



Analysis of Target Data-Dependent Greedy Kernel Algorithms: Convergence Rates for f -, $f \cdot P$ - and f/P -Greedy

Tizian Wenzel¹ · Gabriele Santin²  · Bernard Haasdonk¹

Received: 17 May 2021 / Revised: 24 March 2022 / Accepted: 5 April 2022 /

Published online: 18 October 2022

© The Author(s) 2022

Abstract

Data-dependent greedy algorithms in kernel spaces are known to provide fast converging interpolants, while being extremely easy to implement and efficient to run. Despite this experimental evidence, no detailed theory has yet been presented. This situation is unsatisfactory, especially when compared to the case of the data-independent P -greedy algorithm, for which optimal convergence rates are available, despite its performances being usually inferior to the ones of target data-dependent algorithms. In this work, we fill this gap by first defining a new scale of greedy algorithms for interpolation that comprises all the existing ones in a unique analysis, where the degree of dependency of the selection criterion on the functional data is quantified by a real parameter. We then prove new convergence rates where this degree is taken into account, and we show that, possibly up to a logarithmic factor, target data-dependent selection strategies provide faster convergence. In particular, for the first time we obtain convergence rates for target data adaptive interpolation that are faster than the ones given by uniform points, without the need of any special assumption on the target function. These results are made possible by refining an earlier analysis of greedy algorithms in general Hilbert spaces. The rates are confirmed by a number of numerical examples.

Keywords Kernel methods · Greedy algorithms · Convergence rates · Target data-dependent algorithms

✉ Tizian Wenzel
tizian.wenzel@mathematik.uni-stuttgart.de

Gabriele Santin
gsantin@fbk.eu

Bernard Haasdonk
haasdonk@mathematik.uni-stuttgart.de

¹ Institute for Applied Analysis and Numerical Simulation, University of Stuttgart, Stuttgart, Germany

² Digital Society Center, Bruno Kessler Foundation, Trento, Italy

Mathematics Subject Classification Primary 65D15 · 65D05 · Secondary 41A58 · 46E22

1 Introduction

Kernel methods are a well-understood and widely used technique for approximation, regression and classification in machine learning and numerical analysis.

We start by collecting some notation and preliminary results, while more details are provided in Sect. 2. For a non-empty set Ω a kernel is defined as a symmetric function $k : \Omega \times \Omega \rightarrow \mathbb{R}$. The kernel matrix A_{X_n} for a set of points $X_n = \{x_1, \dots, x_n\} \subset \Omega$ is given as $(A_{X_n})_{ij} = (k(x_i, x_j))_{ij} \in \mathbb{R}^{n \times n}$, $i, j = 1, \dots, n$. If the kernel matrix is strictly positive definite for any set $X_n \subset \Omega$ of n distinct points, the kernel is called *strictly positive definite*. Associated to every strictly positive definite kernel there is a unique *Reproducing Kernel Hilbert Space* $\mathcal{H}_k(\Omega)$ (RKHS) with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_k(\Omega)}$, which is also called *native space* of k , and which is a space of real valued functions on Ω where the kernel k acts as a *reproducing kernel*, that is

1. $k(\cdot, x) \in \mathcal{H}_k(\Omega) \forall x \in \Omega$,
2. $f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k(\Omega)} \forall x \in \Omega, \forall f \in \mathcal{H}_k(\Omega)$ (reproducing property).

Strictly positive definite continuous kernels can be used for the interpolation of continuous functions. The theory is developed under the assumption that $f \in \mathcal{H}_k(\Omega)$, and in this case for any set of pairwise distinct interpolation points $X_n \subset \Omega$ there exists a unique minimum-norm interpolant $s_n \in \mathcal{H}_k(\Omega)$ that satisfies

$$s_n(x_i) = f(x_i) \quad \forall i = 1, \dots, n. \quad (1)$$

It can be shown that this interpolant is given by the orthogonal projection $\Pi_{V(X_n)}(f)$ of f onto the linear subspace $V(X_n) := \text{span}\{k(\cdot, x_i), x_i \in X_n\}$, i.e.,

$$s_n(\cdot) = \Pi_{V(X_n)}(f) = \sum_{j=1}^n \alpha_j k(\cdot, x_j).$$

The coefficients α_j , $j = 1, \dots, n$, can be calculated by solving the linear system of equations arising from the interpolation conditions in Eq. (1), which is always invertible due to the assumed strict positive definiteness of the kernel.

