D. Wirtz and B. Haasdonk

A Vectorial Kernel Orthogonal Greedy Algorithm

Stuttgart, December 2012

Institute of Applied Analysis and Numerical Simulation,
University of Stuttgart,
Pfaffenwaldring 57
D-70569 Stuttgart, Germany
{daniel.wirtz,bernard.haasdonk}@mathematik.uni-stuttgart.de
www.agh.ians.uni-stuttgart.de

Abstract This work is concerned with derivation and analysis of a modified vectorial kernel orthogonal greedy algorithm (VKOGA) for approximation of nonlinear vectorial functions. The algorithm pursues simultaneous approximation of all vector components over a shared linear subspace of the underlying function Hilbert space in a greedy fashion [14, 33] and inherits the selection principle of the f/P-Greedy algorithm [18]. For the considered algorithm we perform a limit analysis of the selection criteria for already included subspace basis functions. We show that the approximation gain is bounded globally and for the multivariate case the limit functions correspond to a directional Hermite interpolation. We further prove algebraic convergence similar to [13], improved by a dimension-dependent factor, and introduce a new a-posteriori error bound. Comparison to related variants of our algorithm are presented. Targeted applications of this algorithm are model reduction of multiscale models [40].

Keywords vector approximation, greedy approximation, orthogonal pursuit, kernel methods, adaptive approximation, sparse approximation

Preprint Series
Stuttgart Research Centre for Simulation Technology (SRC SimTech)
SimTech – Cluster of Excellence
Pfaffenwaldring 7a
70569 Stuttgart
publications@simtech.uni-stuttgart.de
www.simtech.uni-stuttgart.de

Sparse approximation of nonlinear functions is a challenging task that arises in many different areas of modern computing. A key element to find sparse representations is the concept of m-term approximation [7, 28], which basically is a measure of how well a function from a given function space can be approximated by linearly combining m functions out of a given set from the same space. This set can either be a basis of the considered space or a redundant (dense) system of functions, where the latter is also called a *dictionary* and is considered in our work. However, direct computation of *the best* m-term approximation is not possible in practice, as the computation has combinatorical complexity dependent on the number of dictionary elements. Hence, the challenge is to find methods and algorithms that provide near-best m-term approximations. In this work, we will consider a type of approximation method that belongs to the family of *greedy algorithms*, which have already been proven to yield near-best m-term approximations under various conditions, see e.g. [6, 7, 33, 34]. Their "greedy" nature has its foundation in a *greedy step*, which determines the next dictionary element to be added to an existing m-term approximant according to certain maximizing criteria, mostly involving residuals. Well known criteria so far roughly distinguish between pure and weak greedy algorithms, where the true or almost true maximum approximation "gain" is considered, respectively. An extensive survey of greedy algorithms can be found in [33], however, greedy approximation methods appear in the literature in different facets like matching pursuit [4, 15, 19] or greedy pursuit [36]. So far, approximation and convergence results have been established for quite general spaces, e.g. Hilbert [7, 33] or Banach spaces [11, 13].

Here, we will consider a special kind of approximation Hilbert space, namely *reproducing kernel Hilbert spaces* (RKHS) induced by *kernels*, which we introduce in detail in Section 1. RKHS and kernel methods have been applied in many different contexts like pattern recognition [30], machine learning [29] or scattered data approximation [38]. We refer to [8] for a current review of approximation methods with positive definite kernels and [26] for a practical guide on kernel methods and their applications. RKHS have hence been successfully used in various contexts, and, seen as Hilbert spaces of functions, readily allow to apply the greedy approximation theory described above. It is also evident that the selection criteria for subsequent new dictionary elements depends on the way the m-term approximant in any current linear subspace is computed. The most natural approach is to use orthogonal projection with respect to the native RKHS scalar product, which guarantees the best possible approximation in each subspace. We shall regard this approach in our work, however, note that there are more choices e.g. using least squares [37] or orthogonal least squares [1, 3].

However, greedy algorithms in the context of RKHS have been already formulated [18, 25] and some results on convergence have been established. In this work, we will focus on a vectorial variant of orthogonal kernel greedy algorithms, more precisely an extension of the so-called f/P-Greedy algorithm from [18, 3.1.1] in the spirit of [13]. We will investigate the selection criteria more closely and show that an improved error bound and a-posteriori bounds can be obtained for the considered vectorial greedy algorithm. For related work on vectorial greedy algorithms see [12, 14]. A vectorial regression approach can be found in [31] or multioutput orthogonal least squares approximations are discussed in [2].

In our field of research we apply kernel approximation methods in the context of model reduction. Potential applications of this algorithm are projection-based model order reduction of nonlinear dynamical systems [21, 39] and model reduction of multiscale models, where the micro-scale model can often be replaced by approximation of the intput-output relation between the involved scales [40].

After establishing the necessary background regarding kernels and the induced Hilbert spaces in Section 1, Section 2 introduces our vectorial greedy algorithm. We shortly discuss computational aspects in Section 3 and present numerical illustrations in Section 4. We conclude with a summarizing discussion in Section 5.

1 Preliminaries

1.1 Kernels and Reproducing Kernel Hilbert Spaces

We start with introducing the basic definitions and concepts used throughout our article. We will indicate matrix- and vector-valued variables by bold upper- and lower-case latin letters and scalar values by normal typesetting. Let $\Omega \subset \mathbb{R}^d, d \in \mathbb{N}$ be a closed domain for the rest of this work.

Definition 1 (Kernels) A function $K : \Omega \times \Omega \to \mathbb{R}$ is called *positive definite kernel* if $\forall\ N \in \mathbb{N}$, $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \Omega$ and $\boldsymbol{\alpha} \in \mathbb{R}^N \backslash \{0\}$ we have

$$\sum_{i,j=1}^{N} \alpha_i \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0.$$

Definition 2 (RKHS) Let $\Omega \subset \mathbb{R}^d$, $K : \Omega \times \Omega \to \mathbb{R}$ be a symmetric, positive definite kernel and $X \subset \Omega$.
Then we denote by

$$\mathcal{H}^X := \langle \{K(\boldsymbol{x}, \cdot) \mid \boldsymbol{x} \in X\} \rangle$$

the $\mathbb{R}$-vector space spanned by all functions $K(\boldsymbol{x}, \cdot)$, where $\langle \cdot \rangle$ is a shorthand for the span operation. We equip $\mathcal{H}^X$ with the scalar product

$$\langle K(\boldsymbol{x}, \cdot), K(\boldsymbol{y}, \cdot) \rangle_{\mathcal{H}^X} := K(\boldsymbol{x}, \boldsymbol{y}), \quad \boldsymbol{x}, \boldsymbol{y} \in X,$$

which naturally extends to functions from $\mathcal{H}^X$. If $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ for $m \in \mathbb{N}$, $\mathcal{H}^X$ is an at most m-dimensional $\mathbb{R}$-Hilbert space spanned by K over X.

We further denote by

$$\mathcal{H} = \overline{\mathcal{H}^\Omega} = \overline{\langle \{K(\boldsymbol{x}, \cdot) \mid \boldsymbol{x} \in \Omega\} \rangle}$$

the Hilbert space induced by K over Ω. In fact, for each symmetric, positive definite K there is a unique such space with the reproducing property

$$\langle f, K(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}} = f(\boldsymbol{x}) \ \forall\ f \in \mathcal{H}, \boldsymbol{x} \in \Omega, \tag{1}$$

which is why those spaces are also known as reproducing kernel Hilbert spaces (RKHS).

For a more complete characterization of RKHS we refer to [38, §10], for example. For the remainder of this work, let K be a symmetric, positive definite and normalized ($K(\boldsymbol{x}, \boldsymbol{x}) = 1\ \forall\ \boldsymbol{x} \in \Omega$) kernel on Ω with induced Hilbert space $\mathcal{H}$ unless explicitly defined otherwise.

Definition 3 (Kernel matrix and vector) For $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subset \Omega$ we denote by

$$\boldsymbol{K}_X := \big(K(\boldsymbol{x}_i, \boldsymbol{x}_j)\big)_{ij}, \quad i, j = 1 \dots m,$$

the *kernel matrix* of K with respect to X. The positive definiteness of a kernel K is equivalent to positive semi-definiteness of the corresponding kernel matrix $\boldsymbol{K}_X$. Further we denote for $\boldsymbol{x} \in \Omega$ by

$$\boldsymbol{k}_X(\boldsymbol{x}) := \big(K(\boldsymbol{x}_1, \boldsymbol{x}), \dots, K(\boldsymbol{x}_N, \boldsymbol{x})\big)^T \in \mathbb{R}^m$$

the *kernel vector* of K at $\boldsymbol{x}$ with respect to X. For ease of reading, we will omit the subindices $\boldsymbol{K}_X, \boldsymbol{k}_X$ whenever it is clear from context.

The following Lemma shows how smoothness of a kernel inherits to the RKHS functions. The proof is along the lines of [38, §10.6] and/or [32, Lemma 4.34].

Lemma 1 (Derivatives of RKHS functions) *If $K \in C^2(\Omega \times \Omega)$, then $\frac{\partial K}{\partial x_i}(\boldsymbol{x}, \cdot) \in \mathcal{H}\ \forall\ \boldsymbol{x} \in \Omega, i = 1 \dots d$ and we have*

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \left\langle f, \frac{\partial K}{\partial x_i}(\boldsymbol{x}, \cdot) \right\rangle_{\mathcal{H}} \qquad \forall\ \boldsymbol{x} \in \Omega, i = 1 \dots d$$

1.2 Projection and orthogonal remainders

As mentioned at the beginning, we will focus on projection of functions into subspaces to obtain the approximants. Therefore we first introduce the projection operator for linear subspaces.

Definition 4 (Orthogonal projection operator) Let $S \subseteq \mathcal{H}$ be a linear subspace of $\mathcal{H}$. Then the orthogonal projection operator is denoted by

$$\begin{aligned} P_S : \mathcal{H} &\to S \\ f &\mapsto P_S[f], \end{aligned}$$

such that

$$\langle f - P_S[f], g \rangle_{\mathcal{H}} = 0 \ \forall \ g \in S. \tag{2}$$

Next we will show some frequently used properties of projections.

Lemma 2 *Let $S \subseteq \mathcal{H}$ be a linear subspace of $\mathcal{H}$. Then*

$$||\mathcal{P}_S[f]||^2_{\mathcal{H}} = \langle f, \mathcal{P}_S[f] \rangle_{\mathcal{H}} \ \forall \ f \in \mathcal{H}.$$

If further $S = \mathcal{H}^X$ for some $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$ so that $\boldsymbol{K}$ is non-singular, we have

$$\mathcal{P}_S[f] = \sum_{i=1}^{N} c_i K(\boldsymbol{x}_i, \cdot), \qquad c = \boldsymbol{K}^{-1}\boldsymbol{f}, \qquad f_i = f(\boldsymbol{x}_i), \ i = 1 \dots N. \tag{3}$$

Proof.

$$\begin{aligned} ||\mathcal{P}_S[f]||^2_{\mathcal{H}} &= \langle \mathcal{P}_S[f], \mathcal{P}_S[f] \rangle_{\mathcal{H}} = \langle f - (f - \mathcal{P}_S[f]), \mathcal{P}_S[f] \rangle_{\mathcal{H}} \\ &= \langle f, \mathcal{P}_S[f] \rangle_{\mathcal{H}} - \underbrace{\langle (f - \mathcal{P}_S[f]), \mathcal{P}_S[f] \rangle_{\mathcal{H}}}_{=0} = \langle f, \mathcal{P}_S[f] \rangle_{\mathcal{H}} \, . \end{aligned}$$

Equation (3) follows directly from the projection conditions (2) and the fact that $\mathcal{P}_S[f] \in \mathcal{H}^X$. □

Remark 1 In this analytically focused work, we consider direct kernel translates as dictionary elements. Lemma 2 shows that, under certain conditions, the projection actually means interpolation at specific function values. It is well known that the condition of the kernel matrix can get arbitrarily bad, even though the interpolation task itself is not unstable. Several approaches, e.g. [18, 20, 24], have been developed since to alleviate those problems by choosing a stable basis. Proper inclusion of those methods to formulate a stable basis of $\mathcal{H}^X$ is future work and outside the scope of this article.

Unfortunately, due to a different scope, there have been developing two different but closely related ways of notation in the context of greedy algorithms. Whilst the classical greedy theory [33] considers scalar products of function residuals and dictionary elements in the greedy step selection criteria, kernel greedy algorithms [18, 25] usually consider pointwise maxima $\max_x |f(x) - s_{f,X}(x)|$ in the greedy step, where $s_{f,X}$ is the interpolant of f on the current $m - th$ point set X. However, this connection will become clear using the concept of *orthogonal remainders*, which we will investigate in the following.

