PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Special issue of the "Seminari Padovani di Analisi Numerica 2018" (SPAN2018), Volume 11 · 2018 · Pages 23–39

Interpolation with uncoupled separable matrix-valued kernels

Dominik Wittwar[a] · Gabriele Santin[a] · Bernard Haasdonk[a]

Communicated by Emma Perracchione

Abstract

In this paper we consider the problem of approximating vector-valued functions over a domain Ω. For this purpose, we use matrix-valued reproducing kernels, which can be related to Reproducing kernel Hilbert spaces of vectorial functions and which can be viewed as an extension of the scalar-valued case. These spaces seem promising, when modelling correlations between the target function components, as the components are not learned independently of each other. We focus on the interpolation with such matrix-valued kernels. We derive error bounds for the interpolation error in terms of a generalized power-function and we introduce a subclass of matrix-valued kernels whose power-functions can be traced back to the power-function of scalar-valued reproducing kernels. Finally, we apply these kind of kernels to some artificial data to illustrate the benefit of interpolation with matrix-valued kernels in comparison to a componentwise approach.

1 Introduction

Kernel methods are useful tools for dealing with a wide variety of different tasks ranging from machine learning e.g. via Support Vector Machines (SVMs) ([4, 23, 27]), function approximation from scattered data ([8, 14]) and many more. Especially the approximation aspect can be employed for generating surrogate models to speed up expensive function evaluation, see [30]. In cases where the given output data or the desired target function is vector-valued, simple approaches which build individual models for each function component can still be very costly, if the output is high dimensional and the component models rely on independent data sets such that the union of those results in overly large sets. Additionally, approximating a vectorial function componentwise with identical ansatz spaces might be the wrong choice, e.g. in case of different frequencies. We thus propose the use of matrix-valued kernels which lead to surrogates that can deal with correlations between function components, respective structural properties of the target function, and therefore provide a more suitable model. For divergence-free kernels, matrix-valued kernel approximations have already been succesfully applied, see e.g. ([7, 13, 20, 9]).

This paper is structured as follows: In Section 2 we begin with an introduction to matrix-valued kernels and extend well-known properties from the scalar-valued case including error estimation. We then introduce a new subclass of matrix-valued kernels and study its properties in relation to the power-function which enables us to perform a-priori interpolation error estimation in Section 3. A numerical example in Section 4 illustrates the benefits of the matrix-valued ansatz when compared to the scalar-valued case. Finally, we conclude with some remarks and an outlook.

2 Reproducing kernel Hilbert spaces for matrix-valued kernels

In this section we want to give a short overview over the theory of matrix-valued kernels and their application in interpolation. As matrix-valued kernels are an extension of the well studied scalar-valued kernels, many of the following notions, properties and concepts are again suitable extensions of their scalar-valued counterparts. For a more extensive overview with regards to this topic and other approximation schemes involving matrix-valued kernels such as regression, we refer to literature, e.g. [1, 16, 21].

Definition 2.1 (Matrix-valued kernel)**.** Let Ω be a non empty set. We call a function $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ a matrix-valued kernel if

$$k(x,y) = k(y,x)^T \quad \forall\, x, y \in \Omega.$$

Definition 2.2 (Reproducing kernel Hilbert space (RKHS))**.** Let $\mathcal{H}$ denote a Hilbert space of $\mathbb{R}^m$-valued functions over a domain Ω with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and induced norm $\|\cdot\|_{\mathcal{H}}$. We call $\mathcal{H}$ an $\mathbb{R}^m$-reproducing kernel Hilbert space ($\mathbb{R}^m$-RKHS), if for all $x \in \Omega$ and $\alpha \in \mathbb{R}^m$ the directional point evaluation functional $\delta_x^\alpha : \mathcal{H} \to \mathbb{R}$ defined by

$$\delta_x^\alpha(f) := f(x)^T\alpha. \tag{1}$$

is bounded, i.e.

$$\|\delta_x^\alpha\|_{\mathcal{H}'} := \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{\delta_x^\alpha(f)}{\|f\|_{\mathcal{H}}} < \infty.$$

Here $\mathcal{H}'$ denotes the space of all linear bounded functionals mapping from $\mathcal{H}$ into $\mathbb{R}$.

[a]University of Stuttgart, Institute of Applied Analysis and Numerical Simulation (IANS)

Similar to the scalar-valued case, see for example [2], there exists a one-to-one correspondence between RKHS of vector-valued functions and positive definite matrix-valued kernels. A necessary concept for this is the notion of positive definiteness which is a straightforward extension from the scalar-valued case and is given as follows:

Definition 2.3 (Definiteness)**.** Let Ω be non empty and $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a matrix-valued kernel. For a finite set $X := \{x_1, \dots, x_n\} \subset \Omega$, $n \in \mathbb{N}$, we define the Gramian matrix $K \in \mathbb{R}^{mn\times mn}$ as the block matrix given by

$$K := k(X,X) := (k(x_i,x_j))_{i,j=1}^n = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{bmatrix}. \tag{2}$$

The kernel k is denoted as positive definite (p. d.), if for all $n \in \mathbb{N}$ and $X = \{x_1, \dots, x_n\} \subset \Omega$ the Gramian matrix K is positive semi-definite, i.e. it holds

$$\alpha^T K \alpha \geq 0 \quad \forall\, \alpha \in \mathbb{R}^{mn}. \tag{3}$$

The kernel is called strictly positive definite (s.p.d.) if for all $n \in \mathbb{N}$ and pairwise distinct $X = \{x_1, \dots, x_n\} \subset \Omega$ the Gramian matrix K is positive definite, i.e. it holds

$$\alpha^T K \alpha > 0 \quad \forall\, \alpha \in \mathbb{R}^{nm} \setminus \{0\}. \tag{4}$$

Furthermore, we will introduce the abbreviation

$$k(X,x)^T := k(x,X) := (k(x,x_i))_{i=1}^n = \begin{bmatrix} k(x,x_1) & \cdots & k(x,x_n) \end{bmatrix} \in \mathbb{R}^{m\times mn} \tag{5}$$

as it will be useful later on.

Going forward, for A, B symmetric matrices, we will use the notation $A \succeq B$ if $A - B$ is positive semi-definite and $A \succ B$ if $A - B$ is positive definite.

As mentioned before, every RKHS corresponds to a positive definite matrix-valued kernel and vice versa. We state this in the following theorem. A proof for operator-valued kernels, which include the finite dimensional matrix-case, can be found for example in [12].

Theorem 2.1 (One-to-one correspondence)**.** *Let $\mathcal{H}$ be an $\mathbb{R}^m$-RKHS. Then there exists a unique positive definite matrix-valued kernel $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ such that for all $x \in \Omega$, $\alpha \in \mathbb{R}^m$ and $f \in \mathcal{H}$*

$$k(\cdot,x)\alpha \in \mathcal{H} \quad \text{and} \quad \langle f, k(\cdot,x)\alpha\rangle_{\mathcal{H}} = f(x)^T\alpha. \tag{6}$$

Conversely, if $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ is a positive definite matrix-valued kernel, then there exists a unique Hilbert space $\mathcal{H}$ of $\mathbb{R}^m$-valued functions on Ω such that (6) holds.

In the scalar-valued case, there is an alternative characterization by means of feature maps, i.e. for a p.d. kernel there exists a mapping $\Phi : \Omega \to V$, where V is some Hilbert space, such that the reproducing kernel is given by

$$k(x,y) = \langle \Phi(x), \Phi(y)\rangle_V, \quad \forall\, x, y \in \Omega.$$

In the matrix-valued case this is no longer possible, as inner products are scalar-valued. Nonetheless, the concept can be adapted by allowing mappings $\Phi : \Omega \times \mathbb{R}^m \to V$ such that

$$\alpha^T k(x,y)\beta = \langle \Phi(x,\alpha), \Phi(y,\beta)\rangle_V, \quad \forall\, x, y \in \Omega,\ \alpha, \beta \in \mathbb{R}^m.$$

For further details we refer to [5, 15].

Lemma 2.2 (Closed subspaces are RKHS)**.** *Let $\mathcal{H}$ be an $\mathbb{R}^m$-RKHS. If $\mathcal{N} \subset \mathcal{H}$ is a closed subspace then $\mathcal{N}$ is also an $\mathbb{R}^m$-RKHS. Furthermore, if $\mathcal{N}$ is finite dimensional with orthonormal basis $(v_n)_{n=1}^{\dim\mathcal{N}}$, then the reproducing kernel $k_{\mathcal{N}} : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ of $\mathcal{N}$ is given by*

$$k_{\mathcal{N}}(x,y) = \sum_{n=1}^{\dim\mathcal{N}} v_n(x)v_n(y)^T. \tag{7}$$

Proof. By Definition 2.2 it is sufficient to show that the directional point evaluation functionals $\delta_x^\alpha : \mathcal{N} \to \mathbb{R}$ are bounded:

$$\|\delta_x^\alpha\|_{\mathcal{N}'} = \sup_{f\in\mathcal{N}\setminus\{0\}} \frac{\delta_x^\alpha(f)}{\|f\|_{\mathcal{N}}} \leq \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{\delta_x^\alpha(f)}{\|f\|_{\mathcal{H}}} = \|\delta_x^\alpha\|_{\mathcal{H}'} < \infty.$$

We now show that $k_{\mathcal{N}}$ as defined in (7) satisfies the reproducing property (6): Let $\alpha \in \mathbb{R}^m$, it holds

$$k_{\mathcal{N}}(\cdot,x)\alpha = \sum_{n=1}^{\dim\mathcal{N}} v_n(\cdot)\underbrace{v_n(x)^T\alpha}_{\in\mathbb{R}} \in \mathcal{N}$$

and for v_i with $i = 1, \dots, \dim\mathcal{N}$

$$\langle v_i, k_{\mathcal{N}}(\cdot,x)\alpha\rangle_{\mathcal{H}} = \sum_{n=1}^{\dim\mathcal{N}} \underbrace{\langle v_i, v_n\rangle_{\mathcal{H}}}_{=\delta_{in}} v_n(x)^T\alpha = v_i(x)^T\alpha.$$

Due to the linearity of the inner product $\langle f, k_{\mathcal{N}}(\cdot,x)\alpha\rangle_{\mathcal{H}} = f(x)^T\alpha$ holds for all $f \in \mathcal{N}$. □

In general we are interested in finite dimensional subspaces of $\mathcal{H}$ which are spanned by kernel evaluations $k(\cdot, x_i)\alpha$ for different centers $x_i \in X = \{x_1, \dots, x_n\} \subset \Omega$ and directions $\alpha \in \mathbb{R}^m$, i.e. we are considering subspaces $\mathcal{N}(X) \subset \mathcal{H}$ of the form

$$\mathcal{N}(X) := \operatorname{span}\{k(\cdot, x_i)\alpha |\, x_i \in X,\ \alpha \in \mathbb{R}^m\}. \tag{8}$$

Due to the reproducing property (6) the orthogonal projection operator $\Pi_{\mathcal{N}(X)} : \mathcal{H} \to \mathcal{N}(X)$, which is characterized by

$$\left\langle f - \Pi_{\mathcal{N}(X)}(f), g \right\rangle_{\mathcal{H}} = 0, \quad \forall\, g \in \mathcal{N}(X), \tag{9}$$

coincides with the interpolation operator $I_{\mathcal{N}(X)} : \mathcal{H} \to \mathcal{N}(X)$ which interpolates a given function $f \in \mathcal{H}$ on the set X by a function $I_{\mathcal{N}(X)}(f) \in \mathcal{N}(X)$, i.e.

