem 6. *For a fixed CPD kernel Φ, fixed set of center points $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$, general data-dependent bases U, and $f \in \mathcal{N}_\Phi$, the following stability estimate holds for the approximate function values $\mathcal{F}_Y$ at the set of test points $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_s\}$,*

$$\|\mathcal{F}_Y\|_2^2 \le s \cdot \rho(\widetilde{\mathbf{K}}) \cdot \mathrm{cond}_2(\mathrm{G}_{\mathcal{N}_\Phi}) \cdot \|\mathrm{f}\|_{\mathcal{N}_\Phi}^2,$$

where $\mathrm{cond}_2(\mathrm{G}_{\mathcal{N}_\Phi})$ *is the spectral condition number of the $\mathcal{N}_\Phi-$Gramian, ρ is the spectral radius, and* $\widetilde{\mathbf{K}} = [K(\mathbf{y}_i, \mathbf{y}_i)]_{\substack{1 \le i \le s \\ 1 \le i \le s}}$ *for the corresponding reproducing kernel K.*

Proof. Since Frobenius norm is compatible with the euclidean norm, according to (21), we have

$$\|\mathcal{F}_Y\|_2^2 = \|E_Y \cdot \alpha\|_2^2 \le \|E_Y\|_F^2 \, \|\alpha\|_2^2 \,. \tag{22}$$

Now according to (20), we get

$$\alpha^T G_{\mathcal{N}_\Phi} \alpha \le \|f\|_{\mathcal{N}_\Phi}^2.$$

Since

$$\alpha^T G_{\mathcal{N}_\Phi} \alpha = \langle \alpha, G_{\mathcal{N}_\Phi} \alpha \rangle \le \|\alpha\|_2 \|G_{\mathcal{N}_\Phi} \alpha\|_2 \le \|\alpha\|_2^2 \|G_{\mathcal{N}_\Phi}\|_2 = \rho(G_{\mathcal{N}_\Phi}) \|\alpha\|_2^2,$$

then

$$\rho(G_{\mathcal{N}_\Phi}) \|\alpha\|_2^2 \le \|f\|_{\mathcal{N}_\Phi}^2.$$

Therefore

$$\|\alpha\|_2^2 \leq \|f\|_{\mathcal{N}_\Phi}^2 \rho(G_{\mathcal{N}_\Phi}^{-1}). \tag{23}$$

Moreover, we have

$$\mathbf{K}_Y^{(i)} \mathbf{K}^{-1} \left(\mathbf{K}_Y^{(i)}\right)^T = K(\mathbf{y}_i, \mathbf{y}_i) - P_{\Phi,X}^2(\mathbf{y}_i), \quad i = 1, \ldots, s$$

where $\mathbf{K}_Y^{(i)}$ is the $i-$th row of the matrix $\mathbf{K}_Y$ and $P_{\Phi,X}$ is the so-called power function. Now according to (8), we get

$$E_Y^{(i)} C^{-1} \mathbf{K}^{-1} \left(C^{-1}\right)^T \left(E_Y^{(i)}\right)^T = K(\mathbf{y}_i, \mathbf{y}_i) - P_{\Phi,X}^2(\mathbf{y}_i), \quad i = 1, \ldots, s$$

where $E_Y^{(i)}$ is the i-th row of the matrix E_Y. Therefore

$$E_Y^{(i)} \left(G_{\mathcal{N}_\Phi}\right)^{-1} \left(E_Y^{(i)}\right)^T = K(\mathbf{y}_i, \mathbf{y}_i) - P_{\Phi,X}^2 \mathbf{y}_i) \leq K(\mathbf{y}_i, \mathbf{y}_i), \quad i = 1, \ldots, s$$

which leads to

$$\|E_Y^{(i)}\|_2^2 \leq K(\mathbf{y}_i, \mathbf{y}_i) \rho(G_{\mathcal{N}_\Phi}), \quad i = 1, \ldots, s$$