Definition 5 (Orthogonal/orthonormal remainders) Let $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subseteq \Omega$, $\boldsymbol{x} \in \Omega$ and define $\Omega_X := \{\boldsymbol{x} \in \Omega \mid K(\boldsymbol{x}, \cdot) \in \mathcal{H}^X\}$. Then we define the $\mathcal{H}^X$-orthogonal remainder $\tilde{\phi}_{\boldsymbol{x}}$ of $K(\boldsymbol{x}, \cdot)$ as

$$\tilde{\phi}_{\boldsymbol{x}} := K(\boldsymbol{x}, \cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)],$$

and for $\boldsymbol{x} \in \Omega \backslash \Omega_X$ the $\mathcal{H}^X$-orthonormal remainder

$$\phi_{\boldsymbol{x}} := \tilde{\phi}_{\boldsymbol{x}} \Big/ \left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|_{\mathcal{H}} .$$

The next Lemma shows some interesting properties of orthogonal/normal remainders.

Lemma 3 (Properties for remainders) *Let the conditions of Definition 5 hold and let $f \in \mathcal{H}$. Then*

$$\mathcal{P}_{\mathcal{H}^X \oplus \langle \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}}[f] = \mathcal{P}_{\mathcal{H}^X}[f] + \mathcal{P}_{\langle \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}}[f] \qquad \forall\ f \in \mathcal{H}. \tag{4}$$

$$\left\langle f, \tilde{\phi}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} = \langle f - \mathcal{P}_{\mathcal{H}^X}[f], K(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}} = f(\boldsymbol{x}) - \mathcal{P}_{\mathcal{H}^X}[f](\boldsymbol{x}) \qquad \forall\ f \in \mathcal{H} \tag{5}$$

Furthermore, $\tilde{\phi}_{\boldsymbol{x}}$ is the Riesz-representant of the linear Functional $\delta_{\boldsymbol{x}} - \mathcal{P}_{\mathcal{H}^X}[\cdot](\boldsymbol{x}) : \mathcal{H} \to \mathbb{R}$.

Proof. At first, if $K(\boldsymbol{x}, \cdot) \in \mathcal{H}^X$ then condition (4) holds trivially. So, assume $K(\boldsymbol{x}, \cdot) \notin \mathcal{H}^X$ and $\{\varphi_1, \ldots, \varphi_l\}$ to be an orthonormal basis (ONB) of $\mathcal{H}^X$. From this and the condition $\left\langle f - \mathcal{P}_{\langle \phi_{\boldsymbol{x}} \rangle}[f], \phi_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} = 0$ we get

$$\mathcal{P}_{\mathcal{H}^X}[f] = \sum_{i=1}^{l} \langle f, \varphi_i \rangle_{\mathcal{H}}\, \varphi_i,$$
$$\mathcal{P}_{\langle \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}}[f] = \langle f, \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}\, \phi_{\boldsymbol{x}}.$$

As $\phi_{\boldsymbol{x}}$ is orthogonal to $\mathcal{H}^X$ by definition we know that $\{\varphi_1, \ldots, \varphi_l, \phi_{\boldsymbol{x}}\}$ is ONB of $\mathcal{H}^X \oplus \langle \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}$ and directly obtain

$$\mathcal{P}_{\mathcal{H}^X \oplus \langle \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}}[f] = \sum_{i=1}^{l} \langle f, \varphi_i \rangle_{\mathcal{H}}\, \varphi_i + \langle f, \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}\, \phi_{\boldsymbol{x}} = \mathcal{P}_{\mathcal{H}^X}[f] + \mathcal{P}_{\langle \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}}[f].$$

Next, equality (5) follows straightforwardly as both $\langle \mathcal{P}_{\mathcal{H}^X}[f], \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}} = 0 = \langle f - \mathcal{P}_{\mathcal{H}^X}[f], \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)] \rangle_{\mathcal{H}}$ by projection properties:

$$\begin{aligned} \left\langle f, \tilde{\phi}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} &= \left\langle f - \mathcal{P}_{\mathcal{H}^X}[f], \tilde{\phi}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} = \langle f - \mathcal{P}_{\mathcal{H}^X}[f], K(\boldsymbol{x}, \cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)] \rangle_{\mathcal{H}} \\ &= \langle f - \mathcal{P}_{\mathcal{H}^X}[f], K(\boldsymbol{x}, \cdot) \rangle_{\mathcal{H}} = f(\boldsymbol{x}) - \mathcal{P}_{\mathcal{H}^X}[f](\boldsymbol{x}). \end{aligned}$$

Finally, the Riesz representation directly follows from (5). □

So, Lemma 3 allows to show the connection between both ways of notation, which is established mainly via the reproducing property of the RKHS. In equation (5) we see the scalar product of f with a dictionary element $K(\boldsymbol{x}, \cdot)$ in the spirit of general greedy algorithms, but also the pointwise difference $f(\boldsymbol{x}) - \mathcal{P}_{\mathcal{H}^X}[f](\boldsymbol{x})$, which is the same as $f(\boldsymbol{x}) - s_{f,X}(\boldsymbol{x})$ by Lemma 2 for nonsingular kernel matrices. The expression $\langle f, \tilde{\phi}_{\boldsymbol{x}} \rangle_{\mathcal{H}}$ is equivalent to both, but isolates the f-dependency nicely by using a *modified* dictionary element, i.e. the orthogonal remainder of $K(\boldsymbol{x}, \cdot)$.

Remark 2 For any $X \subset \Omega$ and $\boldsymbol{x} \in \Omega \backslash \Omega_X$ the orthogonal remainder $\tilde{\phi}_{\boldsymbol{x}}$ actually corresponds to the direct translate $K_P(\boldsymbol{x}, \cdot)$ of the *Power-Kernel* K_P, see [16, 17].

1.3 The f/P-Greedy algorithm

In order to establish the link to the formalism of the f/P-Greedy algorithm [18], we show the relation of orthogonal remainders to the concept of power functions.

Proposition 1 (Power function and orthogonal remainders) *If for $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ the kernel matrix $\boldsymbol{K}$ is nonsingular, then*

$$\left|\left| \tilde{\phi}_{\boldsymbol{x}} \right|\right|_{\mathcal{H}} = P_{K,X}(\boldsymbol{x}), \tag{6}$$

where $P_{K,X}(\boldsymbol{x})$ denotes the power function defined by (see [38, 11.2] or [18, 2.2.11])

$$P_{K,X}(\boldsymbol{x})^2 := K(\boldsymbol{x}, \boldsymbol{x}) - 2 \sum_{i=1}^{N} u_i(\boldsymbol{x}) K(\boldsymbol{x}, \boldsymbol{x}_i) + \sum_{i,j}^{N} u_i(\boldsymbol{x}) u_j(\boldsymbol{x}) K(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

where $u_i, i = 1 \ldots N$ denotes the Lagrange-Basis of $\mathcal{H}^X$ satisfying $u_i(\boldsymbol{x}_j) = \delta_{ij}$.

Proof. The Lagrange-Basis is given by

$$u_j(\boldsymbol{x}) = \sum_{i=1}^{N} \beta_{ji} K(\boldsymbol{x}_i, \boldsymbol{x}),$$

where the condition $u_i(\boldsymbol{x}_j) = \delta_{ij}$ is satisfied when $\boldsymbol{K}\boldsymbol{\beta}_j = \boldsymbol{e}_j$, i.e. $\boldsymbol{\beta}_j = (\boldsymbol{K}^{-1})_j$. Next, using the kernel column vector shorthand $\boldsymbol{k}(\boldsymbol{x}) = \big(K(\boldsymbol{x}, \boldsymbol{x}_1), \dots, K(\boldsymbol{x}, \boldsymbol{x}_N)\big)^T \in \mathbb{R}^N$ and Lemma 2 we see that

$$\begin{aligned}\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)] &= \sum_{j=1}^{N} (\boldsymbol{K}^{-1})_j^T \boldsymbol{k}(\boldsymbol{x}) K(\boldsymbol{x}_j, \cdot) = \sum_{j=1}^{N}\sum_{i=1}^{N} \boldsymbol{K}^{-1}_{ij} K(\boldsymbol{x}, \boldsymbol{x}_i) K(\boldsymbol{x}_j, \cdot) \\ &= \sum_{i=1}^{N} K(\boldsymbol{x}, \boldsymbol{x}_i) \sum_{j=1}^{N} \boldsymbol{K}^{-1}_{ij} K(\boldsymbol{x}_j, \cdot) = \sum_{i=1}^{N} K(\boldsymbol{x}, \boldsymbol{x}_i) u_j.\end{aligned}$$

By definition of $\tilde{\phi}_{\boldsymbol{x}}$ this yields

$$\begin{aligned}\left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|^2_{\mathcal{H}} &= \left\langle \tilde{\phi}_{\boldsymbol{x}}, \tilde{\phi}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} = \langle K(\boldsymbol{x}, \cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)], K(\boldsymbol{x}, \cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)] \rangle_{\mathcal{H}} \\ &= \left\langle K(\boldsymbol{x}, \cdot) - \sum_{i=1}^{N} u_j K(\boldsymbol{x}, \boldsymbol{x}_i), K(\boldsymbol{x}, \cdot) - \sum_{i=1}^{N} u_j K(\boldsymbol{x}, \boldsymbol{x}_i) \right\rangle_{\mathcal{H}} = P_{K,X}(\boldsymbol{x})^2,\end{aligned}$$

showing (6). □

For more background on power functions see e.g. [4, 18, 23, 38]. Even though both concepts are closely related, we will use the notion of orthogonal remainders as it will prove useful in our algorithm analysis.

With the necessary background established, we now state the scalar f/P-Greedy algorithm [18, 3.1.1] using the adopted notation in Algorithm 1. The equivalency in (7) the can be easily verified

Algorithm 1 f/P-Greedy algorithm

Let $f \in \mathcal{H}$ and define $X_0 := \emptyset, f^0 := 0$ and for $m > 0$ the sequences

$$\begin{aligned}\boldsymbol{x}_m &:= \arg\max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \frac{|f(\boldsymbol{x}) - f_m(\boldsymbol{x})|}{P_{K,X_m}(\boldsymbol{x})} = \arg\max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \left|\left\langle f, \phi^{m-1}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}}\right|, \\ X_m &:= X_{m-1} \cup \{\boldsymbol{x}_m\}, \\ f^m &:= \mathcal{P}_{\mathcal{H}^{X_m}}[f],\end{aligned} \tag{7}$$

where $\phi^m_{\boldsymbol{x}}$ denotes the orthonormal remainder of $K(\boldsymbol{x}, \cdot)$ with respect to X_m for any $m, \boldsymbol{x}$.

using Proposition 1 and Lemma 3.

2 Vectorial kernel orthogonal greedy algorithm

As mentioned in the introduction, we want to consider approximations of functions from vectorial RKHS. Before we can state our vectorial greedy algorithm, we introduce the vectorial kernel spaces we will be dealing with.

Definition 6 (Vectorial Hilbert Spaces) Let $q \in \mathbb{N}$. Then we denote by

$$\mathcal{H}^q := \{f : \Omega \to \mathbb{R}^q \mid f_j \in \mathcal{H}, j = 1 \dots q\}$$

the function space of vectorial functions from $\mathcal{H}$ which we equip with the canonical scalar product and norm

$$\langle f, g\rangle_{\mathcal{H}^q} := \sum_{j=1}^{q} \langle f_j, g_j\rangle_{\mathcal{H}}\,, \qquad ||f||_{\mathcal{H}^q} = \sqrt{\langle f, f\rangle_{\mathcal{H}^q}} = \sqrt{\sum_{j=1}^{q} ||f_j||^2_{\mathcal{H}}}$$

For this type of vectorial spaces, it is clear that any scalar greedy approximation strategy can be straightforwardly applied to each component function f_j for some $f \in \mathcal{H}^q$. But if one does not somehow connect the extension choices over the different component functions, the algorithms will most likely produce different subspaces $\mathcal{H}^{X_j}$ for each component f_j, leading to q disjoint kernel expansions in the worst case. This will be computationally infeasible, so the first and most obvious choice is to force all component approximations to stem from one global approximation subspace, i.e. $f_j \in \mathcal{H}^X\ \forall\ j$ for some base space $\mathcal{H}^X$. This restriction is given if we use the following vectorial projection operator on $\mathcal{H}^q$.

Definition 7 (Vectorial component-wise projection operator) Let $S \subseteq \mathcal{H}$ be a linear subspace and $q \in \mathbb{N}$. Then we define the vectorial orthogonal projection operator

$$\begin{aligned}\mathcal{P}^q_S : \mathcal{H}^q &\to S^q\\ f &\mapsto \left(\mathcal{P}_S[f_j]\right)_j\,, \quad j = 1 \ldots q.\end{aligned}$$

It is easily verifiable that for this definition we have $\langle f - \mathcal{P}^q_S[f], g\rangle_{\mathcal{H}^q} = 0\ \forall\ g \in S^q$ as

$$S^q = \langle \{e_i g \mid i = 1 \ldots q,\ g \in S\}\rangle\,,$$

with e_i denoting the i-th unit vector in $\mathbb{R}^q$.

Consequently, Algorithm 1 can be formulated straightforwardly also in the vectorial context, which is done in Algorithm 2.