$$f(x_i) = I_{\mathcal{N}(X)}(f)(x_i), \quad \forall\, x_i \in X.$$

Indeed, using $g = k(\cdot, x_i)\alpha \in \mathcal{N}(X)$ in (9) results in

$$\left(f(x_i) - \Pi_{\mathcal{N}(X)}(f)(x_i)\right)^T \alpha = 0, \quad \forall\, x_i \in X,\ \forall\, \alpha \in \mathbb{R}^m,$$

and therefore

$$f(x_i) = \Pi_{\mathcal{N}(X)}(f)(x_i), \quad \forall x_i \in X.$$

In summary, this means that the interpolant

$$\Pi_{\mathcal{N}(X)}(f) = \sum_{i=1}^{N} k(\cdot, x_i)\alpha_i \tag{10}$$

is characterized by solutions of the linear system

$$k(X,X)\boldsymbol{\alpha} := \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} =: f(X). \tag{11}$$

If the kernel k is strictly positive definite, system (11) admits a unique solution as the system matrix $k(X,X)$ is invertible. Therefore, an interpolant is always well defined even if the right hand side in (11) does not stem from the evaluation of a function $f \in \mathcal{H}$ on the set of centers X. In cases where the kernel is only positive definite, i.e. $k(X,X)$ is positive semi-definite, the system has in general no unique solution for arbitrary right hand sides. However, a solution still exists when $f \in \mathcal{H}$:

Lemma 2.3. *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a matrix-valued positive definite kernel, $X = \{x_1, \dots, x_n\} \subset \Omega$, $f \in \mathcal{H}$. Furthermore, let*

$$\Pi_{\mathcal{N}(X)}(f) = \sum_{i=1}^{n} k(\cdot, x_i)\alpha_i$$

be the orthogonal projection of f onto $\mathcal{N}(X)$, where $\boldsymbol{\alpha} := \begin{bmatrix} \alpha_1^T & \cdots & \alpha_n^T \end{bmatrix}^T \in \mathbb{R}^{mn}$. Then it holds

$$\boldsymbol{\alpha} \in k(X,X)^+ f(X) + \operatorname{null}(k(X,X)). \tag{12}$$

Here $k(X,X)^+$ denotes the Moore-Penrose pseudo inverse of $k(X,X)$ and $\operatorname{null}(k(X,X))$ denotes the null space of the matrix $k(X,X)$ consisting of all vectors which are mapped to the zero vector, i.e. $\operatorname{null}(k(X,X)) = \operatorname{span}\{\alpha \in \mathbb{R}^m | k(X,X)\alpha = 0\}$.

Proof. Let $e_1, \dots, e_m$ denote the standard basis of $\mathbb{R}^m$. By (9) the interpolant satisfies

$$\begin{aligned} f(x_l)^T e_j = \left\langle \Pi_{\mathcal{N}(X)}(f), k(\cdot, x_l)e_j \right\rangle_{\mathcal{H}} &= \left\langle k(\cdot, x_l)e_j, \Pi_{\mathcal{N}(X)}(f) \right\rangle_{\mathcal{H}} \\ &= \left\langle k(\cdot, x_l)e_j, k(\cdot, X)\boldsymbol{\alpha} \right\rangle_{\mathcal{H}} \\ &= e_j^T k(X, x_l)^T \boldsymbol{\alpha} \\ &= e_j^T k(x_l, X)\boldsymbol{\alpha}. \end{aligned}$$

Since this holds for all $l = 1, \dots, n$ and $j = 1, \dots, m$ we conclude

$$k(X,X)\boldsymbol{\alpha} = f(X), \tag{13}$$

i.e. $\boldsymbol{\alpha}$ solves (11). Let $\boldsymbol{\alpha}^* := k(X,X)^+ f(X)$ and by use of (13) we get

$$k(X,X)\boldsymbol{\alpha}^* = k(X,X)k(X,X)^+ f(X) = k(X,X)k(X,X)^+ k(X,X)\boldsymbol{\alpha} = k(X,X)\boldsymbol{\alpha} = f(X)$$

and therefore $\boldsymbol{\alpha}^*$ also solves (11) which implies $\boldsymbol{\alpha} - \boldsymbol{\alpha}^* \in \operatorname{null}(k(X,X))$. □

Following the above property it seems reasonable to define an approximation to a given function f in the subspace $\mathcal{N}(X)$ by

$$g(x) := k(x,X)k(X,X)^+ f(X) \approx f(x)$$

even if $f \notin \mathcal{H}$. In this case the interpolation property at the centers X can no longer be guaranteed as in the strictly positive definite case, as the linear system (13) might not have a solution.

Before we further investigate how the error between a function $f \in \mathcal{H}$ and its interpolant $\Pi_{\mathcal{N}(X)}(f)$ can be quantified, we will present a direct corollary in which we derive an alternative representation of the reproducing kernel on $\mathcal{N}(X)$:

Corollary 2.4 (Reproducing kernel of $\mathcal{N}(X)$)**.** *It holds*

$$k_{\mathcal{N}(X)}(x,y) = k(x,X)k(X,X)^{+}k(X,y).$$

Proof. By Lemma 2.3 we have

$$\Pi_{\mathcal{N}(X)}k(\cdot,x)\alpha = k(\cdot,X)k(X,X)^{+}k(X,x)\alpha.$$

It is therefore sufficient to show that for any $\alpha \in \mathbb{R}^m$

$$\Pi_{\mathcal{N}(X)}k(\cdot,x)\alpha = k_{\mathcal{N}(X)}(\cdot,x)\alpha.$$

To this end, we first show that $\Pi_{\mathcal{N}} : \mathcal{H} \to \mathcal{N}(X)$ is self-adjoint. For this purpose, let $f, g \in \mathcal{H}$, then it holds

$$\begin{aligned}\langle \Pi_{\mathcal{N}(X)}(f), g\rangle &= \underbrace{\langle \Pi_{\mathcal{N}(X)}(f), g - \Pi_{\mathcal{N}(X)}(g)\rangle}_{=0} + \langle \Pi_{\mathcal{N}(X)}(f), \Pi_{\mathcal{N}(X)}(g)\rangle \\ &= -\underbrace{\langle f - \Pi_{\mathcal{N}(X)}(f), \Pi_{\mathcal{N}(X)}(g)\rangle}_{=0} + \langle f, \Pi_{\mathcal{N}(X)}(g)\rangle \\ &= \langle f, \Pi_{\mathcal{N}(X)}(g)\rangle.\end{aligned}$$

By definition of the projection operator $\Pi_{\mathcal{N}(X)}k(\cdot,x)\alpha \in \mathcal{N}(X)$ and by the above it holds for any $f \in \mathcal{N}(X)$:

$$\left\langle f, \Pi_{\mathcal{N}(X)}k(\cdot,x)\alpha\right\rangle_{\mathcal{H}} = \left\langle \Pi_{\mathcal{N}(X)}(f), k(\cdot,x)\alpha\right\rangle_{\mathcal{H}} = \langle f, k(\cdot,x)\alpha\rangle_{\mathcal{H}} = f(x)^T\alpha.$$

□

The above corollary extends a well known result for scalar-valued kernels, see [19], which states that the reproducing kernel on a closed subspace is equal to the projection of the reproducing kernel on the entire space with regard to either argument. However, in the matrix-valued case this does not carry over immediately, as the kernel has to be weighted with a direction, since the kernel itself is not an element of the RKHS.

As a tool to measure the error between f and its interpolant we want to present the so called power-function, which for example was used in [22] for scalar-valued kernels:

Definition 2.4 (Power-function)**.** Let $\mathcal{H}$ be an $\mathbb{R}^m$-RKHS and $\mathcal{N} \subset \mathcal{H}$ be a closed subspace. Furthermore, let $\Pi_{\mathcal{N}} : \mathcal{H} \to \mathcal{N}$ denote the orthogonal projection onto $\mathcal{N}$. We define the power-function $\mathcal{P}_{\mathcal{N}} : \mathcal{H}' \to \mathbb{R}$ by

$$\mathcal{P}_{\mathcal{N}}(\lambda) := \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{|\lambda(f) - \lambda(\Pi_{\mathcal{N}}(f))|}{\|f\|_{\mathcal{H}}} \quad \text{for } \lambda \in \mathcal{H}'. \tag{14}$$

In the case where $\lambda = \delta_x^\alpha$, we might also use the notation

$$\mathcal{P}_{\mathcal{N}}^\alpha(x) := \mathcal{P}_{\mathcal{N}}(\delta_x^\alpha).$$

In other words, the power-function maps a linear operator λ to the norm of the composition of λ with the orthogonal projection onto $\mathcal{N}^\perp$:

$$\begin{aligned}\mathcal{P}_{\mathcal{N}}(\lambda) &= \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{|\lambda(f) - \lambda(\Pi_{\mathcal{N}}(f))|}{\|f\|_{\mathcal{H}}} \\ &= \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{|\lambda\circ(\mathrm{id} - \Pi_{\mathcal{N}})(f)|}{\|f\|_{\mathcal{H}}} \\ &= \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{|\lambda\circ\Pi_{\mathcal{N}^\perp}(f))|}{\|f\|_{\mathcal{H}}} \\ &= \|\lambda\circ\Pi_{\mathcal{N}^\perp}\|_{\mathcal{H}'}\end{aligned} \tag{15}$$

We want to remark that the above definition of the power-function is, in contrast to the power-function introduced in [25], independent of the function f and can be utilized to derive a-priori error bounds which we show in Corollary 2.7.

It is easy to see that for a nested sequence of closed subspaces $\mathcal{N}_1 \subset \mathcal{N}_2 \subset \dots$ the power-function is non-increasing, i.e. $P_{\mathcal{N}_1}(\lambda) \geq P_{\mathcal{N}_2}(\lambda) \geq \dots$. For general $\lambda \in \mathcal{H}'$ the evaluation of $P_{\mathcal{N}}(\lambda)$ is nontrivial, however, using the Riesz representer $v_\lambda \in \mathcal{H}$ of λ we obtain an alternative representation of $\mathcal{P}_{\mathcal{N}}(\lambda)$:

Corollary 2.5 (Alternative representation of the power-function)**.** *Let $\mathcal{H}$ be an $\mathbb{R}^m$-RKHS and $\mathcal{N} \subset \mathcal{H}$ be a closed subspace. Furthermore, let $\Pi_{\mathcal{N}} : \mathcal{H} \to \mathcal{N}$ denote the orthogonal projection onto $\mathcal{N}$ and $\mathcal{P}_{\mathcal{N}} : \mathcal{H}' \to \mathbb{R}$ the power-function. For any $\lambda \in \mathcal{H}'$ let $v_\lambda \in \mathcal{H}$ denote its Riesz representer. Then it holds*

$$\mathcal{P}_{\mathcal{N}}(\lambda) = \|v_\lambda - \Pi_{\mathcal{N}}(v_\lambda)\|_{\mathcal{H}} = \|\Pi_{\mathcal{N}^\perp}(v_\lambda)\|_{\mathcal{H}}.$$

Proof. It follows from the definition of the power-function (14)

$$\mathcal{P}_{\mathcal{N}}(\lambda) = \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{\langle v_\lambda, f-\Pi_{\mathcal{N}}(f)\rangle_{\mathcal{H}}}{\|f\|_{\mathcal{H}}}.$$