Now since

$$\|E_Y\|_F^2 = \sum_{i=1}^{s} \|E_Y^{(i)}\|_2^2,$$

we have

$$\|E_Y\|_F^2 \leq \mathrm{tr}(\widetilde{\mathbf{K}}) \rho(G_{\mathcal{N}_\Phi}) \leq s \cdot \rho(\widetilde{\mathbf{K}}) \cdot \rho(G_{\mathcal{N}_\Phi}). \tag{24}$$

So by substituting (23) and (24) in (22), the proof is completed. □

Theorem 7. *For a fixed CPD kernel Φ, general data-dependent bases U, and $f \in \mathcal{N}_\Phi$, the following error bound holds*

$$\|\mathbf{f}_Y - \mathcal{F}_Y\|_2^2 \leq \left(s \cdot \rho(\widetilde{\mathbf{K}}) - \rho(G_{\mathcal{N}_\Phi}^{-1}) \|E_Y\|_F^2\right) \|f\|_{\mathcal{N}_\Phi}^2$$

Proof.

$$\begin{aligned}
\|\mathbf{f}_Y - \mathcal{F}_Y\|_2^2 = \sum_{i=1}^{s} |f(\mathbf{y}_i) - s_f(\mathbf{y}_i)|^2 &\leq \sum_{i=1}^{s} P_{\Phi,X}^2(\mathbf{y}_i)\|f\|_{\mathcal{N}_\Phi}^2 \\
&= \sum_{i=1}^{s} \left(K(\mathbf{y}_i, \mathbf{y}_i) - E_Y^{(i)} \left(G_{\mathcal{N}_\Phi}\right)^{-1} \left(E_Y^{(i)}\right)^T \right) \|f\|_{\mathcal{N}_\Phi}^2 \\
&\leq \left(\mathrm{tr}(\widetilde{\mathbf{K}}) - \sum_{i=1}^{s} \left(\rho(G_{\mathcal{N}_\Phi}^{-1}) \|E_Y^{(i)}\|_2^2 \right) \right) \|f\|_{\mathcal{N}_\Phi}^2 \\
&\leq \left(s \cdot \rho(\widetilde{\mathbf{K}}) - \rho(G_{\mathcal{N}_\Phi}^{-1}) \|E_Y\|_F^2 \right) \|f\|_{\mathcal{N}_\Phi}^2.
\end{aligned}$$

□

Remark 7. *For the pointwise behaviour of the $\mathcal{N}_\Phi-$orthonormal basis U, the bounds obtained above become*

$$\begin{aligned}
|s_f(\mathbf{y})| &\leq \sqrt{K(\mathbf{y}, \mathbf{y})} \cdot \|f\|_{\mathcal{N}_\Phi}, \\
|f(\mathbf{y}) - s_f(\mathbf{y})| &\leq \sqrt{K(\mathbf{y}, \mathbf{y}) - \|U(\mathbf{y})\|_2^2} \cdot \|f\|_{\mathcal{N}_\Phi},
\end{aligned}$$

for fixed $\mathbf{y} \in \Omega$.

5 Duality

The goal of this section is to construct new class of bases that are dual to the general data-dependent bases $U = [u_1, \ldots, u_n]$, proposed for the finite-dimensional inner product subspace $\mathcal{N}_\Phi$ of the native space $\mathcal{N}_\Phi(\Omega)$ associated to the CPD kernel Φ. The dual space ${\mathcal{N}_\Phi}^*$ consists of all linear functionals on $\mathcal{N}_\Phi$. Consider the dual functionals η_i such that

$$\eta_i(\alpha_1 u_1 + \cdots + \alpha_n u_n) = \alpha_i, \quad \alpha_i \in \mathbb{R},\ i = 1, \ldots, n,$$