Algorithm 2 Vectorial f/P-Greedy

Let $q \in \mathbb{N}$ and $\boldsymbol{f} \in \mathcal{H}^q$, define $X_0 := \emptyset$, $\boldsymbol{f}^0 := 0$ and for $m > 0$ the sequences

$$\boldsymbol{x}_m := \arg \max_{\boldsymbol{x} \in \Omega \backslash \Omega_{X_{m-1}}} \left|\left\langle \boldsymbol{f}, \tilde{\boldsymbol{\phi}}^{m-1}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}^q}\right|, \tag{8}$$

$$X_m := X_{m-1} \cup \{\boldsymbol{x}_m\}, \tag{9}$$

$$\boldsymbol{f}^m := \mathcal{P}^q_{\mathcal{H}^{X_m}}[\boldsymbol{f}], \tag{10}$$

where $\tilde{\boldsymbol{\phi}}^{m-1}_{\boldsymbol{x}} \in \mathcal{H}^q$ denotes the vectorial repetition of $\tilde{\phi}^{m-1}_{\boldsymbol{x}}$, i.e. $(\tilde{\boldsymbol{\phi}}^{m-1}_{\boldsymbol{x}})_i = \tilde{\phi}^{m-1}_{\boldsymbol{x}}, i = 1 \ldots q$.

An important feature of the scalar f/P-Greedy algorithm is that each extension step maximizes the $\mathcal{H}$-norm of $\mathcal{P}_{\mathcal{H}^X}[f]$ (the interpolant in [18, 3.1.4], see also [27, Thm 6.], [5]), which is equivalent to achieving the largest possible "gain" in approximation due to the orthogonality conditions $||\mathcal{P}_{\mathcal{H}^X}[f]||^2_{\mathcal{H}} = ||f||^2_{\mathcal{H}} - ||f - \mathcal{P}_{\mathcal{H}^X}[f]||^2_{\mathcal{H}}$. This aspect is not taken into account by the vectorial greedy algorithms proposed in [12, 13], which pursue a vectorial greedy search in the fashion of the standard scalar f-Greedy variant. However, it remains to verify that the canonical vectorial selection criteria of Algorithm 2 inherits this property. The concept of a *gain function* will prove useful in this context.

Definition 8 (Gain function) Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} \subseteq \Omega$ and $\boldsymbol{f} \in \mathcal{H}^q$. Then we denote the vectorial gain function with respect to X and $\boldsymbol{f}$ by

$$\begin{aligned}G_{X,\boldsymbol{f}} : \Omega\backslash\Omega_X &\to \mathbb{R},\\ \boldsymbol{x} &\mapsto \sum_{j=1}^{q} \langle f_j, \phi_{\boldsymbol{x}}\rangle^2_{\mathcal{H}}\,,\end{aligned} \tag{11}$$

where $\phi_{\boldsymbol{x}}$ denotes the orthonormal remainder of $K(\boldsymbol{x}, \cdot)$ with respect to X.

For our choice of vectorial RKHS, the following Lemma characterizes the $\mathcal{H}^q$-norm maximizing aspect of the f/P-Greedy algorithm using the gain function.

Lemma 4 (Locally optimal vectorial subspace extension) *Let $\boldsymbol{f} \in \mathcal{H}^q$, $X = \{x_1, \dots, x_m\} \subseteq \Omega$, $q \in \mathbb{N}$ and $\boldsymbol{x} \in \Omega \backslash \Omega_X$. Then*

$$\left|\left|\boldsymbol{f} - \mathcal{P}^q_{\mathcal{H}^X \oplus \langle K(\boldsymbol{x},\cdot)\rangle}[\boldsymbol{f}]\right|\right|^2_{\mathcal{H}^q} = \left|\left|\boldsymbol{f} - \mathcal{P}^q_{\mathcal{H}^X}[\boldsymbol{f}]\right|\right|^2_{\mathcal{H}^q} - G_{X,\boldsymbol{f}}(\boldsymbol{x}), \tag{12}$$

where $f_j \in \mathcal{H}$ denotes the $j-th$ component function of $\boldsymbol{f} \in \mathcal{H}^q$.

Proof. At first we note that in fact $\mathcal{H}^X \oplus \langle K(\boldsymbol{x},\cdot)\rangle = \mathcal{H}^X \oplus \langle \phi_{\boldsymbol{x}}\rangle$, which follows directly from the definition of $\phi_{\boldsymbol{x}}$ as $\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x},\cdot)] \in \mathcal{H}^X$. Moreover, for an $f \in \mathcal{H}$ we have $\mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f] = \langle f, \phi_{\boldsymbol{x}}\rangle_{\mathcal{H}}\, \phi_{\boldsymbol{x}}$ by (2). Then using Lemma 2 and 3 we deduce

$$\begin{aligned}
\left|\left|f - \mathcal{P}_{\mathcal{H}^X \oplus \langle K(\boldsymbol{x},\cdot)\rangle}[f]\right|\right|^2_{\mathcal{H}} &= \left|\left|f - \mathcal{P}_{\mathcal{H}^X \oplus \langle \phi_{\boldsymbol{x}}\rangle}[f]\right|\right|^2_{\mathcal{H}} \\
&= \left|\left|f - \mathcal{P}_{\mathcal{H}^X}[f] - \mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f]\right|\right|^2_{\mathcal{H}} \\
&= ||f - \mathcal{P}_{\mathcal{H}^X}[f]||^2_{\mathcal{H}} - 2\left\langle f - \mathcal{P}_{\mathcal{H}^X}[f], \mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f]\right\rangle_{\mathcal{H}} + \left|\left|\mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f]\right|\right|^2_{\mathcal{H}} \\
&= ||f - \mathcal{P}_{\mathcal{H}^X}[f]||^2_{\mathcal{H}} - 2\left\langle f, \mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f]\right\rangle_{\mathcal{H}} \\
&\quad - 2\underbrace{\left\langle \mathcal{P}_{\mathcal{H}^X}[f], \mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f]\right\rangle_{\mathcal{H}}}_{=0} + \left\langle f, \mathcal{P}_{\langle\phi_{\boldsymbol{x}}\rangle}[f]\right\rangle_{\mathcal{H}} \\
&= ||f - \mathcal{P}_{\mathcal{H}^X}[f]||^2_{\mathcal{H}} - \langle f, \phi_{\boldsymbol{x}}\rangle^2_{\mathcal{H}}
\end{aligned}$$

Using the definition of $\mathcal{P}^q$ and $G_{X,\boldsymbol{f}}$ we obtain

$$\begin{aligned}
\left|\left|\boldsymbol{f} - \mathcal{P}^q_{\mathcal{H}^X \oplus \langle K(\boldsymbol{x},\cdot)\rangle}[\boldsymbol{f}]\right|\right|^2_{\mathcal{H}^q} &= \sum_{j=1}^{q} \left|\left|f_j - \mathcal{P}_{\mathcal{H}^X \oplus \langle K(\boldsymbol{x},\cdot)\rangle}[f_j]\right|\right|^2_{\mathcal{H}} \\
&= \sum_{j=1}^{q} ||f_j - \mathcal{P}_{\mathcal{H}^X}[f_j]||^2_{\mathcal{H}} - \langle f_j, \phi_{\boldsymbol{x}}\rangle^2_{\mathcal{H}} \\
&= ||\boldsymbol{f} - \mathcal{P}_{\mathcal{H}^X}[\boldsymbol{f}]||^2_{\mathcal{H}^q} - G_{X,\boldsymbol{f}}(\boldsymbol{x}).
\end{aligned}$$

□

Since the projection into a linear subspace always gives the best possible approximation in that space, a direct consequence is the following Corollary.

Corollary 1 *Let the conditions from Lemma 4 hold. Then*

$$\inf_{\boldsymbol{x} \in \Omega \backslash \Omega_X} \min_{\boldsymbol{g} \in (\mathcal{H}^X \oplus \langle K(\boldsymbol{x},\cdot)\rangle)^q} ||\boldsymbol{f} - \boldsymbol{g}||^2_{\mathcal{H}^q} = C - \sup_{\boldsymbol{x} \in \Omega \backslash \Omega_X} G_{X,\boldsymbol{f}}(\boldsymbol{x})$$

with $C := ||\boldsymbol{f} - \mathcal{P}_{\mathcal{H}^X}[\boldsymbol{f}]||^2_{\mathcal{H}^q}$.

So, in fact, any $\boldsymbol{x} \in \Omega \backslash \Omega_X$ that yields the best approximation in $\mathcal{H}^{X \cup \{\boldsymbol{x}\}}$ (or largest possible gain with respect to the $\mathcal{H}^q$-norm) also maximizes $G_{X,\boldsymbol{f}}$. Consequently, we state in Algorithm 3 a modified variant of Algorithm 2, which we will consider for the rest of this work. We note here that in Algorithm 3 is closely related to the vectorial vector greedy algorithm "WSOGA2" introduced in [13, §3, (3.4)], with the difference of having a gain-maximizing extension selection instead of maximizing element selection. We will compare those algorithms in Section 4.

A recursive application of Lemma 4 yields the following result on the error decay and a Parseval-type identity.

Corollary 2 *Then the $\mathcal{H}^q$-approximation error is monotoneously decreasing and we have the identity*

$$||\boldsymbol{f} - \boldsymbol{f}^m||^2_{\mathcal{H}^q} = ||\boldsymbol{f}||^2_{\mathcal{H}^q} - \sum_{i=1}^{m}\sum_{j=1}^{q} \left\langle f_j, \phi^{i-1}_{\boldsymbol{x}_i}\right\rangle^2_{\mathcal{H}} = ||\boldsymbol{f}||^2_{\mathcal{H}^q} - \sum_{i=1}^{m} G_{X_{i-1},\boldsymbol{f}}(\boldsymbol{x}_i), \qquad \forall\ m > 0.$$

Algorithm 3 Vectorial kernel orthogonal greedy algorithm (VKOGA)

Let $q \in \mathbb{N}$ and $\boldsymbol{f} \in \mathcal{H}^q$, define $X_0 := \emptyset$, $\boldsymbol{f}^0 := 0$ and for $m > 0$ the sequences

$$\boldsymbol{x}_m := \arg \max_{\boldsymbol{x} \in \Omega \backslash \Omega_{X_{m-1}}} G_{X_{m-1},\boldsymbol{f}}(\boldsymbol{x}) = \arg \max_{\boldsymbol{x} \in \Omega \backslash \Omega_{X_{m-1}}} \sum_{j=1}^{q} \left\langle f_j, \phi_{\boldsymbol{x}}^{m-1} \right\rangle_{\mathcal{H}}^2, \tag{13}$$

$$X_m := X_{m-1} \cup \{\boldsymbol{x}_m\}, \tag{14}$$

$$\boldsymbol{f}^m := \mathcal{P}^q_{\mathcal{H}^{X_m}}[\boldsymbol{f}]. \tag{15}$$

2.1 Algorithm analysis

Until now we have only considered points $\boldsymbol{x} \in \Omega \backslash \Omega_X$ for possible extension of $\mathcal{H}^X$ via the induced kernel translate $K(\boldsymbol{x}, \cdot)$. So far in literature, it remains unanswered what happens in the vicinity of points $\boldsymbol{x} \in \Omega_X$. In numerical applications, one always works on discrete sets of points and hence this issue if of less importance. However, it is of general analytical interest to determine the behaviour of the selection criteria (13) for cases where $\boldsymbol{x} \to \boldsymbol{y} \in \Omega_X$. More precise, in this work we want to put some effort into analyzing the behaviour of $G_{X,\boldsymbol{f}}$ in the neighborhood of Ω_X.

The following Theorem yields an explicit expression for the limit functions $\phi_{\boldsymbol{x}}$ when $\boldsymbol{x} \to \boldsymbol{y} \in \Omega_X$. Assuming at least C^2-smoothness of K (C^1 w.r.t. each argument), it turns out that the limits in these cases can be described by orthonormal remainders of directional derivatives of $K(\boldsymbol{x}, \cdot)$.

Theorem 1 (Directional limit of orthonormal remainders) *Let $K \in C^2(\Omega \times \Omega)$ and $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \Omega$ so that the kernel matrix $\boldsymbol{K}$ is nonsingular. Further choose $\boldsymbol{x} \in \Omega_X$. Then $\forall\, \boldsymbol{v} \in \mathbb{R}^d$ we have*

$$\lim_{h \to 0} \phi_{\boldsymbol{x}+h\boldsymbol{v}} = \phi_{\boldsymbol{x}}^{\nabla v} \in \mathcal{H} \tag{16}$$

with

$$\phi_{\boldsymbol{x}}^{\nabla v} := \begin{cases} \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} \Big/ \left\| \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} \right\|_{\mathcal{H}} & , \ \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} \neq 0 \\ 0 & , \ \textit{else} \end{cases}, \qquad \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} := \boldsymbol{v}^T \nabla_1 K(\boldsymbol{x}, \cdot) - \mathcal{P}_{\mathcal{H}^X}[\boldsymbol{v}^T \nabla_1 K(\boldsymbol{x}, \cdot)],$$

where ∇_1 denotes the gradient operator w.r.t. the first argument.