Since both $\Pi_{\mathcal{N}}$ and $\mathrm{id}-\Pi_{\mathcal{N}}$ are orthogonal projections by assumption and therefore self-adjoint, the Cauchy-Schwarz inequality yields

$$\mathcal{P}_{\mathcal{N}}(\lambda) = \sup_{f\in\mathcal{H}\setminus\{0\}} \frac{\langle v_\lambda - \Pi_{\mathcal{N}}(v_\lambda), f\rangle_{\mathcal{H}}}{\|f\|_{\mathcal{H}}} \le \|v_\lambda - \Pi_{\mathcal{N}}(v_\lambda)\|_{\mathcal{H}},$$

and equality is reached for $f = v_\lambda - \Pi_{\mathcal{N}}(v_\lambda)$. □

For the directional point evaluation functional δ_x^α the Riesz representer is given by the reproducing kernel $k(\cdot,x)\alpha$. Therefore, we can easily compute the power-function using the reproducing property of k on $\mathcal{H}$ and $k_{\mathcal{N}}$ on $\mathcal{N}$:

Corollary 2.6. *For any $x\in\Omega$ and $\alpha\in\mathbb{R}^m$ it holds*

$$\mathcal{P}_{\mathcal{N}}^\alpha(x)^2 = \alpha^T\left(k(x,x)-k_{\mathcal{N}}(x,x)\right)\alpha = \alpha^T k_{\mathcal{N}^\perp}(x,x)\alpha.$$

Proof. By Corollary 2.5 it holds

$$\begin{aligned}
\mathcal{P}_{\mathcal{N}}^\alpha(x)^2 &= \|k(\cdot,x)\alpha - \Pi_{\mathcal{N}}(k(\cdot,x)\alpha)\|_{\mathcal{H}}^2\\
&= \|k(\cdot,x)\alpha - k_{\mathcal{N}}(\cdot,x)\alpha\|_{\mathcal{H}}^2\\
&= \langle k(\cdot,x)\alpha - k_{\mathcal{N}}(\cdot,x)\alpha, k(\cdot,x)\alpha - k_{\mathcal{N}}(\cdot,x)\alpha\rangle_{\mathcal{H}}\\
&= \langle k(\cdot,x)\alpha, k(\cdot,x)\alpha\rangle_{\mathcal{H}} - 2\langle k(\cdot,x)\alpha, k_{\mathcal{N}}(\cdot,x)\alpha\rangle_{\mathcal{H}} + \langle k_{\mathcal{N}}(\cdot,x)\alpha, k_{\mathcal{N}}(\cdot,x)\alpha\rangle_{\mathcal{H}}\\
&= \alpha^T k(x,x)\alpha - 2\alpha^T k_{\mathcal{N}}(x,x)\alpha + \alpha^T k_{\mathcal{N}}(x,x)\alpha\\
&= \alpha^T\left(k(x,x)-k_{\mathcal{N}}(x,x)\right)\alpha\\
&= \alpha^T k_{\mathcal{N}^\perp}(x,x)\alpha.
\end{aligned}$$

Here the identity $\alpha^T(k(x,x)-k_{\mathcal{N}}(x,x))\alpha = \alpha^T k_{\mathcal{N}^\perp}\alpha$ follows from Corollary 2.4 as $\Pi_{\mathcal{N}^\perp} = \mathrm{id}-\Pi_{\mathcal{N}}$. □

Corollary 2.7 (Bound on the interpolation error)**.** *Let $\mathcal{H}$ be an $\mathbb{R}^m$-RKHS with reproducing kernel k, let $\mathcal{N}\subset\mathcal{H}$ be a closed subspace with reproducing kernel $k_{\mathcal{N}}$ and $\Pi_{\mathcal{N}}:\mathcal{H}\to\mathcal{N}$ the orthogonal projection onto $\mathcal{N}$. Then it holds for any $f\in\mathcal{H}$ and $\alpha\in\mathbb{R}^m$*

$$\left|\left(f(x)-(\Pi_{\mathcal{N}}(f))(x)\right)^T\alpha\right| \le \mathcal{P}_{\mathcal{N}}^\alpha(x)\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}} \le \mathcal{P}_{\mathcal{N}}^\alpha(x)\|f\|_{\mathcal{H}}, \quad \forall\, x\in\Omega \tag{16}$$

and

$$\begin{aligned}
\|f(x)-(\Pi_{\mathcal{N}}(f))(x)\|_2 &\le \|k(x,x)-k_{\mathcal{N}}(x,x)\|_2^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}},\\
\|f(x)-(\Pi_{\mathcal{N}}(f))(x)\|_\infty &\le \max_{i=1,\dots,m}|k(x,x)_{ii}-k_{\mathcal{N}}(x,x)_{ii}|^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}},\\
\|f(x)-(\Pi_{\mathcal{N}}(f))(x)\|_1 &\le \sqrt{m}\|k(x,x)-k_{\mathcal{N}}(x,x)\|_2^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}.
\end{aligned}$$

Here $\|\cdot\|_2$ denotes the spectral norm on $\mathbb{R}^{m\times m}$.

Proof. It holds

$$\begin{aligned}
\left|\left(f(x)-(\Pi_{\mathcal{N}}(f))(x)\right)^T\alpha\right| &= \left|\langle f-\Pi_{\mathcal{N}}(f), k(\cdot,x)\alpha\rangle_{\mathcal{H}}\right|\\
&= \left|\langle(\mathrm{id}-\Pi_{\mathcal{N}})(f), k(\cdot,x)\alpha\rangle_{\mathcal{H}}\right|\\
&= \left|\left\langle(\mathrm{id}-\Pi_{\mathcal{N}})^2(f), k(\cdot,x)\alpha\right\rangle_{\mathcal{H}}\right|\\
&= \left|\langle(\mathrm{id}-\Pi_{\mathcal{N}})(f), (\mathrm{id}-\Pi_{\mathcal{N}})(k(\cdot,x)\alpha)\rangle_{\mathcal{H}}\right|\\
&\le \|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}\|k(\cdot,x)\alpha - k_{\mathcal{N}}(\cdot,x)\alpha\|_{\mathcal{H}}\\
&= \|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}\mathcal{P}_{\mathcal{N}}^\alpha(x).
\end{aligned} \tag{17}$$

Choosing $\alpha = f(x)-(\Pi_{\mathcal{N}}(f))(x)$ and applying Corollary 2.6 we get

$$\begin{aligned}
\|f(x)-(\Pi_{\mathcal{N}}(f))(x)\|_2^2 &\le \left(\alpha^T\left(k(x,x)-k_{\mathcal{N}}(x,x)\right)\alpha\right)^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}\\
&\le \|\alpha\|_2\|(k(x,x)-k_{\mathcal{N}}(x,x)\|_2^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}
\end{aligned}$$

and after dividing by $\|\alpha\| = \|f(x)-(\Pi_{\mathcal{N}}(f))(x)\|_2$

$$\|f(x)-(\Pi_{\mathcal{N}}(f))(x)\|_2 \le \|k(x,x)-k_{\mathcal{N}}(x,x)\|_2^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}$$

Choosing $\alpha = e_i$ results in

$$|f(x)_i - (\Pi_{\mathcal{N}}(f)(x))_i| \le \|f\|_{\mathcal{H}}\mathcal{P}_{\mathcal{N}}^{e_i}(x) \le |k(x,x)_{ii}-k_{\mathcal{N}}(x,x)_{ii}|^{1/2}\|f-\Pi_{\mathcal{N}}(f)\|_{\mathcal{H}}$$

Maximization over $i\in\{1,\dots,m\}$ gives the desired bound.
The last inequality follows directly from the first inequality and the inequality $\|\cdot\|_1\le\sqrt{m}\|\cdot\|_2$ on $\mathbb{R}^m$. □

3 Separable matrix-valued kernels

In order to practically solve interpolation problems, we need to take a look at how matrix-valued kernels can be constructed. To this end, we consider matrix-valued kernels which stem from scalar-valued kernels. In particular, we focus on the notion of separable kernels, see [1], and we introduce a new subtype for which error estimation via the power-function can be traced back to the power-functions of the scalar-valued kernels that were used to generate the matrix-valued kernel. For further details and different construction methods we refer to previous works, e.g. [3, 18, 6], in this field.

Definition 3.1 (Separable Kernels)**.** Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a matrix-valued kernel, let $Q_1, \dots, Q_p \in \mathbb{R}^{m\times m}$ be a collection of symmetric matrices and $k_1, \dots, k_p : \Omega \times \Omega \to \mathbb{R}$ a collection of scalar-valued kernels, such that

$$k(x,y) = \sum_{i=1}^{p} k_i(x,y)Q_i, \quad \text{for all } x, y \in \Omega. \tag{18}$$

We call $(k_i, Q_i)_{i=1}^p$ a decomposition of k and p its length. If p is minimal then the kernel k is called separable of order p.

To guarantee the (strict) positive definiteness of the kernel k further assumptions on the scalar-valued kernels k_i and symmetric matrices Q_i have to be made. Taking a closer look at the Gramian matrix $k(X,X)$ for some set $X = \{x_1, \dots, x_n\} \subset \Omega$ it is easy to see that the identity

$$k(X,X) = \sum_{i=1}^{p} k_i(X,X) \otimes Q_i \tag{19}$$

holds, where $k_i(X,X) \otimes Q_i$ denotes the Kronecker product. Since sums and Kronecker products of positive (semi-)definite matrices are positive (semi-)definite, we can conclude that the positive definiteness of k_i and positive (semi-)definiteness of Q_i is sufficient to guarantee that the kernel k is positive definite. In order to guarantee strict positive definiteness of k further assumptions on k_i and Q_i have to be made:

Lemma 3.1 (Separable kernel is s.p.d)**.** *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a separable kernel of order p with decomposition $(k_i, Q_i)_{i=1}^p$. If the kernels $k_1, \dots, k_p$ are s.p.d. and the matrices $Q_1, \dots, Q_p$ are positive semi-definite, such that $\sum_{i=1}^p Q_i$ is positive definite, then k is s.p.d.*

Proof. Let $X = \{x_1, \dots x_n\} \subset \Omega$ be a set of pairwise distinct points. Furthermore, let $K_i := k_i(X,X) \in \mathbb{R}^{n\times n}$ and $K := k(X,X) \in \mathbb{R}^{mn\times mn}$. It holds

$$K = \sum_{i=1}^{p} K_i \otimes Q_i.$$

Since the kernels k_i, $i = 1, \dots, p$ are s.p.d the matrices $K_1, \dots, K_p$ are positive definite which implies $K_i \succeq \lambda I_n$, where

$$\lambda = \min\left\{\lambda' \,|\, \lambda' \text{ is an eigenvalue of } K_i \text{ for some } i \in \{1, \dots, p\}\right\} > 0.$$

Therefore,

$$K = \sum_{i=1}^{p} K_i \otimes Q_i \succeq \sum_{i=1}^{p} \lambda I_n \otimes Q_i = \lambda I_n \otimes \left(\sum_{i=1}^{p} Q_i\right) \succ 0.$$

□

It is worthwhile to mention that the assumption $\sum_{i=1}^p Q_i \succ 0$ also guarantees that the kernel k is universal, c.f [26, 17], if the scalar-valued kernels k_i are universal. This means that for every compact subset $\Omega_c \subset \Omega$ the space

$$\operatorname{span}\{k(\cdot, x)\alpha \,|\, x \in \Omega_c,\, \alpha \in \mathbb{R}^m\}$$

is dense in the set of continuous function $C(\Omega_c)$ over Ω_c. For a proof we refer to [5].

In the above case of Definition 3.1 we call p minimal if any other decomposition of k has at least length p. This minimality can be directly related to the linear independency of the set of scalar-valued kernels $\{k_i\}_{i=1}^p$ and symmetric matrices $\{Q_i\}_{i=1}^p$:

Lemma 3.2 (Sufficient and necessary minimality condition)**.** *Let k be a separable kernel such that there exists a decomposition of length p. Then the following properties are equivalent*

i) p is minimal.

ii) For any decomposition $(k_i, Q_i)_{i=1}^p$ of length p the sets $\{k_1, \dots, k_p\}$ and $\{Q_1, \dots, Q_p\}$ are linearly independent, respectively.