which in turn leads to

$$\eta_i(u_j) = \delta_{ij}.$$

Then any linear functional $\eta \in {\mathcal{N}_\Phi}^*$ can be written as

$$\eta = \eta(u_1)\eta_1 + \eta(u_2)\eta_2 + \cdots + \eta(u_n)\eta_n.$$

Now by the Riesz Representation Theorem, every linear functional on ${\mathcal{N}_\Phi}^*$ has a representer in $\mathcal{N}_\Phi$. That is, for each η_i, there exists $d_i \in \mathcal{N}_\Phi$ such that

$$\eta_i(u_j) = \langle u_j, d_i \rangle = \delta_{ij}. \tag{25}$$

Therefore, we associate $\Lambda = [\eta_1, \ldots, \eta_n]$ with the representers $\mathcal{D} = [d_1, \ldots, d_n]$. Since Λ is linearly independent in ${\mathcal{N}_\Phi}^*$ and dual to U, then the so-called dual basis $\mathcal{D}$ is linearly independent in $\mathcal{N}_\Phi$ and also dual to U. Now let $(\mathcal{N}_\Phi, U, \mathcal{D})$ with basis $U = [u_1, \ldots, u_n]$ and dual basis $\mathcal{D} = [d_1, \ldots, d_n]$, then we can view the basis U as the map

$$\begin{aligned} U: \ & \mathbb{R}^n \to \mathcal{N}_\Phi \\ & \alpha \to U(\alpha) = \sum_{j=1}^{n} \alpha_j U_j, \end{aligned}$$

and likewise, the dual basis $\mathcal{D}$ as

$$\begin{aligned} \mathcal{D}: \ & \mathbb{R}^n \to \mathcal{N}_\Phi \\ & \alpha \to \mathcal{D}(\alpha) = \sum_{j=1}^{n} \alpha_j d_j. \end{aligned}$$

Also the following dual map for identifying the dual space ${\mathcal{N}_\Phi}^*$ with $\mathcal{N}_\Phi$

$$\begin{aligned} \mathcal{D}^*: \ & \mathcal{N}_\Phi \to \mathbb{R}^n \\ & f \to \mathcal{D}^*(f) = \left[\langle f, d_1 \rangle, \ldots, \langle f, d_n \rangle\right]^T. \end{aligned}$$

Then according to (25), $\mathcal{D}$ is dual to U exactly when

$$\mathcal{D}^*(U) = \left[\langle u_j, d_i \rangle_{\mathcal{N}_\Phi}\right]_{\substack{1 \le i \le n \\ 1 \le i \le n}} = I.$$

Theorem 8. *Let U be a general data-dependent basis, then for $(\mathcal{N}_\Phi, U, \mathcal{D})$, the dual basis $\mathcal{D}$ can be expressed in terms of the basis U as*

$$\mathcal{D} = U\mathcal{C},$$

where

$$\mathcal{C} = (U^*(U))^{-1},$$

is a symmetric, positive definite and full-rank $n \times n$ matrix.

Proof. Let $\mathcal{D} = U\mathcal{C}$, then by applying U^* to both sides, we get

$$U^*(\mathcal{D}) = U^*(U)\mathcal{C},$$

which leads to

$$\mathcal{C} = (U^*(U))^{-1}U^*(\mathcal{D}).$$

Since $\mathcal{D}$ is dual to U, this reduces to $\mathcal{C} = (U^*(U))^{-1}$, which is nothing with the inverse of the $\mathcal{N}_\Phi$-Gramian matrix as

$$\mathcal{C} = \left([\langle u_i, u_j \rangle_{\mathcal{N}_\Phi}]_{\substack{1 \le i \le n \\ 1 \le j \le n}} \right)^{-1} = (G_{\mathcal{N}_\Phi})^{-1},$$

that is symmetric and positive definite with rank n. □

Remark 8. *For $(\mathcal{N}_\Phi, T, \mathcal{D})$, with the basis of translates*

$$T = [K(\cdot, \mathbf{x}_1), \ldots, K(\cdot, \mathbf{x}_n))],$$

we have

$$\mathcal{C} = (T^*(T))^{-1} = \mathbf{K}^{-1},$$

then

$$\mathcal{D} = T\mathbf{K}^{-1}.$$

So the Lagrange basis and the basis T of translates are a dual pair.