A standard way of estimating the error between the function f and the interpolant in the $\|\cdot\|_{L^\infty(\Omega)}$ -norm makes use of the *power function*, which is given as

$$\begin{aligned} P_{X_n}(x) &:= \|k(\cdot, x) - \Pi_{V(X_n)}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)} \\ &= \sup_{0 \neq \tilde{f} \in \mathcal{H}_k(\Omega)} \frac{|(\tilde{f} - \Pi_{V(X_n)}(\tilde{f}))(x)|}{\|\tilde{f}\|_{\mathcal{H}_k(\Omega)}}. \end{aligned} \quad (2)$$

Obviously it holds $P_{X_n}(x_i) = 0$ for all $i = 1, \dots, n$, and the standard power function estimate bounds the interpolation error as

$$\begin{aligned} |(f - s_n)(x)| &\leq P_{X_n}(x) \cdot \|f - s_n\|_{\mathcal{H}_k(\Omega)} \\ &= P_{X_n}(x) \cdot \|r_n\|_{\mathcal{H}_k(\Omega)} \quad \forall x \in \Omega, \end{aligned} \tag{3}$$

where we denoted the residual as $r_n := r_n(f) := f - s_n$.

Observe that any worst-case error bound on $|(f - \Pi_{V(X_n)}(f))(x)|$ over the entire $\mathcal{H}_k(\Omega)$ transfers to the same decay of the power function via the second equality in Eq. (2). For the large class of translational invariant kernels, that we will introduce below and that includes the notable class of radial basis function (RBF) kernels, it is possible to refine this error estimate by bounding the decay of the power function in terms of the fill distance

$$h_{X_n} := h_{X_n, \Omega} := \sup_{x \in \Omega} \min_{x_j \in X_n} \|x - x_j\|_2.$$

Depending on certain properties of the kernel, one may obtain in this way both algebraic and exponential rates in terms of h_{X_n} . Especially in the case of kernels whose RKHS is norm equivalent to a Sobolev space, these algebraic rates are provably quasi-optimal and may even be extended to certain functions that are outside of $\mathcal{H}_k(\Omega)$ (see [17]).

These results are nevertheless bounded by the dependence on the filling of the space and by the independence on the target function f . Namely, the fill distance is at most decaying as $h_{X_n, \Omega} \asymp c_\Omega n^{-1/d}$ for *quasi-uniform* points, which are space-filling and target-independent. On the other hand, a global target-dependent optimization of the interpolation points is a combinatorial and practically infeasible task, and thus approximated strategies have been proposed, and in particular greedy algorithms.

Greedy algorithms in general are studied in various branches of mathematics, and we point to [29] for a general treatment of their use in approximation. In kernel interpolation, a greedy algorithm starts with the empty set $X_0 := \emptyset$ and adds points incrementally as $X_{n+1} := X_n \cup \{x_{n+1}\}$ according to some selection criterion $\eta^{(n)}$, that is

$$x_{n+1} := \operatorname{argmax}_{x \in \Omega \setminus X_n} \eta^{(n)}(x).$$

Commonly used selection criteria in the greedy kernel literature are the P -greedy [3], $f \cdot P$ -greedy¹ [6], f -greedy [26], and f/P -greedy [14] criteria, and they choose the next point according to the following strategies. From now on, we use the short-hand notation $P_n(\cdot) := P_{X_n}(\cdot)$ whenever the power function is determined by some greedy algorithm.

- i. P -greedy: $\eta_P^{(n)}(x) = P_n(x)$,
- ii. $f \cdot P$ -greedy: $\eta_{f \cdot P}^{(n)}(x) = |r_n(x)| \cdot P_n(x)$,
- iii. f -greedy: $\eta_f^{(n)}(x) = |r_n(x)|$,

¹ We remark that we use the notation $f \cdot P$ -greedy algorithm here because it fits better to our notation, while in the original publication it was called *psr*-greedy (power scaled residual greedy).

iv. f/P -greedy: $\eta_{f/P}^{(n)}(x) = |r_n(x)|/P_n(x)$.