Proof. Fix $\boldsymbol{v}$. By Lemma 1 we know that $\boldsymbol{v}^T \nabla_1 K(\boldsymbol{x}, \cdot) \in \mathcal{H}$. Further, as $K(\boldsymbol{x}, \cdot) \in \mathcal{H}^X$, Lemma 2 / (3) with $f = K(\boldsymbol{x}, \cdot)$ gives

$$K(\boldsymbol{x}, \cdot) = \sum_{j=1}^{N} (\boldsymbol{K}^{-1})_j^T \boldsymbol{k}(\boldsymbol{x}) K(\boldsymbol{x}_j, \cdot), \tag{17}$$

where $\boldsymbol{K}_j$ denotes the j-th column of $\boldsymbol{K}$. Note that for $\boldsymbol{x} = \boldsymbol{x}_k, k \in \{1 \ldots N\}$, equation (17) simplifies to $(\boldsymbol{K}^{-1})_j^T \boldsymbol{k}(\boldsymbol{x}_k) = \delta_{jk}$ as $\boldsymbol{k}(\boldsymbol{x}_k) = \boldsymbol{K}_k$. Now, for $h > 0$, the first order multivariate Taylor series of K at $\boldsymbol{x}$ gives

$$K(\boldsymbol{x} + h\boldsymbol{v}, \cdot) = K(\boldsymbol{x}, \cdot) + h\boldsymbol{v}^T \nabla_1 K(\boldsymbol{x}, \cdot) + \mathcal{O}\left(h^2 C(\boldsymbol{x}, \cdot)\right), \tag{18}$$

where $C(\boldsymbol{x}, \cdot)$ is independent of h, which we will thus omit in the following. Next, with the shorthand

$$\nabla \boldsymbol{K}(\boldsymbol{x}) = \left(\nabla_1 K(\boldsymbol{x}, \boldsymbol{x}_1) \ldots \nabla_1 K(\boldsymbol{x}, \boldsymbol{x}_N)\right) \in \mathbb{R}^{d \times N},$$

equation (18) directly gives the representation

$$\boldsymbol{k}(\boldsymbol{x} + h\boldsymbol{v}) = \boldsymbol{k}(\boldsymbol{x}) + h \nabla \boldsymbol{K}(\boldsymbol{x})^T \boldsymbol{v} + \mathcal{O}\left(h^2\right),$$

and together with (3) and (17) we see that

$$\begin{aligned}
\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}+h\boldsymbol{v},\cdot)] &= \sum_{j=1}^{N}(\boldsymbol{K}^{-1})_j^T\boldsymbol{k}(\boldsymbol{x}+h\boldsymbol{v})K(\boldsymbol{x}_j,\cdot)\\
&= \sum_{j=1}^{N}(\boldsymbol{K}^{-1})_j^T\left(\boldsymbol{k}(\boldsymbol{x})+h\nabla\boldsymbol{K}(\boldsymbol{x})^T\boldsymbol{v}\right)K(\boldsymbol{x}_j,\cdot)+\mathcal{O}\left(h^2\right)\\
&= K(\boldsymbol{x},\cdot)+h\sum_{j=1}^{N}(\boldsymbol{K}^{-1})_j^T\nabla\boldsymbol{K}(\boldsymbol{x})^T\boldsymbol{v}K(\boldsymbol{x}_j,\cdot)+\mathcal{O}\left(h^2\right)\\
&= K(\boldsymbol{x},\cdot)+h\sum_{j=1}^{N}(\boldsymbol{K}^{-1})_j^T\begin{pmatrix}\boldsymbol{v}^T\nabla_1K(\boldsymbol{x},\boldsymbol{x}_1)\\ \vdots\\ \boldsymbol{v}^T\nabla_1K(\boldsymbol{x},\boldsymbol{x}_N)\end{pmatrix}K(\boldsymbol{x}_j,\cdot)+\mathcal{O}\left(h^2\right)\\
&= K(\boldsymbol{x},\cdot)+h\mathcal{P}_{\mathcal{H}^X}[\boldsymbol{v}^T\nabla_1K(\boldsymbol{x},\cdot)]+\mathcal{O}\left(h^2\right)
\end{aligned}$$

Using (18) again we obtain the representation

$$\begin{aligned}
\tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}} &= K(\boldsymbol{x}+h\boldsymbol{v},\cdot)-\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}+h\boldsymbol{v},\cdot)]\\
&= h\boldsymbol{v}^T\nabla_1K(\boldsymbol{x},\cdot)-\mathcal{P}_{\mathcal{H}^X}[\boldsymbol{v}^T\nabla_1K(\boldsymbol{x},\cdot)]+\mathcal{O}\left(h^2\right)\\
&= h\tilde{\phi}_{\boldsymbol{x}}^{\nabla v}+\mathcal{O}\left(h^2\right)
\end{aligned}\tag{19}$$

Now if $\tilde{\phi}_{\boldsymbol{x}}^{\nabla v}\neq 0$ we see (16) by

$$\lim_{h\to 0}\phi_{\boldsymbol{x}+h\boldsymbol{v}}=\lim_{h\to 0}\frac{\tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}}}{\left|\left|\tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}}\right|\right|_{\mathcal{H}}}=\lim_{h\to 0}\frac{\tilde{\phi}_{\boldsymbol{x}}^{\nabla v}+\mathcal{O}\left(h\right)}{\left|\left|\tilde{\phi}_{\boldsymbol{x}}^{\nabla v}\right|\right|_{\mathcal{H}}+\mathcal{O}\left(h\right)}=\phi_x^{\nabla v}.$$

□

Now, Theorem 1 allows to draw interesting conclusions with regard to the situations where $\boldsymbol{x}\in\Omega_X$, i.e. $K(\boldsymbol{x},\cdot)\in\mathcal{H}^X$. At first we see that for any $X=\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N\},\boldsymbol{x}\in\Omega_X,\boldsymbol{v}\in\mathbb{R}^d$ and $\boldsymbol{f}\in\mathcal{H}^q$ we have

$$\lim_{h\to 0}G_{X,\boldsymbol{f}}(\boldsymbol{x}+h\boldsymbol{v})=\sum_{j=1}^{q}\left\langle f_j,\phi_{\boldsymbol{x}}^{\nabla v}\right\rangle_{\mathcal{H}}^2.$$

The resulting limit value depends on the direction $\boldsymbol{v}$ from which the limit is taken, which implies that $G_{X,\boldsymbol{f}}$ cannot be continuously extended on Ω in general. To illustrate the occuring discontinuities, Figure 1 shows the values of $\langle f,\phi_{\boldsymbol{x}}\rangle_{\mathcal{H}}$ (which corresponds to $G_{X,\boldsymbol{f}}$ in $q=1$ dimensions without the squared scalar product) for the test settings $q=1,\Omega=[-4,4]^2,X=\{(0,1),(-0.5,0),(2,-1),(-1,3),(-1.5,-3)\}$ and a suitable $f\in\mathcal{H}$. The discontinuities are clearly recognizable around any point $\boldsymbol{x}\in X$, which are marked by red dots. Furthermore, for each $\boldsymbol{v}\in\mathbb{R}^d$ equation (12) now reads as

$$\lim_{h\to 0}\left|\left|\boldsymbol{f}-\mathcal{P}^q_{\mathcal{H}^X\oplus\langle K(\boldsymbol{x}+h\boldsymbol{v},\cdot)\rangle}[\boldsymbol{f}]\right|\right|_{\mathcal{H}^q}^2=\left|\left|\boldsymbol{f}-\mathcal{P}^q_{\mathcal{H}^X}[\boldsymbol{f}]\right|\right|_{\mathcal{H}^q}^2-\sum_{j=1}^{q}\left\langle f_j,\phi_{\boldsymbol{x}}^{\nabla v}\right\rangle_{\mathcal{H}}^2,\qquad\forall\ \boldsymbol{f}\in\mathcal{H}^q.$$

Hence, the possible different $\phi_{\boldsymbol{x}}^{\nabla v}$ at any $\boldsymbol{x}\in\Omega_X$ imply that the limit of the left hand side also differs with changing $\boldsymbol{v}$. This raises the question about the projection limit as $h\to 0$, which is answered satisfactory by the following corollary.

Corollary 3 *Let the conditions from Theorem 1 hold and let $\boldsymbol{f}\in\mathcal{H}^q$. Then we have*

$$\lim_{h\to 0}\mathcal{P}^q_{\mathcal{H}^X\oplus\langle K(\boldsymbol{x}+h\boldsymbol{v},\cdot)\rangle}[\boldsymbol{f}]=\mathcal{P}^q_{\mathcal{H}^X\oplus\langle\boldsymbol{v}^T\nabla_1K(\boldsymbol{x},\cdot)\rangle}[\boldsymbol{f}].$$

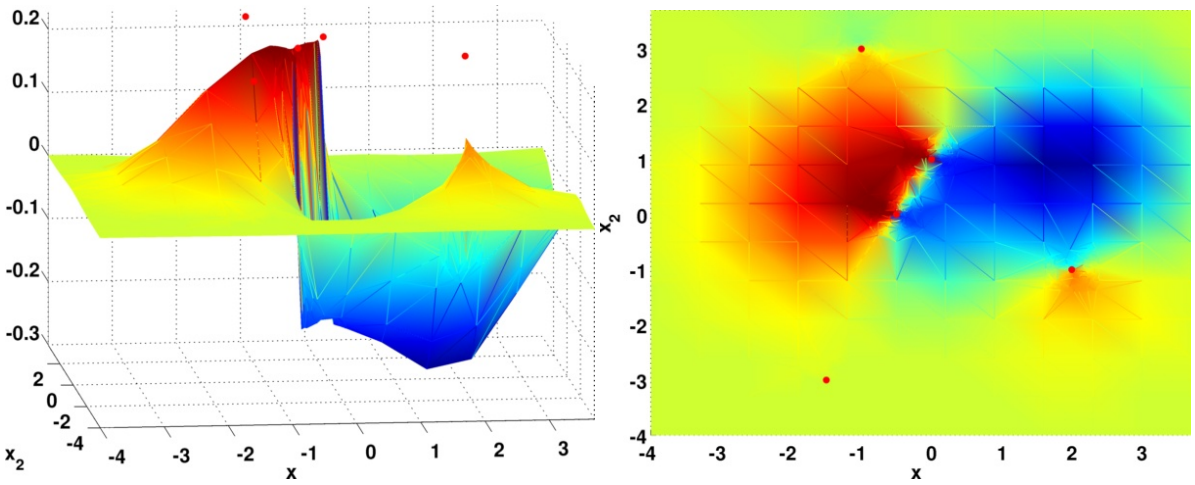

Fig. 1 Example of $\langle f, \phi_{\boldsymbol{x}} \rangle_{\mathcal{H}}$ on $\Omega \backslash \Omega_X$. Red dots: X

Proof. Using the relation (19) we obtain for each component function $f_j, j = 1 \dots q$ that

$$\begin{aligned}
\mathcal{P}_{\mathcal{H}^X \oplus \langle K(\boldsymbol{x}+h\boldsymbol{v},\cdot)\rangle}[f_j] &= \mathcal{P}_{\mathcal{H}^X}[f_j] + \mathcal{P}_{\langle \tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}} \rangle}[f_j] = \mathcal{P}_{\mathcal{H}^X}[f_j] + \frac{\left\langle f_j, \tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}} \right\rangle_{\mathcal{H}}}{\left|\left|\tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}}\right|\right|^2_{\mathcal{H}}} \tilde{\phi}_{\boldsymbol{x}+h\boldsymbol{v}} \\
&= \mathcal{P}_{\mathcal{H}^X}[f_j] + \frac{\left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} \right\rangle_{\mathcal{H}}}{\left|\left|\tilde{\phi}_{\boldsymbol{x}}^{\nabla v}\right|\right|^2_{\mathcal{H}}} \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} + \mathcal{O}\left(h^2\right) = \mathcal{P}_{\mathcal{H}^X}[f_j] + \mathcal{P}_{\langle \tilde{\phi}_{\boldsymbol{x}}^{\nabla v} \rangle}[f_j] + \mathcal{O}\left(h^2\right) \\
&= \mathcal{P}_{\mathcal{H}^X \oplus \langle \boldsymbol{v}^T \nabla_1 K(\boldsymbol{x},\cdot)\rangle}[f_j] + \mathcal{O}\left(h^2\right).
\end{aligned}$$

Taking the limit $h \to 0$ and recalling the component-wise action of $\mathcal{P}^q$ finishes the proof. □

Remark 3 The expression $\mathcal{P}^q_{\mathcal{H}^X \oplus \langle \boldsymbol{v}^T \nabla_1 K(\boldsymbol{x},\cdot)\rangle}[\boldsymbol{f}]$ actually corresponds to a simultaneous, component-wise directional Hermite interpolation in $\mathcal{H}^q$ since

$$\left\langle f_j, \boldsymbol{v}^T \nabla_1 K(\boldsymbol{x},\cdot) \right\rangle_{\mathcal{H}} = \boldsymbol{v}^T \nabla_1 f_j(\boldsymbol{x}), \quad j = 1 \dots N,$$

which can be easily verified using Lemma 1. Interestingly enough, this means that the closer a considered point approaches an already included one from the direction $\boldsymbol{v}$, the "direct extension" gain is converging towards the gain that would be achieved adding the directional derivative $\boldsymbol{v}^T \nabla_1 K(\boldsymbol{x},\cdot)$ to $\mathcal{H}^X$.