Proof. "⇒" Let $(k_i, Q_i)_{i=1}^p$ be a decomposition of length p. Assume that either $\{k_1, \dots, k_p\}$ or $\{Q_1, \dots, Q_p\}$ is linearly dependent. Without loss of generality (w.l.o.g.) we can assume that either

$$k_1 = \sum_{i=2}^{p} \alpha_i k_i \quad \text{or} \quad Q_1 = \sum_{i=2}^{p} \beta_i Q_i.$$

Therefore,

$$k = \sum_{i=1}^{p} k_i Q_i = \sum_{i=2}^{p} k_i(Q_i + \alpha_i Q_1) \quad \text{or} \quad k = \sum_{i=2}^{p} (k_i + \beta_i k_1) Q_i.$$

In either case we found a smaller decomposition which contradicts the minimality of p since $Q_i + \alpha_i Q_1$ is still symmetric and $k_i + \beta_i k_1$ is still a matrix-valued kernel.
"$\Leftarrow$" Let $(k_i, Q_i)_{i=1}^p$ be a decomposition of length p such that $\{k_1, \dots, k_p\}$ and $\{Q_1, \dots, Q_p\}$ are linearly independent. Assume there exists a decomposition $(\hat{k}_i, \hat{Q}_i)_{i=1}^q$ of length $q < p$. Let $\mathrm{vec} : \mathbb{R}^{m\times m} \to \mathbb{R}^{m^2}$ denote the vectorization operator. We have

$$\sum_{i=1}^{p} k_i Q_i = k = \sum_{j=1}^{q} \hat{k}_j \hat{Q}_j$$

and thus

$$\sum_{i=1}^{p} \mathrm{vec}(Q_i) k_i = \sum_{j=1}^{q} \mathrm{vec}(\hat{Q}_j) \hat{k}_j.$$

Setting $Q := [\mathrm{vec}(Q_1) \dots \mathrm{vec}(Q_p)] \in \mathbb{R}^{m^2\times p}$, $\hat{Q} := [\mathrm{vec}(\hat{Q}_1) \dots \mathrm{vec}(\hat{Q}_q)] \in \mathbb{R}^{m^2\times q}$ we get

$$Q \begin{pmatrix} k_1 \\ \vdots \\ k_p \end{pmatrix} = \hat{Q} \begin{pmatrix} \hat{k}_1 \\ \vdots \\ \hat{k}_q \end{pmatrix}. \tag{20}$$

Since $\{Q_1, \dots, Q_p\}$ is linearly indendent it holds $\mathrm{rank}(Q) = p$ and, therefore, there exists a left inverse $A \in \mathbb{R}^{p\times m^2}$, i.e. $AQ = I_p$. Multiplying both sides in (20) with A from the left, we get

$$\begin{pmatrix} k_1 \\ \vdots \\ k_p \end{pmatrix} = A\hat{Q} \begin{pmatrix} \hat{k}_1 \\ \vdots \\ \hat{k}_q \end{pmatrix} = \hat{A} \begin{pmatrix} \hat{k}_1 \\ \vdots \\ \hat{k}_q \end{pmatrix}$$

with $\hat{A} := A\hat{Q} \in \mathbb{R}^{p\times q}$. Ultimately, we get $\mathrm{span}\{k_1, \dots, k_p\} \subset \mathrm{span}\{\hat{k}_1, \dots, \hat{k}_q\}$ which contradicts the linear independency of $\{k_1, \dots, k_p\}$. □

It is clear that k given by (18) is a matrix-valued kernel with regards to Definition 2.1, as

$$k(x,y)^T = \left(\sum_{i=1}^{p} k_i(x,y) Q_i\right)^T = \sum_{i=1}^{p} k_i(x,y) Q_i^T = \sum_{i=1}^{p} k_i(y,x) Q_i = k(y,x).$$

However, the minimality of p by no means implies the uniqueness of the decomposition in the sense that for decompositions $(k_i, Q_i)_{i=1}^p$ and $(\hat{k}_i, \hat{Q}_i)_{i=1}^p$ there exists a permutation ι of $\{1, \dots, p\}$ such that

$$k_i Q_i = \hat{k}_{\iota(i)} \hat{Q}_{\iota(i)}.$$

This is illustrated by the following example:

Example 3.1. Let $k_1, k_2 : \Omega \times \Omega \to \mathbb{R}$ denote two linearly independent scalar-valued kernels. Then $k : \Omega \times \Omega \to \mathbb{R}^2$ given by

$$k(x,y) := \begin{pmatrix} k_1(x,y) & 0 \\ 0 & k_1(x,y) + k_2(x,y) \end{pmatrix}$$

denotes a matrix-valued-kernel which has infinitely many minimal decompositions. Let $\lambda \in \mathbb{R}$, then

$$k(x,y) = k_1(x,y) Q_1(\lambda) + ((1-\lambda)k_1 + k_2)(x,y) Q_2,$$

where

$$Q_1(\lambda) = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

We note that there exists only one decomposition for which the spaces spanned by the columns of $Q_1(\lambda)$ and Q_2 have zero intersection. This leads us to the definition of a subclass of separable kernels:

Definition 3.2 (Uncoupled separable kernels)**.** Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a separable matrix-valued kernel and $(k_i, Q_i)_{i=1}^p$ be a decomposition. The decomposition is called uncoupled if

$$\mathrm{rank}\left(\sum_{i=1}^{p} Q_i\right) = \sum_{i=1}^{p} \mathrm{rank}(Q_i). \tag{21}$$

If there exists at least one uncoupled decomposition, the kernel is also called uncoupled.

Using the abbreviation $Q := \sum_{i=1}^{p} Q_i$, the rank condition (21) is equivalent to the assumption that the range $R(Q) := \mathrm{span}\{Q\alpha | \alpha \in \mathbb{R}^m\}$ is equal to the direct sum of the ranges $R(Q_i)$ of the individual matrices. We will state this in the following Lemma:

Lemma 3.3. *Let $Q_1,\dots,Q_p \in \mathbb{R}^{m\times m}$ be symmetric matrices. Then the following statements are equivalent*

i) $\operatorname{rank}\left(\sum_{i=1}^{p} Q_i\right) = \sum_{i=1}^{p} \operatorname{rank}(Q_i)$

ii) $R\left(\sum_{i=1}^{p} Q_i\right) = \bigoplus_{i=1}^{p} R(Q_i)$.

Proof. "⇒" We first show that the sum is direct. It is sufficient to show that $R(Q_i)\cap R(Q_j) = \{0\}$ for $i \neq j$. W.l.o.g. we assume $i = 1$ and $j = 2$ and $R(Q_1)\cap R(Q_2) \neq \{0\}$. It follows $\dim(R(Q_1)\cap R(Q_1)) \geq 1$ and thus

$$\begin{aligned}\operatorname{rank}\left(\sum_{i=1}^{p} Q_i\right) &= \dim R\left(\sum_{i=1}^{p} Q_i\right)\\ &\leq \dim R(Q_1) + \dim R(Q_2) - \dim(R(Q_1)\cap R(Q_2)) + \dim R\left(\sum_{i=3}^{p} Q_i\right)\\ &< \sum_{i=1}^{p} \dim R(Q_i) = \sum_{i=1}^{p} \operatorname{rank}(Q_i),\end{aligned}$$

which contradicts i). It is obvious that $R\left(\sum_{i=1}^{p} Q_i\right) \subset \bigoplus_{i=1}^{p} R(Q_i)$ and by i) the vector spaces have the same dimension and are therefore equal.
"⇐" Since the sum is direct it holds

$$\operatorname{rank}\left(\sum_{i=1}^{p} Q_i\right) = \dim R\left(\sum_{i=1}^{p} Q_i\right) = \dim\left(\bigoplus_{i=1}^{p} R(Q_i)\right) = \sum_{i=1}^{p} \dim R(Q_i) = \sum_{i=1}^{p} \operatorname{rank}(Q_i).$$

□

With the notion of uncoupledness we can now impose a sufficient condition for the uniqueness of a minimal decomposition up to permutations and scalings:

Theorem 3.4 (Uniqueness of uncoupled decompositions)**.** *Let k be a separable matrix-valued kernel with uncoupled decomposition $(k_i, Q_i)_{i=1}^{p}$. If p is minimal, then the decomposition is unique, up to permutations and scalings.*

Proof. Since the decomposition is uncoupled we have $R(Q_1)\cap R(Q_2+\cdots+Q_p) = \{0\}$ and $R(Q_1+Q_2+\cdots+Q_p) = R(Q_1)+R(Q_2+\cdots+Q_p)$. Therefore, there exists a $c \in \mathbb{R}^m$ such that $Q_1 c \neq 0$ and $Q_2 c,\dots,Q_p c = 0$. We get

$$\begin{aligned}k_1 \underbrace{Q_1 c}_{\neq 0} &= k_1 Q_1 c + k_2 Q_2 c + \cdots + k_p Q_p c\\ &= \hat{k}_1 \hat{Q}_1 c + \dots \hat{k}_p \hat{Q}_p c.\end{aligned}$$

Thus, k_1 can be written as a linear combination of $\hat{k}_1,\dots,\hat{k}_p$ such that

$$k_1 = \sum_{i=1}^{p} \hat{k}_i a_{1i}.$$

Similarly, the same holds for $k_2,\dots,k_p$ and therefore there exists a matrix $A = (a_{ij})_{i,j=1}^{p}$ such that

$$\begin{pmatrix} k_1 \\ \vdots \\ k_p \end{pmatrix} = A \begin{pmatrix} \hat{k}_1 \\ \vdots \\ \hat{k}_p \end{pmatrix}.$$

Furthermore, it holds for $i = 1,\dots,p$:

$$\hat{Q}_j = \sum_{i=1}^{p} a_{i,j} Q_i$$

and therefore

$$R(\hat{Q}_j) = \bigoplus_{i=1}^{p} a_{i,j} R(Q_i).$$

Since $(\hat{k}_j, \hat{Q}_j)_{j=1}^{p}$ is uncoupled it holds for $j \neq j'$:

$$R(\hat{Q}_j)\cap R(\hat{Q}_{j'}) = \{0\},$$

from which we conclude that $a_{i,j}$ or $a_{i,j'}$ is equal to 0. Thus, for every i there is exactly one $j = j(i)$ such that $a_{i,j(i)} \neq 0$ and the mapping $i \mapsto j(i)$ is bijective and it holds

$$k_i = \sum_{j=1}^{p} a_{i,j}\hat{k}_j = a_{i,j(i)}\hat{k}_{j(i)}.$$

Since both $(k_i, Q_i)_{i=1}^p$ and $(\hat{k}_j, \hat{Q}_j)_{j=1}^p$ are decompositions of k we get

$$\begin{aligned} 0 = k - k &= \sum_{i=1}^{p} k_i Q_i - \sum_{j=1}^{p} \hat{k}_j \hat{Q}_j \\ &= \sum_{i=1}^{p} k_i Q_i - \sum_{i=1}^{p} \hat{k}_{j(i)} \hat{Q}_{j(i)} \\ &= \sum_{i=1}^{p} \hat{k}_{j(i)} (a_{i,j(i)} Q_i - \hat{Q}_{j(i)}). \end{aligned}$$