These algorithms have been used in a series of applications (see, e.g., [6, 9–11, 18, 21, 26–28]), and overwhelming numerical evidence points to the fact that criteria which incorporate a residual-dependent term provide faster convergence, even if sometimes at the price of stability (see [32] for a discussion of this fact for f/P -greedy, and [6] for $f \cdot P$ -greedy).

The faster convergence is fully understandable since function adaptivity should clearly be beneficial to convergence speed. Nevertheless, the theoretical results are of opposite nature. Namely, for the P -greedy algorithm it is possible to prove quasi-optimality statements (see [20]), namely that whatever is the best known decay rate of the power function for arbitrarily optimized points, this transfers to the same decay of the power function associated to the points selected by P -greedy. Especially in the case of Sobolev spaces, these results can be proven to be optimal [32]. On the other hand, the convergence theory for the target data-dependent algorithms is much weaker: The known results (see Sect. 2 for a detailed account) provide convergence of order at most $n^{-1/2}$, which is generally not only largely missing the practical observations, but also slower than the rates proven for P -greedy.

We remark that existing techniques to prove convergence of greedy algorithms in general Hilbert spaces are not directly transferable to this setting. Indeed, the first results on similar algorithms have been obtained in Matching Pursuit, and they work for finite dimensional spaces [2, 13]. When transferred to the kernel setting (see [26]) these require a norm equivalence between the $\mathcal{H}_k(\Omega)$ - and the ∞ -norm, which hold only for finite n . Subsequent general results on greedy algorithms (see [5]) require special assumptions on the target function, and the resulting rates are only of order $n^{-1/2}$. Another common strategy in the greedy literature makes use of the Restricted Isometry Property (see, e.g., [1]), which in the kernel setting translates to the requirement that the smallest eigenvalue λ_n of the kernel matrix is bounded away from zero uniformly in n . This is not the case here, since it is known that $\lambda_n \leq \min_{1 \leq j \leq n} P_{X_n \setminus \{x_j\}}(x_j)^2$ (see [23]), and we will see later that a fast convergence to zero of the right hand side of this inequality is the key of our analysis. Especially, all these results prove convergence in the Hilbert space norm, which is generally too strong (to obtain convergence rates) since the interpolation operator is an orthogonal projection in $\mathcal{H}_k(\Omega)$. We work instead with the ∞ -norm, which allows to derive fast convergence, even if it introduces an additional difficulty since the norm of the error is not monotonically decreasing. Furthermore we point to the empirical interpolation method (EIM) [12], which is also a greedy technique, that however aims at interpolating a set of functions (instead of a single function) using a subset of these functions as basis elements (instead of kernel evaluations $k(\cdot, x)$).

The paper is organized as follows. After recalling additional facts on kernel greedy interpolation in Sect. 2, we derive a new analysis of general greedy algorithms in general Hilbert spaces based on [4] (Sect. 3).

In Sect. 4 we frame the four selection rules into a joint scale of greedy algorithms by introducing β -greedy algorithms (Definition 4) which include P -greedy ($\beta = 0$), $f \cdot P$ -greedy ($\beta = 1/2$), f -greedy ($\beta = 1$), and f/P -greedy ($\beta = \infty$), and we study them within a novel error analysis.

These results are combined in Sect. 5 to obtain precise convergence rates of the minimum error $e_{\min}(f, n) := \min_{1 \leq i \leq n} \|f - s_i\|_{\infty}$. This measure allows us to circumvent the non-monotonicity of the error, and we remark in particular that $e_{\min}(f, n) < \varepsilon$ for some $\varepsilon > 0$ means that an error smaller than ε is achieved using *at most* n points. As an exemplary result, we mention here the case where the rate of worst-case convergence in $\mathcal{H}_k(\Omega)$ for a fixed set of n interpolation points is $n^{-\alpha}$ for a given $\alpha > 0$. In this case, for $\beta \in [0, 1]$ we get new convergence rates of the form

$$e_{\min}(f, n) \leq c \log(n)^{\alpha} n^{-\beta/2} n^{-\alpha}, \quad n \geq n_0 \in \mathbb{N},$$

with $c > 0$. These results prove in particular that the worst case decay of the error that can be obtained in $\mathcal{H}_k(\Omega)$ with a fixed sequence of points transfers to the β -greedy algorithms with an additional multiplicative factor of $\log(n)^{\alpha} n^{-\beta/2}$. Namely, adaptively selected points provide faster convergence than any fixed set of points.