We conclude our analysis of $G_{X,\boldsymbol{f}}$ with the following Theorem.

Theorem 2 (Gain function characterization) *Let $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subseteq \Omega$ and $\boldsymbol{f} \in \mathcal{H}^q$. Then $G_{X,\boldsymbol{f}}$ is continuous on $\Omega \backslash \Omega_X$ and $\forall\ \boldsymbol{x} \in \Omega_X$ exists a neighborhood of $\boldsymbol{x}$ on which $G_{X,\boldsymbol{f}}$ is bounded.*

Proof. Let $\boldsymbol{x} \in \Omega \backslash \Omega_X$. Then for any $\boldsymbol{v} \in \mathbb{R}^d$ a similar argumentation with Taylor series as in the proof of Theorem 1 shows that

$$\lim_{h \to 0} ||\phi_{\boldsymbol{x}} - \phi_{\boldsymbol{x}+h\boldsymbol{v}}||_{\mathcal{H}} = 0,$$

from which we obtain continuity as $\lim_{\boldsymbol{y} \to \boldsymbol{x}} G_{X,\boldsymbol{f}}(\boldsymbol{y}) = G_{X,\boldsymbol{f}}(\boldsymbol{x})$. Next, for $\boldsymbol{x} \in \Omega_X$ and an ϵ with $B_\epsilon(\boldsymbol{x}) \subset \Omega$ we see that

$$\sup_{\boldsymbol{v} \in B_\epsilon(0)} \lim_{h \to 0} G_{X,\boldsymbol{f}}(\boldsymbol{x}+h\boldsymbol{v}) = \sup_{\boldsymbol{v} \in B_\epsilon(0)} \sum_{j=1}^{q} \left\langle f_j, \phi_{\boldsymbol{x}}^{\nabla v} \right\rangle^2_{\mathcal{H}} \leq \sup_{\boldsymbol{v} \in B_\epsilon(0)} \sum_{j=1}^{q} ||f_j||^2_{\mathcal{H}} \left|\left|\phi_{\boldsymbol{x}}^{\nabla v}\right|\right|^2_{\mathcal{H}} = ||\boldsymbol{f}||^2_{\mathcal{H}^q} < \infty,$$

which shows the boundedness. □

The most important conclusion from the above analysis for applications is, that the gain function $G_{X,\boldsymbol{f}}$ does not have any poles. This boundedness allows to actually obtain an analytic maxima in (13), presuming a suitable extension of $G_{X,\boldsymbol{f}}$ onto Ω. Furthermore, should it occur that the maximum of $G_{X,\boldsymbol{f}}$ is achieved at some $\boldsymbol{x} \in \Omega_X$ for suitable $\boldsymbol{v} \in \mathbb{R}^d$, this means that the extension of $\mathcal{H}^X$ with $\boldsymbol{v}^T \nabla_1 K(\boldsymbol{x}, \cdot)$ yields a better improvement than inclusion of *any* direkt kernel translate $K(\boldsymbol{x}, \cdot)$. However, in practical applications of this algorithm we yet consider only direct kernel translates for inclusion. This is why we will assume for the remainder of this work to extend $G_{X,\boldsymbol{f}}$ onto Ω by setting $G_{X,\boldsymbol{f}}(\boldsymbol{x}) := 0 \ \forall \ \boldsymbol{x} \in \Omega_X$.

Remark 4 If in the context of Theorem 1 K is actually induced by a radial basis function ϕ via $K(\boldsymbol{x}, \boldsymbol{y}) = \phi(||\boldsymbol{x} - \boldsymbol{y}||)$, we directly obtain

$$\boldsymbol{g}_x := \phi'(||\boldsymbol{x} - \cdot||) \left\langle \boldsymbol{v}, \frac{\boldsymbol{x} - \cdot}{||\boldsymbol{x} - \cdot||} \right\rangle - \sum_{j=1}^{N} K(\boldsymbol{x}_j, \cdot) \sum_{i=1}^{N} K_{ij}^{-1} \phi'(||\boldsymbol{x} - \boldsymbol{x}_i||) \left\langle \boldsymbol{v}, \frac{\boldsymbol{x} - \boldsymbol{x}_i}{||\boldsymbol{x} - \boldsymbol{x}_i||} \right\rangle,$$

with K_{ij}^{-1} denoting the ij-th entry of $\boldsymbol{K}^{-1}$. It is interesting to see that, for all directions that are free of any part towards the other centers (i.e. $\boldsymbol{v} \perp \boldsymbol{x} - \boldsymbol{x}_i$ for all i), the contribution of the projection vanishes completely.

Remark 5 If we assume $d = 1$ in the context of Proposition 2, then the limit of ϕ_x for $x \to \tilde{x} \in \Omega_X$ is unique and $G_{X,\boldsymbol{f}}$ is continuous on $\Omega \ \ \forall \ f \in \mathcal{H}$.

2.2 Convergence analysis

In this Section we want to investigate the convergence behaviour of Algorithm 3, where we will prove a slightly improved convergence bound similar to the one established in [13] with a yet simplified proof. Note that this type of convergence rate stems from the more general theory of greedy algorithms in vectorial Hilbert/Banach spaces [13, 14]. On the other hand, there are various results for greedy algorithm convergence/approximation error bounds in the scalar setting for RKHS, which usually involve the concept of a *fill distance*, see [18, 22, 38] to name a few. Due to their generality, the foremost mentioned Temlyakov-style error bounds are often too conservative, while the fill distance-related bounds provide excellent convergence rates in many situations. However, the latter also suffer from the condition that a sufficiently small fill distance is hard to achieve in practice. Hence, it remains an open question if this "gap" can be closed in the future, as the practical convergence rates of kernel greedy algorithms are mostly much faster than the Temlyakov-bounds.

We will need some auxiliary lemmata before we can state our convergence results. The following Lemma was stated first in [35, 3.1] with proof in [6, Lemma 3.4], however as the referred proof is only similar we state it here for completeness.

Lemma 5 (Lemma 3.1 from [35]) *Let $M > 0, t_m, a_m \geq 0$ be non-negative sequences satisfying $a_0 \leq M, \ a_{m+1} \leq a_m(1 - t_{m+1}\frac{a_m}{M})$. Then*

$$a_m \leq M \left(1 + \sum_{k=1}^{m} t_k\right)^{-1} \qquad \forall \ m \geq 0 \tag{20}$$

Proof. If we have $a_{m_0} = 0$ for an $m_0 \geq 0$ we have $a_m = 0 \ \forall \ m \geq m_0$ and thus (20) holds trivially. So assume $a_m \neq 0$ for all $m \geq 0$ and we continue by induction. Then for $m = 0$ equation (20) is given by prerequisite. The induction step $m \to m + 1$ can then be performed using the third binomial formula $(1 - b)(1 + b) = 1 - b^2 \leq 1$ and the prerequisites:

$$\begin{aligned} a_{m+1}^{-1} &\geq a_m^{-1} \left(1 - t_{m+1}\frac{a_m}{M}\right)^{-1} \geq a_m^{-1}(1 + t_{m+1}\frac{a_m}{M}) \\ &= a_m^{-1} + \frac{t_{m+1}}{M} \geq \frac{1}{M}\left(1 + \sum_{k=1}^{m} t_k\right) + \frac{t_{m+1}}{M} \\ &= \frac{1}{M}\left(1 + \sum_{k=1}^{m+1} t_k\right). \end{aligned}$$

□

Lemma 6 (Vectorial young's inequality) *Let $a_n \in \mathbb{R}, n \in \mathbb{N}$ be an arbitrary sequence. Then*

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leq n \sum_{i=1}^{n} a_i^2 \ \forall\ n \in \mathbb{N}. \tag{21}$$

Proof. Case $n = 1$ holds trivially. So (21) hold for $n \in \mathbb{N}$ arbitrary but fixed. By the third binomial formula or young's inequality for products and exponent 2 we have $2ab \leq a^2 + b^2\ \forall\ a, b \in \mathbb{R}$. Applying this n times gives

$$\begin{aligned}
\left(\sum_{i=1}^{n+1} a_i\right)^2 &= \sum_{i,j}^{n+1} a_i a_j = \sum_{i,j}^{n} a_i a_j + 2\sum_{i=1}^{n} a_i a_{n+1} + a_{n+1}^2 \\
&= \left(\sum_{i=1}^{n} a_i\right)^2 + 2\sum_{i=1}^{n} a_i a_{n+1} + a_{n+1}^2 \\
&\leq n\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n}(a_i^2 + a_{n+1}^2) + a_{n+1}^2 \\
&= n\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} a_i^2 + (n+1)a_{n+1}^2 \\
&= (n+1)\left(\sum_{i=1}^{n} a_i^2 + a_{n+1}^2\right) = (n+1)\sum_{i=1}^{n+1} a_i^2
\end{aligned}$$

□

A key aspect of the Temlyakov-type estimations is to consider a certain subclass of functions

$$\mathcal{H}_M^q := \left\{ \boldsymbol{f} \in \mathcal{H}^q \;\middle|\; f_j = \sum_{k=0}^{\infty} \alpha_k^j K(\boldsymbol{x}_k, \cdot), \sum_{k=0}^{\infty} |\alpha_k^j| \leq M, j = 1 \dots q \right\}$$

for $M > 0$. It is easy to see that especially $||\boldsymbol{f}||_{\mathcal{H}^q} \leq M\ \forall\ \boldsymbol{f} \in \mathcal{H}_M^q$. For more background on this methodology we refer to [6, §3].

Theorem 3 (Convergence rates of the VKOGA algorithm) *Let the conditions of Algorithm 3 hold and let $M > 0$. Then for any $\boldsymbol{f} \in \mathcal{H}_M^q$, $\boldsymbol{f}^m$ converges to $\boldsymbol{f}$ no slower than*

$$||\boldsymbol{f} - \boldsymbol{f}^m||_{\mathcal{H}^q} \leq \sqrt{q} M \left(1 + \frac{m}{q}\right)^{-\frac{1}{2}}, \quad m \geq 0 \tag{22}$$

Further, with the definition

$$c_m := \max_{\boldsymbol{x} \in \Omega} \tilde{\phi}_{\boldsymbol{x}}^{m-1}(\boldsymbol{x}) = \max_{\boldsymbol{x} \in \Omega} \left|\left|\tilde{\phi}_{\boldsymbol{x}}^{m-1}\right|\right|_{\mathcal{H}}^2, \quad m > 0$$

we obtain the a-posteriori convergence bound

$$||\boldsymbol{f} - \boldsymbol{f}^m||_{\mathcal{H}^q} \leq \sqrt{q} M \left(1 + \frac{1}{q}\sum_{k=1}^{m} \frac{1}{c_k}\right)^{-\frac{1}{2}}, \quad m \geq 0 \tag{23}$$

Proof. Since $c_m \leq 1 \ \forall \ m > 0$ by Lemma 3 we obtain the a-priori bound (22) from (23) by setting $c_m := 1 \ \forall \ m > 0$, which leaves us to prove (23). First, using Lemma 3 we see for $j = 1 \dots q$ that

$$\begin{aligned}
\left|\left|f_j - f_j^{m-1}\right|\right|_{\mathcal{H}}^2 &= \left\langle f_j - f_j^{m-1}, f_j - f_j^{m-1}\right\rangle_{\mathcal{H}} = \left\langle f_j, f_j - f_j^{m-1}\right\rangle_{\mathcal{H}} \\
&= \sum_{k=1}^{\infty} \alpha_k^j \left\langle K(\boldsymbol{x}_k, \cdot), f_j - f_j^{m-1}\right\rangle_{\mathcal{H}} \\
&\leq \sum_{k=1}^{\infty} |\alpha_k^j| \left|\left\langle f_j, \tilde{\phi}_{x_k}^{m-1}\right\rangle_{\mathcal{H}}\right| \\
&\leq \sum_{k=1}^{\infty} |\alpha_k^j| \max_{x \in \Omega} \left|\left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}\right| \\
&\leq M \max_{x \in \Omega} \left|\left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}\right|.
\end{aligned}$$