Remark 9. *Among all data-dependent bases, the $\mathcal{N}_\Phi$-orthonormal bases are exactly those which are self-dual, since $\mathcal{C} = I$.*

Theorem 9. *Let U be a general data-dependent basis, then for $(\mathcal{N}_\Phi, U, \mathcal{D})$, the dual basis $\mathcal{D}$ can be expressed in terms of the basis T of translates as*

$$\mathcal{D} = T(E^T)^{-1},$$

where E is the evaluation matrix.

Proof. According to Theorem 8 and equation (7), we have

$$\mathcal{D} = U\mathcal{C} = TC(C^T\mathbf{K}C)^{-1} = T\mathbf{K}^{-1}(C^T)^{-1} = T(E^T)^{-1}.$$

□

Theorem 10. *Let V be the $\ell_{2,w}-$orthonormal basis functions proposed in Section 3.2, then for $(\mathcal{N}_\Phi, V, \mathcal{D})$, the dual basis $\mathcal{D}$ can be expressed in terms of the basis T of translates as*

$$\mathcal{D} = T\sqrt{W}Q.$$

Proof. According to the above theorem, we get

$$\mathcal{D} = T(E_2^T)^{-1} = T(Q^T(\sqrt{W})^{-1})^{-1} = T\sqrt{W}Q.$$

□

6 Numerical Experiments

For the numerical experiments, we consider three different underlying functions and three different types of CPD kernels all of order 2, namely

- Generalized MQ RBF with $\beta = \frac{3}{2}$.
- Cubic RBF $\varphi_c(r) = r^3$ which is shape parameter free;
- Thin plate spline RBF $\varphi_{tps}(r) = r^2 \log(r)$ which is shape parameter free too,

where $r = \|\mathbf{x} - \mathbf{y}\|_2$ with $\mathbf{x}, \mathbf{y} \in \Omega \subset \mathbb{R}^d$. In the following subsection, **standard basis** refers to any of the above RBFs appended by polynomial space of the required degree (1), and **Reproducing kernel** refers to the corresponding PD kernel in (4). Besides, by truncated SVD basis we mean the basis explained in subsection 3.3 such that the evaluation is selected to be E_{1k} in (19).

Moreover, working with generalized MQ RBF, one always needs to find the optimal value of shape parameter ε, which depends on the number and constellation of the data sites. In particular, ε values have significant effects both on the accuracy and stability of the interpolation process. However, we skip this task and we always let $\varepsilon = 1$, since our numerical experiments show that with the suggested alternate bases we obtain good accuracy even without

n	standard gMQ	Reproducing Kernel	SVD Bases
20	1.9687e +17	1.5468e+17	3.9330e+08
50	9.3105e+17	2.3023e+18	1.0946e+08
80	1.2990e+19	5.5690e+18	1.0174e+08
110	1.2637e+19	3.0754e+18	1.2553e+08
150	5.0118e+19	1.911e+19	1.3472e+08

Table 2: 2-norm condition number of the interpolation matrix for different bases; Test problem 1.

optimizing the shape parameter. Moreover, in order to compute the accuracy of the interpolation, the *root mean square error* (RMSE) is computed as

$$RMSE = \sqrt{\frac{1}{s}\sum_{i=1}^{s}(f(\mathbf{z}_i) - s_f(\mathbf{z}_i))^2} \tag{26}$$