Since the kernels $k_1, \dots, k_p$ and $\hat{k}_1, \dots, \hat{k}_p$ are linearly independent by Lemma 3.2, respectively, we conclude that

$$a_{i,j(i)} Q_i = \hat{Q}_{j(i)}$$

which results in

$$k_i Q_i = \hat{k}_{j(i)} a_{i,j(i)} Q_i = \hat{k}_{j(i)} \hat{Q}_{j(i)}.$$

□

In general the existence of an uncoupled or even minimal uncoupled decomposition cannot be guaranteed, as (21) necessitates that the length of any uncoupled decomposition is at most m. Therefore, any separable kernel of order $m+1$ possesses no uncoupled decomposition. In the following we want to present a sufficient criterion for the existence of a minimal uncoupled decomposition. This is motivated by trying to extend the well known fact for scalar-valued kernels that the product of two positive definite kernels is again a positive definite kernel, see [24]. This result does not extend to the matrix-valued case, since the kernels additionally have to commute for every pair of input parameters, i.e. for $k_1, k_2 : \Omega \times \Omega \to \mathbb{R}^{m \times m}$ it must hold

$$k_1(x,y)k_2(x,y) = k_2(x,y)k_1(x,y), \quad \forall\, (x,y) \in \Omega \times \Omega \tag{22}$$

to have that $k := k_1 \cdot k_2$ is a matrix-valued kernel. However, even if (22) is satisfied and both k_1, k_2 are positive definite the kernel k can be indefinite, as the following example shows:

Example 3.2. Let $k_1, k_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by

$$k_1(x,y) := e^{-\frac{1}{10}(x-y)^2} \quad \text{and} \quad k_2(x,y) := e^{-(x-y)^2}$$

and let $Q_1, Q_2 \in \mathbb{R}^{2\times 2}$ be the symmetric matrices

$$Q_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Furthermore, let $k : \Omega \times \Omega \to \mathbb{R}^{2\times 2}$ denote the matrix-valued kernel with decomposition $(k_i, Q_i)_{i=1}^2$ and $X = \{0, 1\}$. By Lemma 3.1 k is a positive definite kernel, but k^2 is not, as

$$k^2(X,X) = \begin{pmatrix} 5 & 3 & 2e^{-\frac{1}{5}} + 2e^{-\frac{11}{10}} + e^{-2} & 2e^{-\frac{1}{5}} + e^{-\frac{11}{10}} \\ 3 & 2 & 2e^{-\frac{1}{5}} + e^{-\frac{11}{10}} & 2e^{-\frac{1}{5}} \\ 2e^{-\frac{1}{5}} + 2e^{-\frac{11}{10}} + e^{-2} & 2e^{-\frac{1}{5}} + e^{-\frac{11}{10}} & 5 & 3 \\ 2e^{-\frac{1}{5}} + e^{-\frac{11}{10}} & 2e^{-\frac{1}{5}} & 3 & 2 \end{pmatrix}$$

has a negative eigenvalue $\lambda \approx -0.044$.

Taking a closer look, the matrix $k^2(X,X)$ can be written as a block-Hadamard product

$$k^2(X,X) = k(X,X) \square k(X,X) := (k(x_i,x_j)k(x_i,x_j))_{i,j}.$$

As it was shown in [10], the block-Hadamard product of two positive (semi-)definite block matrices $A = (A_{ij})_{i,j}$, $B = (B_{ij})_{i,j}$ is positive (semi-)definite if each block of A commutes with each block of B. If this restriction is applied to every possible Gramian matrix of a matrix-valued kernel, this leads to the condition

$$k(x,y)k(\tilde{x},\tilde{y}) = k(\tilde{x},\tilde{y})k(x,y), \quad \forall\, x,y,\tilde{x},\tilde{y} \in \Omega.$$

In this case, the kernel k can be characterized as follows:

Theorem 3.5. *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be matrix-valued kernel such that $k(x,y) = k(y,x)$ for all $x, y \in \Omega$. Then the following statements are equivalent*

i) $k(x,y)k(\tilde{x},\tilde{y}) = k(\tilde{x},\tilde{y})k(x,y)$ *for all* $x, \tilde{x}, y, \tilde{y} \in \Omega$

ii) There exists an orthogonal matrix $P \in \mathbb{R}^{m\times m}$ such that $P^T k(x,y)P$ is diagonal for all $x, y \in \Omega$.

iii) k is separable and there exists an uncoupled decomposition $(k_i, Q_i)_{i=1}^p$ with length $p \leq m$ and for which $Q_i Q_j = 0$ for $i \neq j$.

Proof. "$i) \Rightarrow ii)$" Let $A_1, \dots, A_d$ denote a basis of span $\{k(x,y) | x, y \in \Omega\}$. Then the A_i are symmetric, commute with one other and therefore are simultaneously diagonalizable, i.e. there exists an orthogonal matrix P such that $P^T A_i P$ is diagonal for $i = 1, \dots, m$. It follows, that $k(x,y) \in \text{span}\{A_1, \dots A_d\}$ is diagonalizable for any $x, y \in \Omega$
"$ii) \Rightarrow iii)$" By assumption it holds

$$P^T k(x,y) P = \text{diag}(k_1(x,y), \dots, k_d(x,y))$$

and $k_i : \Omega \times \Omega \to \mathbb{R}$, $i = 1, \dots, d$ are scalar-valued kernels. For $i = 1, \dots, d$ let $J(i) := \left\{ j : k_i = \alpha_{i,j} k_j \text{ for some } \alpha_{i,j} \in \mathbb{R} \right\}$. Then there exist $i_1, \dots, i_p$ with minimal p such that

$$\bigcup_{l=1}^{p} J(i_l) = \{1, \dots, d\} \quad \text{and} \quad J(i) \cap J(i') = \emptyset \text{ for } i \neq i'.$$

It holds

$$k = \sum_{i=1}^{m} k_i (Pe_i)(Pe_i)^T = \sum_{l=1}^{m} k_{i_l} \underbrace{\sum_{j \in J(i_l)} \alpha_{i_l,j} (Pe_j)(Pe_j)^T}_{=:Q_{i_l}} = \sum_{l=1}^{m} k_{i_l} Q_{i_l}.$$

Furthermore,

$$Q_i Q_{i'} = \sum_{j \in J(i)} \sum_{j' \in J(i')} \alpha_{i,j} \alpha_{i',j'} (Pe_j) \underbrace{(Pe_j)^T (Pe_{j'})}_{=0} (Pe_{j'})^T = 0$$

"$iii) \Rightarrow i)$" It holds

$$\begin{aligned}
k(x,y)k(\tilde{x},\tilde{y}) &= \left(\sum_{i=1}^{p} k_i(x,y) Q_i \right) \left(\sum_{j=1}^{p} k_j(\tilde{x},\tilde{y}) Q_j \right) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} k_i(x,y) k_j(\tilde{x},\tilde{y}) Q_i Q_j = \sum_{i=1}^{p} k_i(x,y) k_i(\tilde{x},\tilde{y}) Q_i^2 \\
&= \sum_{i=1}^{p} k_i(\tilde{x},\tilde{y}) k_i(x,y) Q_i^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} k_j(\tilde{x},\tilde{y}) k_i(x,y) Q_j Q_i \\
&= \left(\sum_{j=1}^{p} k_j(\tilde{x},\tilde{y}) Q_j \right) \left(\sum_{i=1}^{p} k_i(x,y) Q_i \right) \\
&= k(\tilde{x},\tilde{y}) k(x,y).
\end{aligned}$$

□

We conclude this subsection with a direct corollary:

Corollary 3.6. *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a positive definite matrix-valued kernel that satisfies $k(x,y) = k(y,x)$ for all $x, y \in \Omega$. If one of the conditions in Theorem 3.5 is met, then k^n is a positive definite matrix-valued kernel for any $n \in \mathbb{N}_0$.*

Proof. By Theorem 3.5 k can be decomposed as

$$k(x,y) = \sum_{l=1}^{p} k_l Q_l$$

with positive-definite scalar-valued kernels k_l and positive semi-definite matrices Q_l satisfying $Q_l Q_{l'} = 0$ for $l \neq l'$. Therefore, for any set $X = \{x_1, \dots, x_n\}$ of p.w. distinct points

$$k^n(X,X) = \sum_{l=1}^{p} \underbrace{\underbrace{k_l^n(X,X)}_{\succ 0} \otimes \underbrace{Q_l}_{\succeq 0}}_{\succeq 0} \succeq 0.$$

□

3.1 RKHS for separable kernels

As we want to consider approximations in the RKHS of separable kernels, we will show how the RKHS of the matrix-valued kernel k relates to the RKHS of the scalar-valued kernels k_i and matrices Q_i which form a decomposition of k. We start with decompositions of order 1:

Lemma 3.7 (RKHS of separable kernels of order 1). *Let k_s be a scalar-valued p.d. kernel and $Q \in \mathbb{R}^{m\times m}$ a positive semi-definite matrix. Then $k := k_s \cdot Q$ is a p.d. matrix-valued kernel and it holds*

$$\mathcal{H}_k = \mathcal{H}_{k_s} e_1 \oplus \cdots \oplus \mathcal{H}_{k_s} e_p. \tag{23}$$

Here $\{e_i\}_{i=1}^p$ denotes a basis of the range of Q.

Proof. We first show that the sum is direct. Let $f_i \in \mathcal{H}_{k_s} e_i$, $i = 1,..,p$. Assume that

$$f_1 + \cdots + f_p = 0$$

and there is at least one $j \in \{1,\dots,p\}$ such that $f_j \neq 0$. It follows

$$\operatorname{span}\{e_j\} \ni f_j(x) = -\sum_{i=1, i\neq j}^{p} f_i(x) \in \operatorname{span}\{e_1,\dots,e_{j-1},e_{j+1},\dots,e_p\}, \quad \forall\, x \in \Omega$$

and therefore

$$f_j(x) \in \operatorname{span}\{e_j\} \cap \operatorname{span}\{e_1,\dots,e_{j-1},e_{j+1},\dots,e_p\} = \{0\}, \quad \forall\, x \in \Omega,$$

i.e. $f_j = 0$. Iteratively we get $f_i = 0$ for $i = 1,\dots,p$ and the sum is direct. We now show that the right hand side of (23) is a subspace of the left hand side. Therefore, let $f_i \in \mathcal{H}_{k_s} e_i$, $i = 1,..,p$. Then there exist sequences $(\alpha_n^{(i)})_{n\in\mathbb{N}} \subset \mathbb{R}$ and $(x_n^{(i)})_{n\in\mathbb{N}} \subset \Omega$ such that

$$\begin{aligned} f_i &= \left(\sum_{n=1}^{\infty} k_s(\cdot, x_n^{(i)})\alpha_n^{(i)}\right) e_i \\ &= \sum_{n=1}^{\infty} k(\cdot, x_n^{(i)})\alpha_n^{(i)} v_i, \end{aligned}$$

where $v_i \in \mathbb{R}^m$ satisfies $Qv_i = e_i$. We conclude that $f_i \in \mathcal{H}_k$ for $i = 1,\dots,p$ and thus

$$f_1 + \cdots + f_p \in \mathcal{H}_k.$$

Assume that $\mathcal{H}_k \neq \mathcal{H}_{k_s} e_1 \oplus \cdots \oplus \mathcal{H}_{k_s} e_p$. Then for any $f \in \left(\mathcal{H}_{k_s} e_1 \oplus \cdots \oplus \mathcal{H}_{k_s} e_p.\right)^{\perp}$ it holds

$$\langle f, k_s(\cdot, x) e_i \rangle_{\mathcal{H}} = 0 \quad \forall\, x \in \Omega, i = 1,\dots,p.$$