Now, Lemma 2 gives

$$\left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|_{\mathcal{H}}^2 = ||K(\boldsymbol{x}, \cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)]||_{\mathcal{H}}^2 = 1 - \langle K(\boldsymbol{x}, \cdot), \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)]\rangle_{\mathcal{H}} \leq 1,$$

and together with Lemma 6 (twice) we estimate the vectorial gain term as

$$\begin{aligned}
G_{X_{m-1},\boldsymbol{f}}(\boldsymbol{x}_m) &= \max_{\boldsymbol{x} \in \Omega} G_{X_{m-1},\boldsymbol{f}}(\boldsymbol{x}) = \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} G_{X_{m-1},\boldsymbol{f}}(\boldsymbol{x}) \\
&= \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \sum_{j=1}^{q} \frac{1}{\left|\left|\tilde{\phi}_{\boldsymbol{x}}^{m-1}\right|\right|_{\mathcal{H}}^2} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 \geq \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \frac{1}{c_m} \sum_{j=1}^{q} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 \\
&\geq \frac{1}{c_m} \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \max_{j=1 \dots q} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 = \frac{1}{c_m} \max_{j=1 \dots q} \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 \\
&= \frac{1}{\sqrt{q} c_m} \left(q \max_{j=1 \dots q} \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^4 \right)^{\frac{1}{2}} \\
&\geq \frac{1}{\sqrt{q} c_m} \left(\sum_{j=1}^{q} \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^4 \right)^{\frac{1}{2}} \\
&\geq \frac{1}{\sqrt{q} c_m} \left(\frac{1}{q} \left(\sum_{j=1}^{q} \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 \right)^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{q c_m} \sum_{j=1}^{q} \max_{\boldsymbol{x} \in \Omega \setminus \Omega_{X_{m-1}}} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 \geq \frac{1}{q c_m} \sum_{j=1}^{q} \frac{\left|\left|f_j - f_j^{m-1}\right|\right|_{\mathcal{H}}^4}{M^2} \\
&\geq \frac{1}{q M^2 c_m} \frac{1}{q} \left(\sum_{j=1}^{q} \left|\left|f_j - f_j^{m-1}\right|\right|_{\mathcal{H}}^2 \right)^2 = \frac{1}{q^2 M^2 c_m} \left|\left|f - f^{m-1}\right|\right|_{\mathcal{H}^q}^4.
\end{aligned}$$

Using Lemma 4, we see that

$$\begin{aligned}
||f - f^m||_{\mathcal{H}^q}^2 &= \left|\left|f - f^{m-1}\right|\right|_{\mathcal{H}^q}^2 - G_{X_{m-1},\boldsymbol{f}}(\boldsymbol{x}_m) \\
&\leq \left|\left|f - f^{m-1}\right|\right|_{\mathcal{H}^q}^2 - \frac{1}{q^2 M^2 c_m} \left|\left|f - f^{m-1}\right|\right|_{\mathcal{H}^q}^4 \\
&= \left|\left|f - f^{m-1}\right|\right|_{\mathcal{H}^q}^2 \left(1 - \frac{\frac{1}{q c_m} \left|\left|f - f^{m-1}\right|\right|_{\mathcal{H}^q}^2}{q M^2}\right).
\end{aligned}$$

Further, with $f_j \in \mathcal{H}^M$ we have

$$\left|\left|f - f^0\right|\right|^2_{\mathcal{H}^q} = ||f||^2_{\mathcal{H}^q} = \sum_{j=1}^{q} ||f_j||^2_{\mathcal{H}} \leq \sum_{j=1}^{q} M^2 = qM^2.$$

Finally, applying Lemma 5 with $a_m = ||f - f^m||^2_{\mathcal{H}^q}$, $a_0 \leq qM^2$ and $t_m := \frac{1}{qc_m}$ gives

$$||f - f^m||^2_{\mathcal{H}^q} \leq qM^2 \left(1 + \sum_{k=1}^{m} \frac{1}{qc_k}\right)^{-1} \quad \forall\, m \in \mathbb{N},$$

and hence (23). □

3 Computational aspects

Before we present some numerical experiments we make some remarks on computational aspects of the VKOGA algorithm. The straightforward approach is, for any given set of points X, to use the standard basis of translates $\{\boldsymbol{K}(\boldsymbol{x}_1, \cdot), \ldots, \boldsymbol{K}(\boldsymbol{x}_N, \cdot)\}$ of $\mathcal{H}^X$ to obtain the projection via solving the system $\boldsymbol{K}\boldsymbol{c} = \boldsymbol{f}_{|X}$. However, it is well known that this "RBF-Direct" method suffers from ill conditioned kernel matrices $\boldsymbol{K}$, especially for point distributions with small distances. In order to alleviate those problems several approaches like preconditioning techniques [24] or RBF-QR [9, 10] have been developed, to name a few. However, in this work we will use the *Newton basis* formulated recently in [20].

Definition 9 (Newton-Basis of $\mathcal{H}^X$) Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\} \subseteq \Omega$ so that $\mathcal{H}^X$ is m-dimensional. Then the Newton basis $N_1, \ldots, N_m$ of $\mathcal{H}^X$ is given by the recursion

$$N_1 := \frac{K(\boldsymbol{x}_1, \cdot)}{\sqrt{K(\boldsymbol{x}_1, \boldsymbol{x}_1)}}, \qquad \tilde{N}_j = K(\boldsymbol{x}_j, \cdot) - \sum_{i=1}^{j-1} N_i(\boldsymbol{x}_i) N_i, \qquad N_j = \frac{\tilde{N}_j}{\left|\left|\tilde{N}_j\right|\right|_{\mathcal{H}}}, \quad j = 2 \ldots m. \tag{24}$$

and satisfies

$$\langle N_i, N_j \rangle_{\mathcal{H}} = \delta_{ij}. \tag{25}$$

Condition (25) is easily verified by induction. With this basis a stable computation of the projection at each greedy step is possible, without having to touch any previously computed coefficients again. Next we state the representations of the involved quantities with respect to the Newton basis and refer to [20] for more details on this approach.

Lemma 7 (Newton basis representations) *Let $N_1, \ldots, N_m$ be the Newton-Basis of $\mathcal{H}^X$ and $f \in \mathcal{H}$. Then*

$$\mathcal{P}_{\mathcal{H}^X}[f] = \sum_{i=1}^{m} \langle f, N_i \rangle_{\mathcal{H}} N_i, \tag{26}$$

$$\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x}, \cdot)] = \sum_{i=1}^{m} N_i(\boldsymbol{x}) N_i, \tag{27}$$

$$\left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|^2_{\mathcal{H}} = K(\boldsymbol{x}, \boldsymbol{x}) - \sum_{i=1}^{m} N_i^2(\boldsymbol{x}) \tag{28}$$

Proof. From the projection conditions (2) and (25) we immediately obtain (26). Equation (27) is a special case of (26) for $f = K(\boldsymbol{x}, \cdot)$.

Using (27) and Lemma 2 gives (28) via

$$\begin{aligned}\left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|_{\mathcal{H}}^2 &= \langle K(\boldsymbol{x},\cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x},\cdot)], K(\boldsymbol{x},\cdot) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x},\cdot)]\rangle_{\mathcal{H}} \\ &= K(\boldsymbol{x},\boldsymbol{x}) - 2\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x},\cdot)](\boldsymbol{x}) + ||\mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x},\cdot)]||_{\mathcal{H}}^2 \\ &= K(\boldsymbol{x},\boldsymbol{x}) - \mathcal{P}_{\mathcal{H}^X}[K(\boldsymbol{x},\cdot)](\boldsymbol{x}) = K(\boldsymbol{x},\boldsymbol{x}) - \sum_{i=1}^{m} N_i^2(\boldsymbol{x}).\end{aligned}$$

□

Finally the following Proposition states how the VKOGA Algorithm can be computed efficiently using the Newton basis.

Proposition 2 (Computation of VKOGA with Newton-Basis) *For $m = 1$ set*

$$\begin{aligned}\boldsymbol{x}_1 &:= \arg\max_{\boldsymbol{x}\in\Omega} G_{\emptyset,\boldsymbol{f}}(\boldsymbol{x}) = \arg\max_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{q} f_j(\boldsymbol{x})^2, \\ \boldsymbol{c}_1 &:= (\langle f_1, N_1\rangle_{\mathcal{H}}, \ldots, \langle f_q, N_1\rangle_{\mathcal{H}})^T = \sqrt{K(\boldsymbol{x}_1,\boldsymbol{x}_1)}^{-1}(f_1(\boldsymbol{x}),\ldots,f_q(\boldsymbol{x}))^T \in \mathbb{R}^q.\end{aligned}$$

Then, at the $m > 1$-th iteration with given $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{m-1}$, $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_{m-1}$ we define

$$\boldsymbol{x}_m := \arg\max_{\boldsymbol{x}\in\Omega\setminus\Omega_{X_{m-1}}} \left|\left|\boldsymbol{f}(\boldsymbol{x}) - \sum_{i=1}^{m-1} \boldsymbol{c}_i N_i(\boldsymbol{x})\right|\right|_2^2 \left(K(\boldsymbol{x},\boldsymbol{x}) - \sum_{i=1}^{m-1} N_i^2(\boldsymbol{x})\right)^{-1}, \tag{29}$$

$$\boldsymbol{c}_m := \begin{pmatrix}\langle f_1, N_m\rangle_{\mathcal{H}} \\ \vdots \\ \langle f_q, N_m\rangle_{\mathcal{H}}\end{pmatrix} = \frac{\boldsymbol{f}(\boldsymbol{x}_m) - \sum_{i=1}^{m-1} \boldsymbol{c}_i N_i(\boldsymbol{x}_m)}{\left(K(\boldsymbol{x}_m,\boldsymbol{x}_m) - \sum_{i=1}^{m-1} N_i^2(\boldsymbol{x}_m)\right)^{\frac{1}{2}}} \in \mathbb{R}^q. \tag{30}$$

Proof. With Lemma 7 we see (29) by

$$G_{X_{m-1},\boldsymbol{f}}(\boldsymbol{x}) = \frac{\sum_{j=1}^{q}\left\langle f_j, \tilde{\phi}_{\boldsymbol{x}}\right\rangle_{\mathcal{H}}^2}{\left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|_{\mathcal{H}}^2} = \frac{\left|\left|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}^{m-1}(\boldsymbol{x})\right|\right|_2^2}{\left|\left|\tilde{\phi}_{\boldsymbol{x}}\right|\right|_{\mathcal{H}}^2} = \frac{\left|\left|\boldsymbol{f}(\boldsymbol{x}) - \sum_{i=1}^{m-1} \boldsymbol{c}_i N_i(\boldsymbol{x})\right|\right|_2^2}{K(\boldsymbol{x},\boldsymbol{x}) - \sum_{i=1}^{m} N_i^2(\boldsymbol{x})}$$

In order to see (30) we note that

$$\begin{aligned}\left|\left|\tilde{N}_m\right|\right|_{\mathcal{H}}^2 &= K(\boldsymbol{x}_m,\boldsymbol{x}_m) - \sum_{i=1}^{m-1} N_i^2(\boldsymbol{x}_m), \\ \langle f_j, N_m\rangle_{\mathcal{H}} &= \frac{\left\langle f_j, \tilde{N}_m\right\rangle_{\mathcal{H}}}{\left|\left|\tilde{N}_m\right|\right|_{\mathcal{H}}} = \frac{f_j(\boldsymbol{x}_m) - \sum_{i=1}^{m-1}\langle f_j, N_i\rangle_{\mathcal{H}} N_i(\boldsymbol{x}_m)}{\left(K(\boldsymbol{x}_m,\boldsymbol{x}_m) - \sum_{i=1}^{m-1} N_i^2(\boldsymbol{x}_m)\right)^{\frac{1}{2}}}.\end{aligned}$$

□

Note that the final approximant $\boldsymbol{f}^m$ will have the structure (26). In order to evaluate the expansion using the direct translate basis, the triangular matrix with values of the computed m Newton basis functions at the selected points X_m can be used to obtain the corresponding coefficients [20].

Remark 6 (Connection to remainders) In the context of Proposition 2 we actually have $\tilde{\phi}_{x_m} = \tilde{N}_m$ and consequently $\phi_{x_m} = N_m$. This means that the orthonormal remainders in each step directly state all possible candidates of new Newton basis functions.

Algorithm 4 WSOGA2

Let K be a symmetric, positive definite and normalized kernel spanning the RKHS $\mathcal{H}$ on a closed $\Omega \subset \mathbb{R}^d$. Further let $\boldsymbol{f} \in \mathcal{H}^q$, define $X_0 := \emptyset, \boldsymbol{f}^0 := 0$ and for $m > 0$ the sequences

$$\begin{aligned} \boldsymbol{x}_m &= \arg\max_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{q} \left\langle f_j - f_j^{m-1}, K(\boldsymbol{x},\cdot)\right\rangle_{\mathcal{H}}^2 \\ X_m &:= X_{m-1} \cup \{\boldsymbol{x}_m\}, \\ f^m &:= \mathcal{P}^q_{\mathcal{H}^{X_m}}[f]. \end{aligned} \tag{31}$$

4 Numerical illustrations

As mentioned earlier, Algorithm 3 describes an algorithm similar to the vectorial algorithms presented in [13, §3], especially the variant "WSOGA2" at (3.4). We state it here fitted to our RKHS setting.