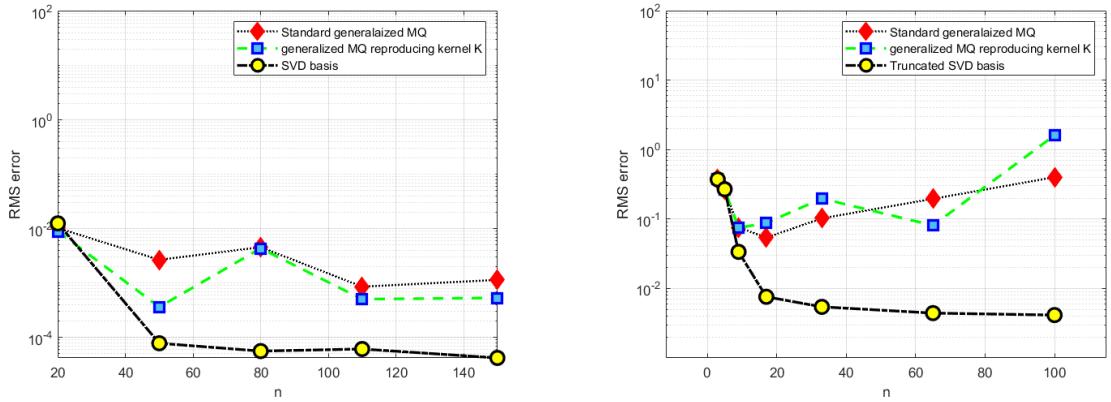
where $\{\mathbf{z}_i\}_{i=1}^{s}$ is the set of evaluation points.

6.1 Test problem 1

Let us consider the *Runge* function

$$f(x) = \frac{1}{1+25x^2}\,, \quad x \in [-1,1].$$

We reconstruct f using *uniformly distributed* center points with different sizes $n = \{20, 50, 80, 110, 150\}$. Regarding the size of data sites, our interpolant is evaluated over an equispaced point set on Ω with size $s = 5n$. To evaluate the reproducing kernel (4) we let $\Xi = \{0, 1\}$ to form the Lagrange linear bases for the polynomial space. Table 2 shows the ℓ_2 condition number of the interpolation matrix using different bases. It is observable that SVD bases lead to better conditioning. Figure 1(a) shows how more stable bases lead to better accuracy, particularly for an underlying function that is prone to inaccurate interpolation due to its intrinsic oscillatory behavior.



(a) (b)

Figure 1: RMSE of Runge's (a) and Franke's functions (b) approximants using different bases; Test problems 1 and 2.

n	standard gMQ	Reproducing Kernel	Truncated SVD
9	2.5275e+05	7.5530e+04	274.8278
25	9.9028e+08	4.5283e+08	2.1280e+04
81	2.7716e+15	1.4380e+15	2.4357e+05
289	2.8478e+19	8.9075e+18	5.1423e+05
1089	5.6001e+20	6.1388+19	9.5946e+05
4225	3.9644e+21	1.1233e+22	1.9277e+06
10000	5.5799e+22	3.2801e+21	2.8522e+06

Table 3: 2-norm condition number of interpolation matrix for different bases; Test problem 2.