Due to the linearity of the inner products it also holds

$$f(x)^T \alpha = \langle f, k(\cdot, x)\alpha \rangle_{\mathcal{H}} = 0 \quad \forall\, x \in \Omega$$

and thus $f = 0$. □

Remark 1. In the special case of $Q = I_m$, which where for example considered in [11, 29] this leads to the RKHS $\mathcal{H}_k = \bigotimes_{i=1}^{m} \mathcal{H}_{k_s}$ with the inner product given by

$$\langle f, g \rangle_{\mathcal{H}_k} = \langle (f_1,\dots,f_m), (g_1,\dots,g_m) \rangle = \sum_{i=1}^{m} \langle f_i, g_i \rangle_{\mathcal{H}_{k_s}}.$$

We have seen, c.f. Corollary 2.7 that the power-function is a valuable tool to provide error estimators to the pointwise error between a function f in $\mathcal{H}$ and its interpolant in a subspace $\mathcal{N}$. For scalar-valued kernels bounds on the decay of the power-functions are known for a wide variety of kernels, see [28] for more details. We want to make use of these bounds, to derive similar bound for the matrix-valued case. Again, we restrict ourself to the separable kernels of order 1 at first:

Lemma 3.8 (Power-function of separable kernels of order 1). *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a separable kernel of order 1 with decomposition (k_1, Q_1), where k_1 is a p.d. kernel and Q_1 is positive semi-definite. Let $X_n = \{x_1,\dots,x_n\} \subset \Omega$ be a set of pairwise distinct points. Furthermore, let $\mathcal{N} := \mathcal{N}_k(X_n)$, $\hat{\mathcal{N}} := \operatorname{span}\{k_1(\cdot, x_j) | \, x_j \in X_n\}$ and let $\mathcal{P}_{\hat{\mathcal{N}}}$ denote the power function of the scalar-valued kernel k_1. Then it holds*

$$\left(\mathcal{P}_{\mathcal{N}}^{\alpha}(x)\right)^2 = \mathcal{P}_{\hat{\mathcal{N}}}(x)^2 \alpha^T Q_1 \alpha. \tag{24}$$

Proof. Since $k = k_1 Q_1$ and due to Corollary 2.6 it is sufficient to show that $k_{\mathcal{N}} = k_{1,\hat{\mathcal{N}}} Q_1$. Let $K_1 := k_1(X_n, X_n)$. It is easy to see that $k(x, X_n) = k_1(x, X_n) \otimes Q_1 \in \mathbb{R}^{mn\times m}$ and $K = k(X_n, X_n) = K_1 \otimes Q_1$ and therefore by applying Corollary 2.4 we get

$$\begin{aligned} k_{\mathcal{N}}(x,y) &= k(x,X_n)^T K^+ k(y,X_n) \\ &= (k_1(x,X_n) \otimes Q_1)^T (K_1 \otimes Q_1)^+ k_1(y,X_n) \otimes (Q) \\ &= \left(k_1(x,X_n)^T K_1^+ k_1(y,X_n)\right) \otimes \left(Q_1 Q_1^+ Q_1\right) \\ &= k_{1,\hat{\mathcal{N}}}(x,y) \otimes Q_1 = k_{1,\hat{\mathcal{N}}}(x,y) Q_1. \end{aligned}$$

□

We now extend this result to separable kernels of higher order. It is easy to see that for k with decomposition $(k_i, Q_i)_{i=1}^p$ it holds

$$\mathcal{H} = \mathcal{H}_1 + \cdots + \mathcal{H}_p \tag{25}$$

where $\mathcal{H}_i$ denotes the RKHS of the separable kernel $k_i Q_i$ of order 1. By Lemma 3.7 we know that $\mathcal{H}_i$ can be written as a direct sum. However, in (25) the sum does no longer need to be direct which causes issues when trying to determine the power-function of k in terms of the power-function of the kernels k_i. This can be traced back to the fact that for a set $X = \{x_1, \dots, x_n\} \subset \Omega$ the space spanned by the functions $k(\cdot, x)\alpha$ for $x \in X$ and $\alpha \in \mathbb{R}^m$ is not equal to the sum of the individual subspaces spanned by $k_i(\cdot, x)Q_i\alpha$.

Lemma 3.9 (Power-function bound of separable kernel of order p). *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a separable matrix-valued kernel with decomposition $(k_i, Q_i)_{i=1}^p$ and $X = \{x_1, \dots, x_n\} \subset \Omega$. Furthermore, let $\hat{k}_i := k_i Q_i$ with $\mathcal{H}_i$ as its respective RKHS and*

$$\mathcal{N}_i := \operatorname{span}\left\{\hat{k}_i(\cdot, x)\alpha \,|\, x \in X,\, \alpha \in \mathbb{R}^m\right\}, \quad i = 1, \dots N,$$
$$\mathcal{N} := \operatorname{span}\left\{k(\cdot, x)\alpha \,|\, x \in X,\, \alpha \in \mathbb{R}^m\right\}$$

Then it holds for all $x \in \Omega$ and $\alpha \in \mathbb{R}^m$:

$$\sum_{i=1}^p \left(\mathcal{P}^\alpha_{\mathcal{N}_i}(x)\right)^2 \le \mathcal{P}^\alpha_{\mathcal{N}}(x)^2. \tag{26}$$

Proof. $\mathcal{N}_i \subset \mathcal{H}_i$ is a closed subspace with reproducing kernel $\hat{k}_{i,\mathcal{N}_i}$ and by Corollary 2.6 it holds

$$\mathcal{P}^\alpha_{\mathcal{N}_i}(x)^2 = \alpha^T \left(\hat{k}_i(x, x) - \hat{k}_{i,\mathcal{N}_i}(x, x)\right)\alpha. \tag{27}$$

We make use of the fact that the sum $\mathcal{M} := \mathcal{N}_1 + \cdots + \mathcal{N}_p$ is an RKHS with reproducing kernel $k_{\mathcal{M}} = \hat{k}_{1,\mathcal{N}_1} + \cdots + \hat{k}_{p,\mathcal{N}_p}$ and norm given by

$$\|f\|_{\mathcal{M}} = \min\left\{\sum_{i=1}^p \|f_i\|^2_{\mathcal{N}_i} \,|\, f = \sum_{i=1}^p f_i,\, f_i \in \mathcal{N}_i\right\}.$$

A proof for this assertion for the scalar-valued case can be found in [2]. The proof for the matrix-valued case only involves minor modifications. For the sake of completeness it is shown in the appendix.

It now holds

$$\begin{aligned}\|f\|_{\mathcal{M}} &= \min\left\{\sum_{i=1}^p \|f_i\|^2_{\mathcal{N}_i} \,|\, f = \sum_{i=1}^p f_i,\, f_i \in \mathcal{N}_i\right\}\\ &= \min\left\{\sum_{i=1}^p \|f_i\|^2_{\mathcal{H}_i} \,|\, f = \sum_{i=1}^p f_i,\, f_i \in \mathcal{N}_i\right\}\\ &= \|f\|_{\mathcal{H}}\end{aligned}$$

and therefore $k_{\mathcal{M}}$ is the reproducing kernel of the subspace $\mathcal{M} \subset \mathcal{H}$. Using (27) and Corollary 2.6 we conclude that

$$\begin{aligned}\mathcal{P}^\alpha_{\mathcal{M}}(x)^2 &= \alpha^T\left(\sum_{i=1}^p \hat{k}_i(x, x) - \sum_{i=1}^p \hat{k}_{i,\mathcal{N}_i}(x, x)\right)\alpha\\ &= \sum_{i=1}^p \alpha^T\left(\hat{k}_i(x, x) - \hat{k}_{i,\mathcal{N}_i}(x, x)\right)\alpha\\ &= \sum_{i=1}^p \mathcal{P}^\alpha_{\mathcal{N}_i}(x)^2.\end{aligned} \tag{28}$$

Since $\mathcal{N} \subset \mathcal{M}$ is a subspace the orthogonal complements satisfy $\mathcal{M}^\perp \subset \mathcal{N}^\perp$ and by applying Corollary 2.5 it follows

$$\sum_{i=1}^p \mathcal{P}^\alpha_{\mathcal{N}_i}(x)^2 = \mathcal{P}^\alpha_{\mathcal{M}}(x)^2 = \|\Pi_{\mathcal{M}^\perp} k(\cdot, x)\alpha\|_{\mathcal{H}} \le \|\Pi_{\mathcal{N}^\perp} k(\cdot, x)\alpha\|_{\mathcal{H}} = \mathcal{P}^\alpha_{\mathcal{N}}(x)^2.$$

□

We see that in general equality cannot be guaranteed. It only holds if the space $\mathcal{M}$ is equal to $\mathcal{N}$. This is equivalent to the fact that all $\hat{k}_i(\cdot, x)\alpha$ with $x \in X$ lie in $\mathcal{N}$. We will see in the following that this can be achieved when the decomposition is uncoupled:

Lemma 3.10 (Power-function of uncoupled separable kernels of order p). *Let $k : \Omega \times \Omega \to \mathbb{R}^{m\times m}$ be a separable matrix-valued kernel with uncoupled decomposition $(k_i, Q_i)_{i=1}^p$ and $X = \{x_1, \dots, x_n\} \subset \Omega$. Furthermore, let $\hat{k}_i := k_i Q_i$ with $\mathcal{H}_i$ as its respective RKHS and*

$$\mathcal{N}_i := \operatorname{span}\left\{\hat{k}_i(\cdot, x)\alpha \,|\, x \in X,\, \alpha \in \mathbb{R}^m\right\}, \quad i = 1, \dots N,$$
$$\mathcal{N} := \operatorname{span}\left\{k(\cdot, x)\alpha \,|\, x \in X,\, \alpha \in \mathbb{R}^m\right\}$$

Then it holds for all $x \in \Omega$ and $\alpha \in \mathbb{R}^m$:

$$\sum_{i=1}^p \left(\mathcal{P}^\alpha_{\mathcal{N}_i}(x)\right)^2 = \mathcal{P}^\alpha_{\mathcal{N}}(x)^2. \tag{29}$$

Proof. As mentioned before, it is sufficient to show that $\hat{k}_i(\cdot,x)\alpha \in \mathcal{N}$ for all $x \in X$, $\alpha \in \mathbb{R}^m$ and $i = 1,\dots,p$. Because the decomposition is uncoupled it holds with Lemma 3.3 that

$$R\left(\sum_{i=1}^{p} Q_i\right) = \bigoplus_{i=1}^{p} R(Q_i).$$

Therefore, for every $\alpha \in \mathbb{R}^m$ there exists a $\beta \in \mathbb{R}^m$ such that $Q_i\alpha = (Q_1 + \cdots + Q_p)\beta$. Since the sum is direct it holds that $Q_j\beta = 0$ for $j \neq i$ and therefore

$$\hat{k}_i(\cdot,x)\alpha = k_i(\cdot,x)Q_i\alpha = \sum_{i=1}^{p} k_i(\cdot,x)Q_i\beta = k(\cdot,x)\beta \in \mathcal{N}.$$

□

Lastly, we want to remark that while a lower bound in terms of the sum of the power-functions for the matrix-valued kernels of order 1 can be achieved, as seen in Lemma 3.9, an upper bound of this kind is not available in general as the following example shows.