Remark 7 In the context of Algorithm 4 we have monotonicity of $\mathcal{H}^q$ error decay and the same a-priori convergence rate as for Algorithm 3 can be shown to apply. Note here that the proof of convergence rates of the WSOGA2-Algorithm has already been performed (in a more general setting) in [13], albeit using a different technique and obtaining a convergence rate which is a factor of $\sqrt{q}$ slower. Furthermore, $\boldsymbol{x}_m$ is chosen using the maximum local L^2 pointwise approximation error in the sense of

$$\boldsymbol{x}_m = \arg\max_{\boldsymbol{x}\in\Omega} ||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}^m(\boldsymbol{x})||_2^2 . \tag{32}$$

4.1 Analytical comparison of VKOGA and WSOGA2

Before we present some illustrating experimental results, we perform an analytical comparison and show how this can be interpreted. Let $X \subset \Omega$ be given and denote by $\boldsymbol{x}^o, \boldsymbol{x}^c$ the subspace extension choices the Algorithms 3 and 4, respectively. Then we see that

$$\begin{aligned} \sum_{j=1}^{q} \left\langle f_j - f_j^{m-1}, K(\boldsymbol{x}^c,\cdot)\right\rangle_{\mathcal{H}}^2 = \sum_{j=1}^{q} \left\langle f_j, \tilde{\phi}_{\boldsymbol{x}^c}\right\rangle_{\mathcal{H}}^2 &\leq \sum_{j=1}^{q} \langle f_j, \phi_{\boldsymbol{x}^c}\rangle_{\mathcal{H}}^2 \\ &\leq \max_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{q} \langle f_j, \phi_{\boldsymbol{x}}\rangle_{\mathcal{H}}^2 = G_{X,\boldsymbol{f}}(\boldsymbol{x}^o) \end{aligned}$$

by the selection criteria definitions. This means the VKOGA algorithm will locally always make as good a choice as the WSOGA2 algorithm. Unfortunately, as the successive spaces constructed by both algorithms will in general be different, it remains an open question to us if we can and how to compare the performance of both variants directly at some given subspace size $m > 0$. However, with the help of Lemma 3, the VKOGA extension choice criteria can also be written as

$$\max_{\boldsymbol{x}\in\Omega} \sum_{j=1}^{q} \left\langle f_j, \phi_{\boldsymbol{x}}^{m-1}\right\rangle_{\mathcal{H}}^2 = \max_{\boldsymbol{x}\in\Omega\setminus\Omega_{X_{m-1}}} \frac{||\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}^m(\boldsymbol{x})||_2^2}{\left|\left|K(\boldsymbol{x},\cdot) - \mathcal{P}_{\mathcal{H}^{X_{m-1}}}[K(\boldsymbol{x},\cdot)]\right|\right|_{\mathcal{H}}^2}$$

Since the numerator equals the WSOGA2 choice (32), which basically consideres *any* maximizing point $\boldsymbol{x}$ to be equally good for extension, the VKOGA choice scales inversely with how well the associated dictionary element $K(\boldsymbol{x},\cdot)$ is already approximated by $\mathcal{H}^X$. This way, identical pointwise approximation errors closer to the points whose dictionary elements span $\mathcal{H}^X$ are considered to be worse than others. Moreover, as the norm of any orthogonal remainder is independent of the considered $\boldsymbol{f} \in \mathcal{H}^q$, this can be interpreted as how well all functions involving the dictionary element $K(\boldsymbol{x},\cdot)$ in general are already approximated in $\mathcal{H}^q$. This concept is pursued directly by the P-Greedy algorithm mentioned e.g. in [18], which aims to create data-independent approximations of the function space and leads to a very uniform distribution of the selected $\boldsymbol{x}_m$. Figure 2 illustrates this issue using two simple scalar examples,

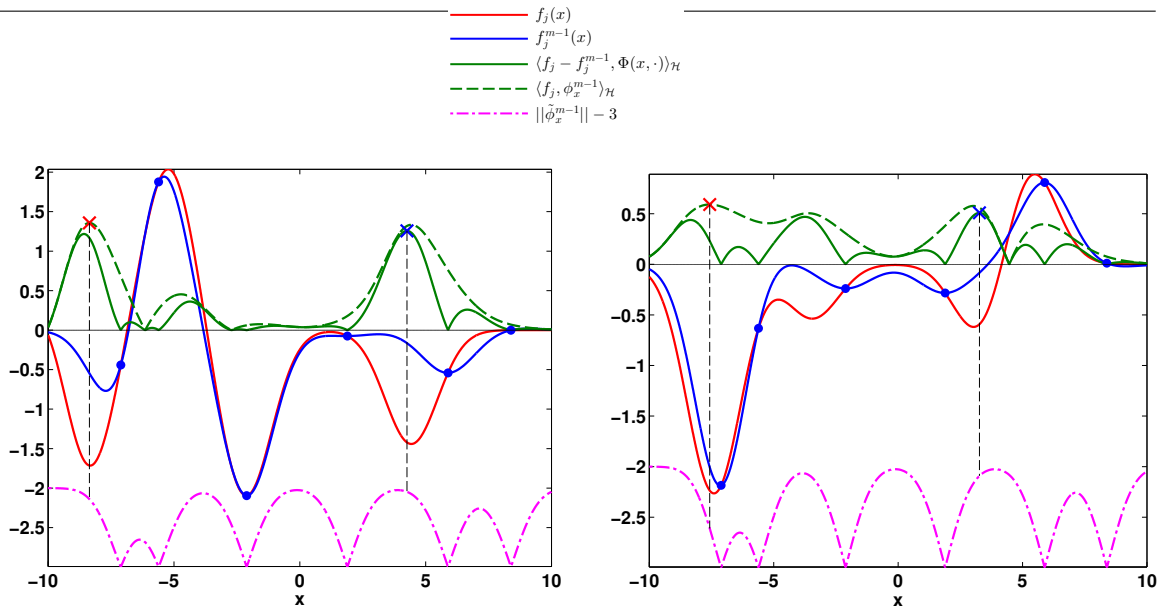

Fig. 2 Example for different selections of points using VKOGA/WSOGA for a scalar setting

which shows the different extension choices at $m = 6$ already given points along with the respective gain functions. The red and blue crosses mark the VKOGA and WSOGA2 selection and the dashed/solid green line the gain functions of the VKOGA/WSOGA2 algorithms, respectively. As test setting $X = \{-7.1, -5.6, -2.1, 1.9, 5.9, 8.4\}$ and a Gaussian with $\gamma^2 = 2.1715$ has been used, whose induced RKHS serves as native space for both dictionary and test function f. The expansion coefficients of f_1 on the left, f_2 on the right hand side are $\boldsymbol{\alpha}_1 = (-2.0465, 2.3066, -0.2428, 0.6805, -2.1213, -1.4411)$ and $\boldsymbol{\alpha}_2 = (1.1702, -0.2113, -0.7158, -0.5346, -1.1990, -1.1459)$ Note that it is clear to see that the gain function of the VKOGA algorithm is indeed continuous as mentioned earlier in Remark 5.

For both cases we see that the VKOGA choice selects spatially very different extension points, even though the gain of both algorithms is not very different. While in the case of f_1 (left) the extension points are selected "well away" from any existing point, the VKOGA extension selects x_7 very close to the existing center at -7.1. There it is evident that, even though the absolute approximation error is bigger elsewhere, the error weighted by the orthogonal remainder norm causes this location to be considered most worth improving. One the downside, especially because this choice involves a function-independent but RKHS-specific part, the choice of VKOGA can be ill-suited if the considered function does not stem from the same RKHS as the dictionary elements. In this situation, adding more points near the same location does not yield an excellent local convergence rate of the nominator as predicted by any fill-distance based approaches. However, as this is the case for the denominator, the rapid decay of the denominator causes even more points to be added in the same area. This effect can also be seen in e.g. [18, 6.1]. In our opinion, instead of considering the VKOGA algorithm a bad choice when the origin of the target function is unknown, we think this effect might be actively used to formulate an indicator for the foremost mentioned situation. If a considered function is "detected" not to belong to the currently chosen RKHS, one can proceed with another choice of RKHS, e.g. a different hyperparameter for the RKHS inducing kernel.

Remark 8 We would like to note that both algorithms can be continuously transferred over to each other by thresholding the orthogonal remainder norms at a certain value. This opens up a large variety of algorithms and the version most suitable for the current situation can be selected.

4.2 Experimental comparison of VKOGA and WSOGA2

Finally we want to pursue some numerical experiments for the truly vectorial case. We use $d = q = 5$, the test domain $\Omega = [-5, 5]^d$ and a Gaussian kernel with $\gamma = 9.8935$, which is chosen so that $K(\boldsymbol{x}, \boldsymbol{y}) <$

$0.6 \; \forall \; ||\boldsymbol{x} - \boldsymbol{y}|| \geq \sqrt{d \cdot \mathrm{diam}(\Omega)} = \sqrt{50}$, i.e. a certain locality of the kernel expansions is ensured. The test functions to approximate are of the structure

$$\boldsymbol{f}(\boldsymbol{x}) := \sum_{k=1}^{N} \boldsymbol{c}_k K(\boldsymbol{x}_k, \boldsymbol{x}) \in \mathcal{H}^q$$

with $N = 20$ random centers within Ω and random expansion coefficients $\boldsymbol{c} \in [0, 15]^q$. Experiments showed that it does not make a considerable difference in performance if we used $\boldsymbol{f} \in \mathcal{H}^q$ (i.e. independent expansions for each dimension) or $\boldsymbol{f} \in \left(\mathcal{H}^X\right)^q$ (a common center set $X \subseteq \Omega$ for each component function), as either way the actual centers are generally not detected/chosen as centers by the greedy algorithms.

For training we use 2500 training points in Ω and we use a validation set of size 1000. The algorithm terminates if the $L^\infty(L^2(\mathbb{R}^d); \mathbb{R}^{2500})$ relative error on training set is $\leq 10^{-4}$ or the expansion size exceeds $N = 200$. In order to avoid numerical issues we used $\sqrt{eps} = 2.2 \times 10^{-8}$ as minimum allowed value for any $\|\tilde{\phi}^m_{\boldsymbol{x}}\|_{\mathcal{H}}$. Figures 3 compares the results for the VKOGA and WSOGA2 algorithms.

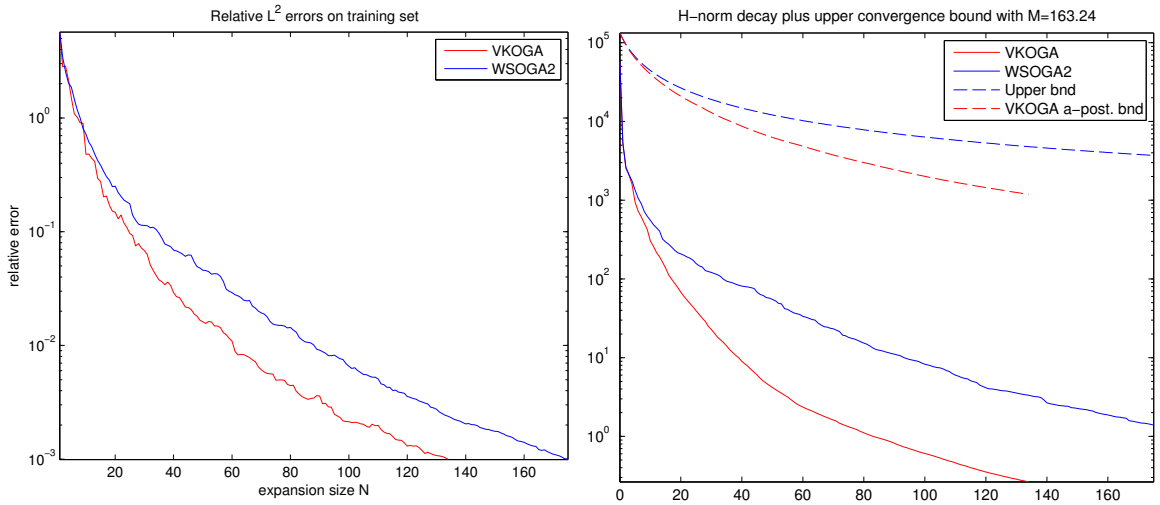

Fig. 3 Left: Relative errors on training set against expansion size. Right: Decay of $||f - f^m||_{\mathcal{H}^q}$ over iterations. Dashed lines: Upper bounds on convergence rates, WSOGA2 (22)/ VKOGA a-posteriori (23)

Now we run 50 tests for random test functions $\boldsymbol{f} \in \mathcal{H}^q$ and extracted some features of the results, which are displayed in Figures 4 and 4.2. Figure 4 shows that in average the VKOGA alrogithm outperforms the WSOGA2 variant on both the training and validation sets. Also, the maximum relative error on the validation set is smaller than one for each run of the VKOGA, while this is rarely the case for WSOGA2. Figure 4.2 shows the expansion sizes on the left and the $\mathcal{H}$-norm errors incl. bounds on the right at a reached relative L^2-error of 10^{-3} on the training data.