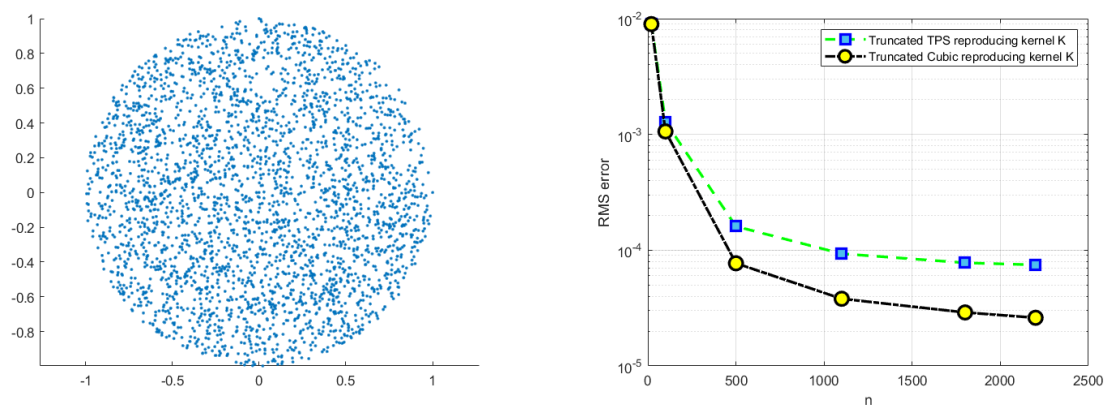
6.2 Test problem 2

For the second test problem, we take the *Franke* function, [14, Chap 2] defined on $\Omega = [0,1]^2 \subset \mathbb{R}^2$ as the target function. The interpolation this time is done at the sequence of *Halton* center points with different sizes $n = \{9, 25, 81, 289, 1089, 4225, 10000\}$. Moreover, let $\Xi = \{(0,0),(0,1),(1,0)\}$ representing the Lagrange linear polynomials. Similarly, for each n, the interpolant is evaluated over a uniform grid with size $s = 2n$ on the domain of interest. We consider truncated SVD bases, with threshold $\delta = 10^{-9}$, obtained by trial and error. It means zeroing all the eigenvalues of the interpolation matrix $\mathbf{K}$ which are smaller than δ. Table 3 shows the ℓ_2 condition number of the interpolation matrix for different bases. In Figure (1)(b), we show the RMSE of the interpolation using these 3 different bases. Once more,

we recall that we avoided any shape parameter value optimization algorithm and we just let $\varepsilon = 1$.

6.3 Test problem 3

Here, we reconstruct the oscillatory function $f(\mathbf{x}, \mathbf{y}) = \cos(20(\mathbf{x}+\mathbf{y}))$ defined on the unit disk with center $(0, 0)$. In order to do so, consider the data set X consisting of 3000 *Halton* points on the unit disk (see Figure 2(a)), Ξ as in the previous example, and the truncation sequence $k = \{20, 100, 500, 1100, 1800, 2400\}$ meaning that, in the first experiment, we take 20 singular values resulted from the SVD decomposition of the kernel matrix $\mathbf{K}$ in (6). We use φ_c and φ_{tps} RBFs to approximate $f(\mathbf{x}, \mathbf{y})$. To measure the accuracy of the reproduction process, we computed the RMSE on an equally spaced grid of evaluation points with size $s = 6000$ on the domain. Figure 2(b) shows that indeed there is a very good low-rank approximation to the problem.



(a) (b)

Figure 2: Data sites X and the RMSE resulted from Truncated SVD approximation for two different bases; Test problem 3.

Remark 10. *We have to highlight that according to our discussion in 2.1, one always needs to make sure that the set $\Xi \subset X$, meaning that the subset used to build the Lagrange polynomials must belong to the set of data sites.*

Remark 11. *In all three experiments, one can see that the RMSE resulting from SVD bases are stuck after some step. In other words, the increase of the number of data sites do not lead to an increase of accuracy. This behavior*

stems from the fact after some steps the singular values of the kernel matrix are too small and so they have only subtle effects on the interpolation.

7 Conclusion

Two different approaches have been presented to construct new stable bases for CPD kernels. Both of these approaches are based on working with reproducing kernel of the corresponding Native Space of CPD kernels. Inspired by [1], we investigated different factorizations of the kernel matrix to obtain other bases with different features. Besides, working with reproducing kernel that is always a PD kernel by definition, we relate CPD kernels to the Mercer theorem to find "natural" class of basis. We also investigated the dual bases of the general data-dependent bases.

Regarding the stability, the experiments confirm the good behavior of the new bases expected from the analysis conducted in the previous sections. More precisely, employing a low-rank approximation of the kernel matrix enables the handling of approximations involving a relatively large number of points also for not optimized shape parameters, and on quite general sets. From a numerical point of view, this procedure can be accomplished without thinning the data sites $X \subset \Omega$, but simply checking if the singular values of the kernel matrix decay under a certain tolerance.