Example 3.3. Let $\Omega \subset \mathbb{R}^d$ and $k_1, k_2 : \Omega \times \Omega \to \mathbb{R}$ be the polynomial kernels given by

$$k_1(x,y) = x^T y \quad \text{and} \quad k_2(x,y) = (x^T y)^2,$$

respectively, then the RKHS $\mathcal{H}_1$ is equal to the space of multivariate polynomials of degree 1 and $\mathcal{H}_2$ to the space of multivariate polynomials of degree 2. In particular, $\dim(\mathcal{H}_1) = d$ and $\dim(\mathcal{H}_2) = d(d+1)/2$ and therefore by choosing $X = \{x_i\}_{i=1}^{d(d+1)/2}$ such that $\{k_2(\cdot,x_i)\}_{i=1}^{d(d+1)/2}$ is linearly independent, the power-functions $\mathcal{P}_{\mathcal{N}_1(X)}$ and $\mathcal{P}_{\mathcal{N}_2(X)}$ vanish. However, the RKHS for $k := k_1 + k_2$ is given by the space of multivariate polynomials of degree 1 or 2 for which $\dim(\mathcal{H}) = d(d+3)/2$ holds. Consequently, $\mathcal{N}(X) \neq \mathcal{H}$ and $\mathcal{P}_{\mathcal{N}(X)}$ does not vanish.

4 Numerical Examples

4.1 Example 1

We now investigate the approximation quality of interpolation with matrix-valued kernels compared to a scalar-valued, i.e. componentwise approach. For this, we consider the target function $f : \Omega := [-2,2] \to \mathbb{R}^3$ given by

$$f(x) := \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} e^{-2.5(x-0.5)^2} + e^{-2.0(x+0.5)^2} \\ e^{-3.5(x-0.7)^2} \\ 1 \end{pmatrix}$$

and the uncoupled separable kernels $k_1,\dots,k_4 : \Omega \times \Omega \to \mathbb{R}^{3\times 3}$ of order 1, 3, 2 and 3, respectively, given by

$$\begin{aligned} k_1(x,y) &:= e^{-\varepsilon_{11}(x-y)^2} I_3 \\ k_2(x,y) &:= e^{-\varepsilon_{21}(x-y)^2} e_1 e_1^T + e^{-\varepsilon_{22}(x-y)^2} e_2 e_2^T + e^{-\varepsilon_{23}(x-y)^2} e_3 e_3^T \\ k_3(x,y) &:= e^{-\varepsilon_{31}(x-y)^2} v_1 v_1^T + e^{-\varepsilon_{32}(x-y)^2} \left(v_2 v_2^T + v_3 v_3^T\right) \\ k_4(x,y) &:= e^{-\varepsilon_{41}(x-y)^2} v_1 v_1^T + e^{-\varepsilon_{42}(x-y)^2} v_2 v_2^T + e^{-\varepsilon_{43}(x-y)^2} v_3 v_3^T, \end{aligned}$$

with shape parameters $\varepsilon_{11},\dots,\varepsilon_{43} \in (0,\infty)$. Here e_i denotes the i-th standard basis vector of $\mathbb{R}^3$ and v_1, v_2, v_3 are an orthonormal basis of eigenvectors of the covariance matrix C of f, which is computed by taking 401 random evaluations of f and setting

$$C := \frac{1}{400} \sum_{i=1}^{401} (f_i - \mu)(f_i - \mu)^T,$$

where $\mu \in \mathbb{R}^3$ contains the componentwise mean given by

$$\mu = \frac{1}{400} \sum_{i=1}^{401} f_i.$$

The kernels k_1 and k_2 handle the data componentwise that is, for the kernel k_1 the same scalar-valued kernel is used for every component, while for k_2 each component is treated by a different scalar-valued kernel. However, for the kernels k_3 and k_4 this is not the case. The shape parameters are determined by minimizing the maximum pointwise interpolation error $e_{k_i}(x) := \|f(x) - s_{k_i}(x)\|_2$ evaluated on a validation set Ω_V of 40 randomly chosen points in Ω for 50 logarithmically equidistantly distributed parameters in $M := [0.1, 100]$, where s_{k_i} is the interpolant on the set of 35 equidistantly distributed centers $X := \left\{-2 + \frac{4}{34} i | i = 0,\dots,34\right\}$ belonging to the RKHS that corresponds to k_i. The resulting parameters are listed in Table 1.

Parameter	ε_{11}	ε_{21}	ε_{22}	ε_{23}	ε_{31}	ε_{32}	ε_{41}	ε_{42}	ε_{43}
Value	1.931	1.931	1.931	1.600	0.244	3.393	0.244	3.393	3.393

Table 1: Results of the parameter selection for the different kernels.

We note that for the kernels k_1 and k_2 the selected shape parameters only differ in the third component, where a smaller parameter and therefore wider Gaussian was choosen for k_2. For the kernels k_3 and k_4 the selected parameters result in the same matrix-valued kernel. This can be explained by the fact that the eigenvectors v_2 and v_3 of the covariance matrix C were a-priori grouped together based on the fact that their corresponding eigenvalues $\lambda_2 = 0.112$ and $\lambda_3 = 0.206$ are of similar magnitude. This is reasonable as the eigenvalues are precisely the standard deviation of the data along the directions v_2 and v_3 and therefore the same Gaussian might be used for both directions.

Using the above parameters we compute the maximum pointwise interpolation error $e_{k_i}(X_N)$ on a test set $\Omega_T \subset \Omega$ of 400 equidistantly distributed points for an increasing number of equidistant training centers, i.e. $X_N := \left\{-2 + \frac{4}{N-1} i | i = 0, \dots, N-1\right\}$. The results for $N = 1, \dots, 35$ are plotted in Figure 1.

We can see that for a small number of centers, the difference in the approximation quality between the scalar-valued and matrix-valued approach is negligible. However, as the number of centers N increases, the kernel $k_3 = k_4$ begins to outperform the componentwise kernel k_1 and k_2. On the one hand, this leads to a higher accuracy for a fixed number of centers, i.e. a difference of almost three orders of magnitude for $N = 21$. On the other hand, this allows for a smaller expansion size while maintaining the same order of accuracy and therefore leads to a sparser approximant. The improved approximation quality for the kernel $k_3 = k_4$ can be traced back to the incorporation of given data in the construction of the kernel.

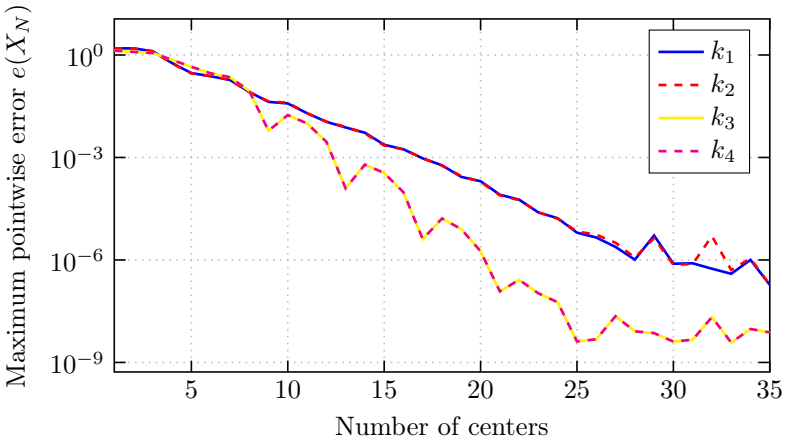

Figure 1: Maximum pointwise error measured in the Euclidean norm for the kernels k_1 to k_4 and for increasing number of centers N.

4.2 Example 2

We now want to verify the validity of the error bounds stated in Corollary 2.7. To this end we consider the domain $\Omega := [-1, 1]^2$ and the separable kernel k with decomposition $(k_1, Q_1)_{i=1}^3$ given by $k_i = e^{-i\|x-y\|^2}$ and

$$Q_1 = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}, Q_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \text{ and } Q_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

We consider the target function $f \in \mathcal{H}$ given by

$$f(x) = \sum_{i=1}^{5} k(x, y_i)\alpha_i,$$

where $y_1, \dots, y_5 \in \Omega$ and $\alpha_1, \dots, \alpha_5 \in \mathbb{R}^4$ were randomly chosen. We further select $X = \{x_1, \dots, x_{100}\}$ random points and for $X_i := \{x_1, \dots, x_i\}$ compute the error in the Euclidean-, infinity- and one-norm as well as the error bounds

$$\Delta_2^1 := \|k_{\mathcal{N}(X_i)^\perp}(x,x)\|_2 \|f - \Pi_{\mathcal{N}(X_i)} f\|, \qquad \Delta_2^2 := \|k_{\mathcal{N}(X_i)^\perp}(x,x)\|_2 \|f\|,$$
$$\Delta_\infty^1 := \max_{j=1,\dots,4} |k_{\mathcal{N}(X_i)^\perp}(x,x)_{jj}| \|f - \Pi_{\mathcal{N}(X_i)} f\|, \qquad \Delta_\infty^2 := \max_{j=1,\dots,4} |k_{\mathcal{N}(X_i)^\perp}(x,x)_{jj}| \|f\|,$$
$$\Delta_1^1 := \|k_{\mathcal{N}(X_i)^\perp}(x,x)\|_2 \|f - \Pi_{\mathcal{N}(X_i)} f\|, \qquad \Delta_1^2 := \|k_{\mathcal{N}(X_i)^\perp}(x,x)\|_2 \|f\|.$$

The results are plotted in Figures 2 - 4, respectively. We can see that the qualitative behaviour of both the error and the error bounds is similar for each norm. However, the second bounds, i.e. $\Delta_2^2, \Delta_\infty^2$ and Δ_1^2 overestimate the true error by a large margin compared to the first bounds $\Delta_2^1, \Delta_\infty^1$ and Δ_1^1 as no information on the residual $f - \Pi_{\mathcal{N}(X_i)}(f)$ is incorporated into the bounds.

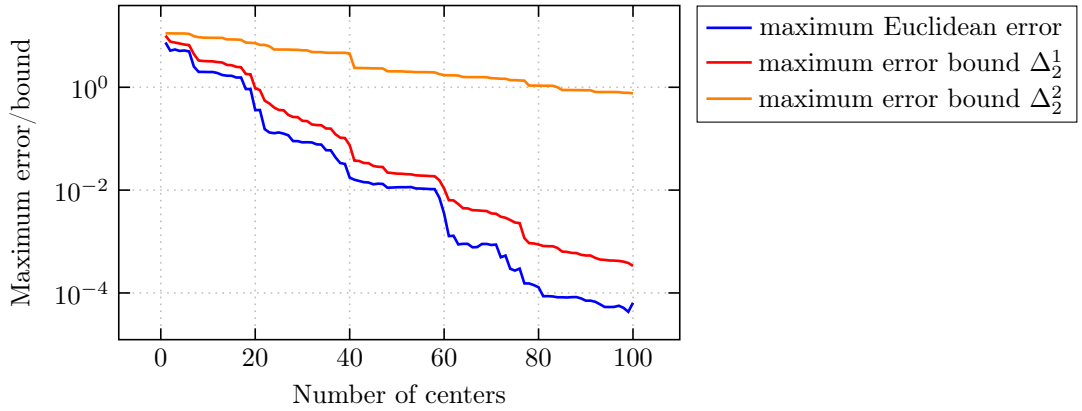

Figure 2: Maximum pointwise error measured in the Euclidean norm for increasing number of centers.