Remark 9 For higher dimensions the a-priori estimation assumptions can cause some problems. Running a test with $d = 20$, Figure 6 shows that even though we have an exponential convergence with N, the $\mathcal{H}$-norm convergence rate is worse than predicted. This is due to the following estimation in the first step of the convergence rate proof in Theorem 3:

$$\sum_{k=1}^{\infty} |\alpha_k^j| \left| \left\langle f_j, \tilde{\phi}^{m-1}_{\boldsymbol{x}_k} \right\rangle_{\mathcal{H}} \right| \leq \sum_{k=1}^{\infty} |\alpha_k^j| \max_{\boldsymbol{x} \in \Omega} \left| \left\langle f_j, \tilde{\phi}^{m-1}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} \right|$$

This estimation holds in theory, but since Ω is replaced by a discrete training set $\Xi \subseteq \Omega$ we might **not** have

$$\left| \left\langle f_j, \tilde{\phi}^{m-1}_{\boldsymbol{x}_k} \right\rangle_{\mathcal{H}} \right| \leq \max_{\boldsymbol{x} \in \Xi} \left| \left\langle f_j, \tilde{\phi}^{m-1}_{\boldsymbol{x}} \right\rangle_{\mathcal{H}} \right| \; \forall \; \boldsymbol{x}_k.$$

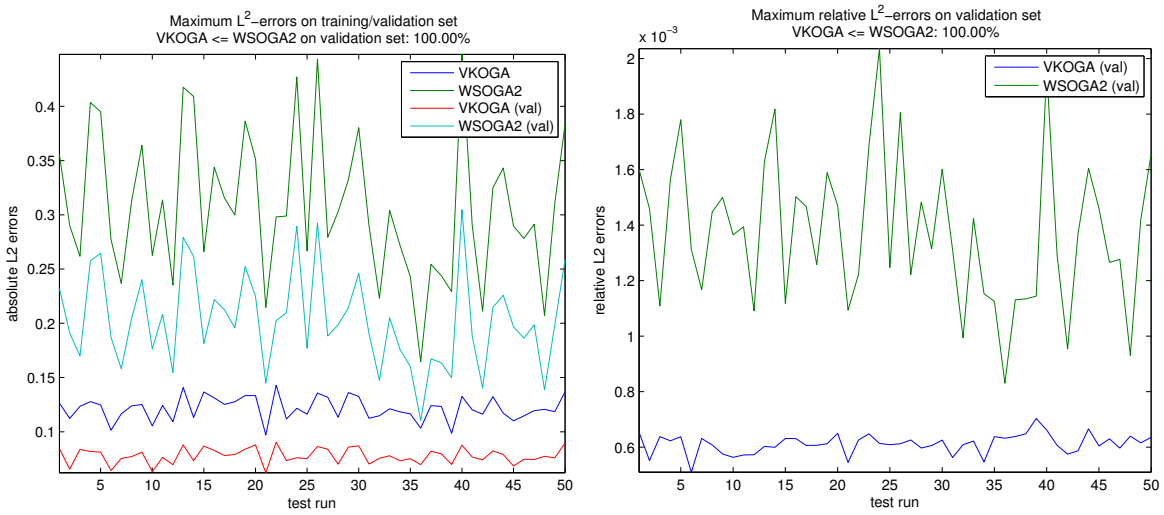

Fig. 4 $L^\infty - L^2$-errors after termination on training and validation sets. L: absolute, R: relative

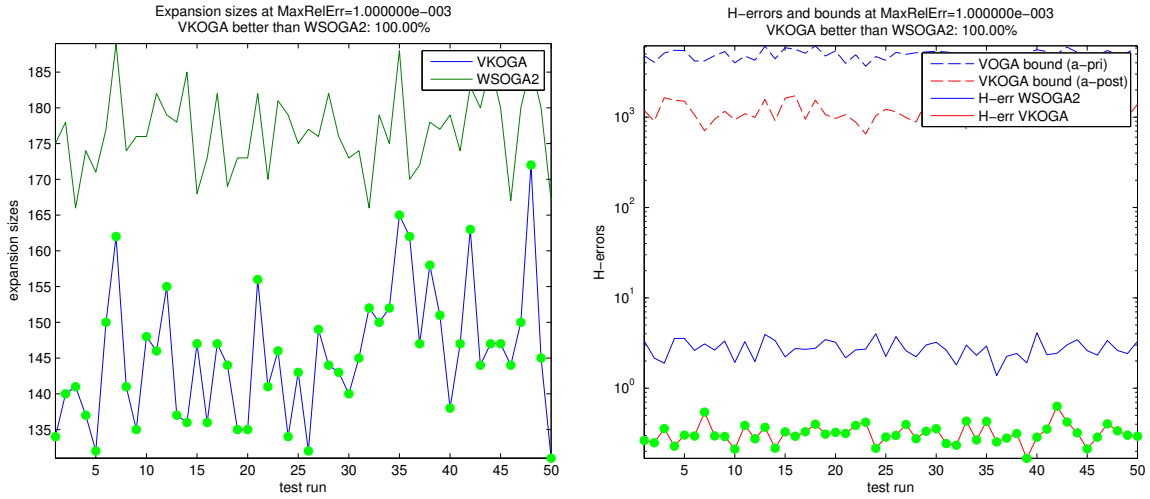

Fig. 5 L: Expansion sizes after termination, R: $||f - f^m||_{\mathcal{H}^q}$-errors and bounds

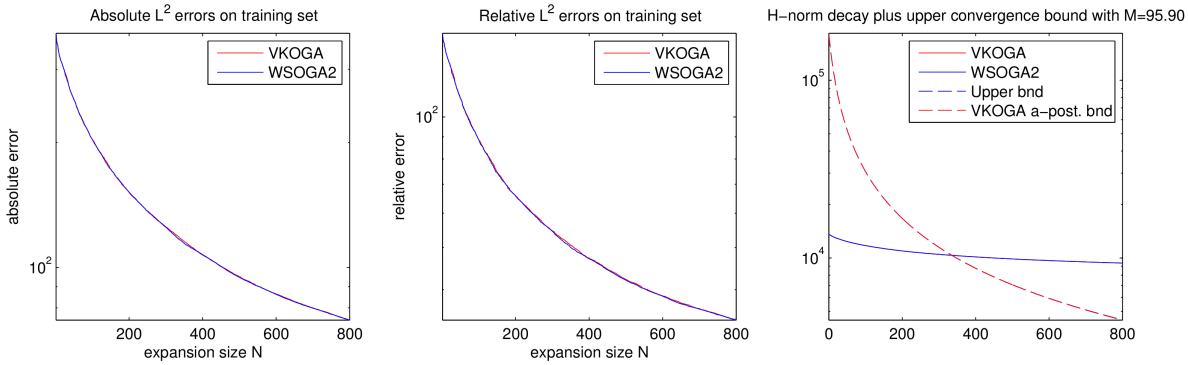

Fig. 6 Plots for $d = 20$ and 25000 training points

5 Conclusion & Perspectives

In this work we considered a extension of the f/P-Greedy algorithm [18] to the vectorial case in the spirit of [13]. The question about the behaviour of the gain function close to already included points has been answered satisfactory and it turned out to be directly related to Hermite interpolation at the repetitively considered points. Moreover, the established Temlyakov-type convergence rates for vectorial greedy algorithms from e.g. [6, 14] could be verified and improved. However, as mentioned already in the convergence analysis, it remains an open question if the discrepancy between the observed and predicted convergence rates can be reduced in future work. The obtained convergence rates for RKHS using fill distances [18, 38] promise room for improvement, standing against some results on lower bounds for the convergence rates [14]. We pursued a comparison of the proposed algorithm to a related existing one and discussed both advantages and disadvantages and their possible remedies.

Future work comprises formulations of the considered algorithms for stable bases, e.g. RBF-QR [9**?**] or other [20]. Even though our conducted experiments have been of a synthetic nature, we are currently also investigating applications of the proposed algorithm and related ones in practical applications [40].

6 Acknowledgements

The authors would like to thank the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart and the Baden-Württemberg Stiftung gGmbH.

We would also like to thank R. Schaback for a fruitful discussion and comments regarding relations to existing work during the DWCAA conference 2012.

References

1. S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2):302 –309, mar 1991.
2. S. Chen, P.M. Grant, and C.F.N. Cowan. Orthogonal least-squares algorithm for training multi-output radial basis function networks. *Radar and Signal Processing, IEE Proceedings F*, 139(6):378 –384, dec 1992.
3. S. Chen and J. Wigger. Fast orthogonal least squares algorithm for efficient subset model selection. *Signal Processing, IEEE Transactions on*, 43(7):1713 –1715, jul 1995.
4. G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximation*, 13:57–98, 1997.
5. S. De Marchi, R. Schaback, and H. Wendland. Near-optimal data-independent point locations for radial basis function interpolation. *Advances in Computational Mathematics*, 23:317–330, 2005.
6. R. DeVore and V. Temlyakov. Some remarks on greedy algorithms. *Advances in Computational Mathematics*, 5:173–187, 1996.
7. R. A. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.
8. G. E. Fasshauer. Positive defnite kernels: Past, present and future. Notes, Department of Applied Mathematics, Illinois Institute of Technology, 2011. Extended notes on presentation at the "Workshop on Kernel Functions and Meshless Methods honoring the 65th birthday of Robert Schaback".
9. B. Fornberg, E. Larsson, and N. Flyer. Stable computations with gaussian radial basis functions. *SIAM J. Sci. Comput.*, 33(2):869–892, April 2011.
10. B. Fornberg and C. Piret. A stable algorithm for flat radial basis functions on a sphere. *SIAM J. Sci. Comput.*, 30(1):60–80, Oct 2007.
11. S. Konyagin and V. Temlyakov. Greedy approximation with regard to bases and general minimal systems. *Serdica Mathematical Journal*, 28(4):305–328, 2002. ISSN: 1310-6600.
12. D. Leviatan and V. Temlyakov. Simultaneous approximation by greedy algorithms. *Advances in Computational Mathematics*, 25:73–90, 2006.
13. D. Leviatan and V.N. Temlyakov. Simultaneous greedy approximation in banach spaces. *Journal of Complexity*, 21(3):275 – 293, 2005.

14. A. Lutoborski and V. N. Temlyakov. Vector greedy algorithms. *Journal of Complexity*, 19(4):458 – 473, 2003.
15. S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397 –3415, dec 1993.
16. M. Mouattamid. *Theory of Power Kernels*. Dissertation, Universität Göttingen, 2005.
17. M. Mouattamid and R. Schaback. Recursive kernels. *Analysis in Theory and Applications*, 25(4):301–316, 2009.
18. S. Müller. *Complexity and Stability of Kernel-based Reconstructions*. Dissertation, Georg-August-Universität Göttingen, Institut für Numerische und Angewandte Mathematik, Lotzestrasse 16-18, D-37083 Göttingen, Jan 2009. Göttinger Online Klassifikation: EDDF 050.
19. Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 40 – 44, Nov 1993.
20. M. Pazouki and R. Schaback. Bases for kernel-based spaces. *J. Comp. Appl. Math.*, 236(4):575 – 588, 2011. International Workshop on Multivariate Approximation and Interpolation with Applications (MAIA 2010).
21. J. Phillips, J. Afonso, A. Oliveira, and L.M. Silveira. Analog macromodeling using kernel methods. In *Proc. of ICCAD 2003*, pages 446–453, November 2003.
22. C. Rieger, R. Schaback, and B. Zwicknagl. Sampling and stability. In *Mathematical Methods for Curves and Surfaces*, volume 5862 of *Lecture Notes in Computer Science*, pages 347–369. Springer Berlin Heidelberg, 2010.
23. R. Schaback. Reconstruction of multivariate functions from scattered data, 1997.
24. R. Schaback. Limit problems for interpolation by analytic radial basis functions. *J. Comp. Appl. Math.*, 212(2):127 – 149, 2008.
25. R. Schaback and H. Wendland. Adaptive greedy techniques for approximate solution of large rbf systems. *Numerical Algorithms*, 24:239–254, 2000.
26. R. Schaback and H. Wendland. Kernel techniques: From machine learning to meshless methods. *Acta Numerica*, 15:543–639, May 2006.
27. R. Schaback and J. Werner. Linearly constrained reconstruction of functions by kernels with applications to machine learning. *Advances in Computational Mathematics*, 25:237–258, 2006. 10.1007/s10444-004-7616-1.
28. E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Mathematische Annalen*, 63:433–476, 1907.
29. B. Schölkopf and A. J. Smola. *Learning with Kernels*. Adaptive Computation and Machine Learning. The MIT Press, 2002.
30. J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004. ISBN: 0521813972.
31. K. Slavakis, P. Bouboulis, and S. Theodoridis. Adaptive multiregression in reproducing kernel hilbert spaces: The multiaccess mimo channel case. *Neural Networks and Learning Systems, IEEE Transactions on*, 23(2):260 –276, feb. 2012.
32. I. Steinwart and A. Christman. *Support Vector Machines*. Science + Business Media. Springer, 2008.
33. V. N. Temlyakov. Greedy approximation. *Acta Numerica*, 17:235–409, 2008.
34. V.N. Temlyakov. The best m-term approximation and greedy algorithms. *Advances in Computational Mathematics*, 8:249–265, 1998.
35. V.N. Temlyakov. Weak greedy algorithms. *Advances in Computational Mathematics*, 12:213–227, 2000.
36. J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Processing*, 86(3):572 – 588, 2006.
37. P. Vincent and Y. Bengio. Kernel matching pursuit. *Machine Learning*, 48:165–187, 2002.
38. H. Wendland. *Scattered data approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, 2005.
39. D. Wirtz and B. Haasdonk. Efficient a-posteriori error estimation for nonlinear kernel-based reduced systems. *Systems and Control Letters*, 61(1):203 – 211, 2012.

40. D. Wirtz, N. Karajan, and B. Haasdonk. Model order reduction of multiscale models using kernel methods. Preprint, SRC SimTech, University of Stuttgart, Germany, 2013. In preparation.