Last but not least, as future work one may consider Remark 6, in order to employ all these new stable bases to solve PDE problems.
Another facet of the newly established foundation through SVD factorization is its inability to employ an adaptive algorithm for singular value computation, instead necessitating a complete matrix factorization for every fixed point distribution. In this case, we can refer to optimized eigenvalue algorithms for finding only a subset of the full spectrum of the kernel matrix such those presented for example in [21, 9].

Acknowledgments. This research has been accomplished within GNCS-INδAM, Rete ITaliana di Approssimazione (RITA), and the topic group on "Approximation Theory and Applications" of the Italian Mathematical Union (UMI).

References

[1] Maryam Pazouki and Robert Schaback. Bases for kernel-based spaces. *Journal of Computational and Applied Mathematics*, 236(4):575–588, 2011.

[2] Holger Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

[3] Stefano De Marchi and Robert Schaback. Stability of kernel-based interpolation. *Advances in Computational Mathematics*, 32:155–161, 2010.

[4] Richard K Beatson, Jeremy Levesley, and CT Mouat. Better bases for radial basis function interpolation problems. *Journal of Computational and Applied Mathematics*, 236(4):434–446, 2011.

[5] Bengt Fornberg and Grady Wright. Stable computation of multiquadric interpolants for all values of the shape parameter. *Computers & Mathematics with Applications*, 48(5-6):853–867, 2004.

[6] Bengt Fornberg and Cécile Piret. A stable algorithm for flat radial basis functions on a sphere. *SIAM Journal on Scientific Computing*, 30(1):60–80, 2008.

[7] Stefan Müller and Robert Schaback. A newton basis for kernel spaces. *Journal of Approximation Theory*, 161(2):645–655, 2009.

[8] Stefano De Marchi and Gabriele Santin. A new stable basis for radial basis function interpolation. *Journal of Computational and Applied Mathematics*, 253:1–13, 2013.

[9] Stefano De Marchi and Gabriele Santin. Fast computation of orthonormal basis for rbf spaces through krylov space methods. *BIT Numerical Mathematics*, 55:949–966, 2015.

[10] Gregory E Fasshauer and Michael J McCourt. Stable evaluation of gaussian radial basis function interpolants. *SIAM Journal on Scientific Computing*, 34(2):A737–A762, 2012.

[11] Richard K Beatson, Jon B Cherrie, and Cameron T Mouat. Fast fitting of radial basis functions: Methods based on preconditioned gmres iteration. *Advances in Computational Mathematics*, 11:253–270, 1999.

[12] Maryam Pazouki and Robert Schaback. Bases for conditionally positive definite kernels. *Journal of Computational and Applied Mathematics*, 243:152–163, 2013.

[13] Richard K Beatson, William A Light, and S Billings. Fast solution of the radial basis function interpolation equations: Domain decomposition methods. *SIAM Journal on Scientific Computing*, 22(5):1717–1740, 2001.

[14] Gregory E Fasshauer. *Meshfree approximation methods with MATLAB*, volume 6. World Scientific, 2007.

[15] Maryam Mohammadi and Maryam Bahrkazemi. Bases for polynomial-based spaces. *Journal of Mathematical Modeling*, 7(1):21–34, 2019.

[16] Gregory E Fasshauer and Michael J McCourt. *Kernel-based approximation methods using Matlab*, volume 19. World Scientific Publishing Company, 2015.

[17] Weimin Han and Kendall E Atkinson. *Theoretical numerical analysis: A functional analysis framework*. Springer, 2009.

[18] Erich Novak and Henryk Woźniakowski. *Tractability of Multivariate Problems: Standard information for functionals*, volume 2. European Mathematical Society, 2008.

[19] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[20] Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

[21] Maryam Bahrkazemi and Maryam Mohammadi. A strategy to estimate the optimal low-rank in incremental svd-based algorithms for recommender systems. *Intelligent Data Analysis*, 26(2):447–467, 2022.