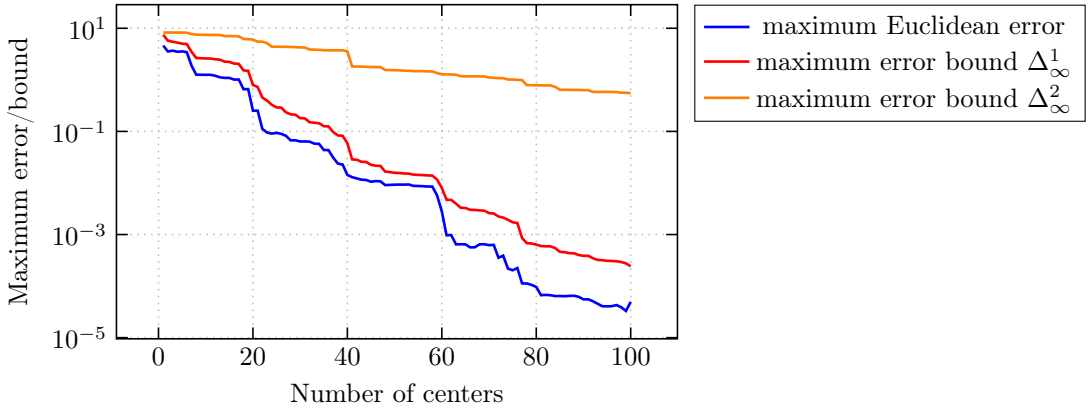

Figure 3: Maximum pointwise error measured in the infinity norm for increasing number of centers.

5 Conclusion

In this paper we recalled the concept of matrix-valued kernels and showed how they can be used to compute approximations or surrogate models for which a-priori error estimate in various norms are available by means of the power-function. Furthermore, we introduced a new subclass of separable matrix-valued kernels, for which the power-function can be traced back to the power-functions of scalar-valued kernels. In an artificial example for a low-dimensional output we illustrated how matrix-valued kernels can be used to encode (linear) correlations between function components which leads to a significant improvement in the quality of the approximation.

Future work will investigate the selection of suitable centers via Greedy algorithms, where we obtained initial results in [31].

Appendix

Theorem: *Let $\mathcal{H}_1, \dots, \mathcal{H}_p$ be RKHS with reproducing kernels $k_1, \dots, k_p$. Then $\mathcal{H} = \bigoplus_{i=1}^p \mathcal{H}_i$ is a RKHS with reproducing kernel $k = \sum_{i=1}^p k_i$ and norm given by*

$$\|f\|_{\mathcal{H}}^2 = \min\left\{ \sum_{i=1}^p \|f_i\|_{\mathcal{H}_i}^2 \mid f = \sum_{i=1}^p f_i, f_i \in \mathcal{H}_i \right\}$$

Proof. By the principle of induction it is sufficient to consider the case $p = 2$. Therefore, let $\mathcal{M} := \mathcal{H}_1 \times \mathcal{H}_2$. One easily verifies that $\mathcal{M}$ equipped with the inner product

$$\langle (f_1, f_2), (g_1, g_2) \rangle_{\mathcal{M}} = \langle f_1, g_1 \rangle_{\mathcal{H}_1} + \langle f_2, g_2 \rangle_{\mathcal{H}_2}$$

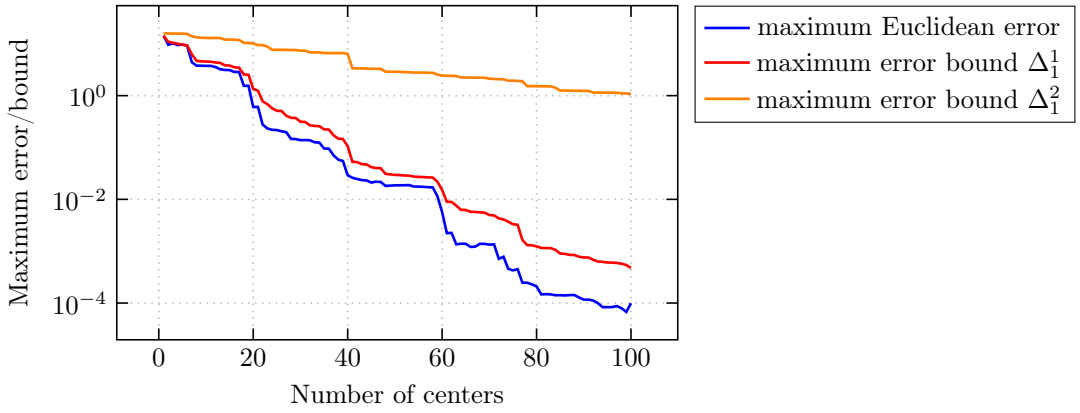

Figure 4: Maximum pointwise error measured in the one-norm for increasing number of centers.

is an RKHS with reproducing kernel $k_{\mathcal{M}} = (k_1, k_2)$. Furthermore, let $S : \mathcal{M} \to \mathcal{H}_1 + \mathcal{H}_2$ be given by

$$S(f_1, f_2) = f_1 + f_2$$

and denote $\mathcal{N} := S^{-1}(\{0\}) = \mathcal{H}_1 \cap \mathcal{H}_2$. □

Then $\mathcal{N}$ is a closed subspace and thus $\mathcal{M} = \mathcal{N} \oplus \mathcal{N}^\perp$. Therefore, $T := S|_{\mathcal{N}^\perp} : \mathcal{N}^\perp \to \mathcal{H}$ is a bijection and we equip $\mathcal{H}$ with the inner product

$$\langle f, g \rangle_{\mathcal{H}} = \langle T^{-1}(f), T^{-1}(g) \rangle_{\mathcal{M}}.$$

For any arbitrary $f \in \mathcal{H}$ we now have

$$S^{-1}(\{f\}) = T^{-1}(f) + \mathcal{N}$$

and therefore

$$\|f\|_{\mathcal{H}}^2 = \|T^{-1}(f)\|_{\mathcal{M}}^2 = \min\{\|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2 | f_1 + f_2 = f, f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2\}.T$$

It remains to show that $k = k_1 + k_2$ satisfies the reproducing property. By definition $k(\cdot, x)\alpha \in \mathcal{H}$ is clear. Let $f \in \mathcal{H}$ and let $(g_1, g_2) = T^{-1}(k(\cdot, x)\alpha)$. It now holds that $(g_1 - k_1(\cdot, x)\alpha, g_2 - k_2(\cdot, x)\alpha) \in \mathcal{N}$ and $(f_1, f_2) := T^{-1}(f) \in \mathcal{N}^\perp$. Therefore,

$$\begin{aligned}
\langle f, k(\cdot, x)\alpha \rangle_{\mathcal{H}} &= \langle T^{-1}(f), (g_1, g_2) \rangle_{\mathcal{M}} \\
&= \langle T^{-1}(f), (g_1 - k_1(\cdot, x)\alpha, g_2 - k_2(\cdot, x)\alpha) \rangle_{\mathcal{M}} \\
&\quad + \langle T^{-1}(f), (k_1(\cdot, x)\alpha, k_2(\cdot, x)\alpha) \rangle_{\mathcal{M}} \\
&= \langle (f_1, f_2), (k_1(\cdot, x)\alpha, k_2(\cdot, x)\alpha \rangle_{\mathcal{M}} \\
&= \langle f_1, k_1(\cdot, x)\alpha \rangle_{\mathcal{H}_1} + \langle f_2, k_2(\cdot, x)\alpha \rangle_{\mathcal{H}_2} \\
&= f_1(x)^T \alpha + f_2(x)^T \alpha \\
&= f(x)^T \alpha.
\end{aligned}$$

Therefore, k is the reproducing kernel of $\mathcal{H}$.

References

[1] M. Alvarez, L. Rosasco, and N. D. Lawrence. Kernels for vector-valued functions: a review. *Foundations and Trends in Machine Learning*, 4(3):195–266, 2012.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[3] R. K. Beatson, W. zu Castell, and S. J. Schrödl. Kernel-based methods for vector-valued data with correlated components. *SIAM Journal on Scientific Computing*, 33(4):1975–1995, 2011.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning*, volume 2. Springer, August 2006. ISBN 978-0-387-31073-2.

[5] A. Caponnetto, C. A. Micchelli, M. Pontil, and Y. Ying. Universal multi-task kernels. *Journal of Machine Learning Research*, 9:1615–1646, August 2008.

[6] C. Carmeli, E. De Vito, and A. Toigo. Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem. *Analysis and Applications*, 4(4):377–408, 2006.

[7] P. Farrell, K. Gillow, and H. Wendland. Multilevel interpolation of divergence-free vector fields. *IMA Journal of Numerical Analysis*, 37(1):332–353, 2017.

[8] R. Franke. *A critical comparison of some methods for interpolation of scattered data*, volume 253. Naval Postgraduate School Tech.Rep., Mar. 1979.

[9] E. J. Fuselier and G. B. Wright. Stability and error estimates for vector field interpolation and decomposition on the sphere with rbfs. *SIAM Journal on Numerical Analysis*, 47(5):3213–3239, 2009.

[10] M. Günther and L. Klotz. Schur's theorem for a block Hadamard product. *Linear Algebra and its Applications*, 437:948–956, 2012.

[11] B. Haasdonk and G. Santin. Greedy kernel approximation for sparse surrogate modeling. In W. Keiper, A. Milde, and S. Volkwein, editors, *Reduced-Order Modeling (ROM) for Simulation and Optimization: Powerful Algorithms as Key Enablers for Scientific Computing*, pages 21–45. Springer International Publishing, Cham, 2018.

[12] H. Kadri, E. Duflos, P. Preux., S. Canu, A. Rakotomamonjy, and J. Audiffren. Operator-valued kernels for learning from functional response data. *Journal of Machine Learning Research*, 17(20), 2016.

[13] S. Lowitzsch. Matrix-valued radial basis functions: stability estimates and applications. *Advances in Computational Mathematics*, 23(3):299–315, Oct 2005.

[14] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2:11–22, 1986.

[15] C. A. Micchelli and M. Pontil. Kernels for multi-task learning. *Advances in Neural Information Processing Systems*, 2004.

[16] C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17(1):177–204, 2005.

[17] C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *Journal of Machine Learning Research*, 7:2651–2667, June 2006.

[18] M. Michelo and J. A. Glaunes. Matrix-valued kernels for shape deformation analysis. *Geometry, Imaging and Computing*, 1(1):57–139, 2014.

[19] M. Mouattamid and R. Schaback. Recursive kernels. *Analysis in Theory and Applications*, 25(4):301–316, 2009.

[20] F. J. Narcowich and J. D. Ward. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Mathematics of Computation*, 63(208):661–687, 1994.

[21] M. Reisert and H. Burkhardt. Learning equivariant functions with matrix valued kernels. *Journal of Machine Learning Research*, 8:385–408, May 2007.

[22] R. Schaback. A comparison of radial basis function interpolants. In K. Jetter and F. Utreras, editors, *Multivariate approximation - from CAGD to wavelets*, pages 293–305. World Scientific, 1993.

[23] R. Schaback and H. Wendland. Kernel techniques: From machine learning to meshless methods. *Acta Numerica*, 15:543–639, May 2006.

[24] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2002.

[25] S. Schrödl. *Operator Valued Reproducing Kernels and Their Application in Approximation and Statistical Learning*. Berichte aus der Mathematik. Shaker, 2009.

[26] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 2:67–93, 2001.

[27] I. Steinwart, D. Hush, and C. Scovel. Training SVMs Without Offset. *Journal of Machine Learning Research*, 12:141–202, February 2011.

[28] H. Wendland. *Scattered Data Approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005.

[29] D. Wirtz and B. Haasdonk. A vectorial kernel orthogonal greedy algorithm. *Dolomites Research Notes on Approximation*, 6:83–100, 2013.

[30] D. Wirtz, N. Karajan, and B. Haasdonk. Surrogate modelling of multiscale models using kernel methods. *International Journal of Numerical Methods in Engineering*, 101(1):1–28, 2015.

[31] D. Wittwar and B. Haasdonk. Greedy Algorithms for Matrix-Valued Kernels. In *Numerical Mathematics and Advanced Applications - ENUMATH 2017*, 2018.