Finally, Sect. 6 illustrates the results with analytical and numerical examples while the final Sect. 7 presents the conclusion and gives an outlook.

2 Background Results on Kernel Interpolation

We recall additional required background information on kernel based approximation and in particular greedy kernel interpolation. For a more detailed overview we refer the reader to [7, 8, 30]. We remark that in this section no special attention is paid to the occurring constants, which can change from line to line.

2.1 Interpolation by Translational Invariant Kernels

In many applications of interest, the domain is a subset of the Euclidean space, i.e., $\Omega \subset \mathbb{R}^d$. In this case, a special kind of kernels is given by *translational invariant kernels*, i.e., there exists a function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ with a continuous Fourier transform $\hat{\Phi}$ such that

$$k(x, y) = \Phi(x - y) \text{ for all } x, y \in \mathbb{R}^d.$$

We remark that the well-known radial basis function kernels are a particular instance of translational invariant kernels.

Depending on the decay of the Fourier transform of the function Φ , two classes of translational invariant kernels can be distinguished:

1. We call the kernel k a kernel of *finite smoothness* $\tau > d/2$, if there exist constants $c_{\Phi}, C_{\Phi} > 0$ such that

$$c_{\Phi}(1 + \|\omega\|_2^2)^{-\tau} \leq \hat{\Phi}(\omega) \leq C_{\Phi}(1 + \|\omega\|_2^2)^{-\tau}.$$

The assumption $\tau > d/2$ is required in order to have a Sobolev embedding in $C^0(\Omega)$.

2. If the Fourier transform $\hat{\Phi}$ decays faster than at any polynomial rate, the kernel is called *infinitely smooth*.

As mentioned in Sect. 1, for these two types of kernels it is possible to derive error estimates by bounding the decay of the power function in terms of the fill distance. We have the following:

1. For kernels of finite smoothness $\tau > d/2$, given appropriate conditions on the domain $\Omega \subset \mathbb{R}^d$ (e.g., Lipschitz boundary and interior cone condition), the native space $\mathcal{H}_k(\Omega)$ is norm equivalent to the Sobolev space $W_2^\tau(\Omega)$. By making use of this connection, error estimates for kernel interpolation can be obtained by using Sobolev bounds [16, 31] that give

$$\|P_{X_n}\|_{L^\infty(\Omega)} \leq \hat{c}_1 h_{X_n}^{\tau-d/2}. \quad (4)$$

2. For kernels of infinite smoothness such as the Gaussian, the multiquadric or the inverse multiquadric, we have

$$\|P_{X_n}\|_{L^\infty(\Omega)} \leq \hat{c}_2 \exp(-\hat{c}_3 h_{X_n}^{-1}), \quad (5)$$

if the domain Ω is a cube. We remark that these error estimates are not limited to these three exemplary kernels. We point to [30, Theorem 11.22] which states a sufficient condition in order to obtain these exponential kind of error estimates.

By looking at well-distributed points such that $h_{X_n, \Omega} \leq c_\Omega n^{-1/d}$, these bounds from Eqs. (4) and (5) can be cast only in terms of the number of interpolation points n , i.e.

$$\begin{aligned} \|P_{X_n}\|_{L^\infty(\Omega)} &\leq \tilde{c}_1 n^{1/2-\tau/d}, \\ \|P_{X_n}\|_{L^\infty(\Omega)} &\leq \tilde{c}_2 \exp(-\tilde{c}_3 n^{1/d}). \end{aligned} \quad (6)$$

2.2 Greedy Kernel Interpolation

We collect the motivation, a few properties, and the existing analysis of the four selection criteria introduced in Sect. 1:

- i. *P-greedy*: The *P-greedy* algorithm is the best analyzed one of the four algorithms named above. It aims at minimizing the error for all functions in the native space simultaneously, which is done by greedily minimizing the upper error bound from Eq. (3), which is the power function. Thus, the selection criterion of the *P-greedy* algorithm is target data independent. For the *P-greedy* algorithm, it holds $P_n(x_{n+1}) = \|P_n\|_{L^\infty(\Omega)}$. Several results on the *P-greedy* algorithm have been derived in [20, 32]:
- (a) Corollary 2.2. in [20] showed convergence statements for the maximal power function value $\|P_n\|_{L^\infty(\Omega)}$ for radial basis function kernels, when $\Omega \subset \mathbb{R}^d$ has a Lipschitz boundary and satisfies an interior cone condition. It states

$$\|P_n\|_{L^\infty(\Omega)} \leq c_1 \cdot n^{1/2-\tau/d} \quad (\text{finite smoothness } \tau > d/2)$$

$$\|P_n\|_{L^\infty(\Omega)} \leq c_2 \exp(-c_3 n^{1/d}) \quad (\text{infinite smoothness}).$$

Via the standard power function bound from Eq. (3) these bounds directly give bounds on the approximation $\|f - s_n\|_{L^\infty(\Omega)}$. A few more details of the proof strategy of [20] will be recalled in Sect. 3.

- (a) The paper [32] showed further results for the case of kernels of finite smoothness $\tau > d/2$: Theorem 12 in [32] showed that the decay rate on $\|P_n\|_{L^\infty(\Omega)}$ is sharp. The sequence of Theorems 15, 19 and 20 of [32] further established that the resulting sequence of points are asymptotically uniformly distributed under some mild conditions. These results implied (optimal) stability statements in [32, Corollary 22].
- ii. *f*-greedy: The *f*-greedy algorithm aims at directly minimizing the residual by setting the currently largest residual to zero by introducing the next interpolation point at this point, i.e. it holds $|(f - s_n)(x_{n+1})| = \|f - s_n\|_{L^\infty(\Omega)}$. Existing results prove convergence of order $n^{-\ell/d}$ for kernels $k \in C^{2\ell}(\Omega \times \Omega)$ in $d = 1$ (see Section 3.4 in [14]), while for general d limited results are known, e.g., [14, Korollar 3.3.8] states that

$$\min_{j=1,\dots,n} \|f - s_j\|_{L^\infty(\Omega)} \leq Cn^{-1/d}$$

if $k \in C^2(\Omega \times \Omega)$. As mentioned before, these convergence results do not reflect the approximation speed of *f*-greedy that can be observed in numerical investigations. Additionally, in [22] convergence of order $n^{-1/2}$ of the $\mathcal{H}_k(\Omega)$ -norm of the error is proven, but only under additional assumptions on *f*.

- iii. *f*/*P*-greedy: The *f*/*P*-greedy selection aims at minimizing the native space error of the residual as much as possible as it can be seen from Eq. (7). We remark as a technical detail that the supremum of $|(f - s_n)(x)|/P_n(x)$ over $x \in \Omega \setminus X_n$ need not be attained as it was exemplified in Example 6 of [32]. However, this can be alleviated by choosing the next point x_{n+1} such that $\frac{|r_n(x_{n+1})|}{P_n(x_{n+1})} \geq (1 - \epsilon) \cdot \sup_{x \in \Omega \setminus X_n} \frac{|r_n(x)|}{P_n(x)}$ for any $0 < \epsilon \ll 1$. As a convergence result, [33, Theorem 3] states

$$\|f - s_n\|_{\mathcal{H}_k(\Omega)} \leq Cn^{-1/2},$$

which, however, only holds for a quite restricted set of functions *f*, which has slightly been extended in [22].

- iv. *f* · *P*-greedy: The idea of the just recently introduced *f* · *P*-greedy algorithm is to have a combination of the power function dependence and the target data dependence in order to balance between the stability of the *P*-greedy algorithm and the target data dependence of the *f*-greedy algorithm. No convergence results were given in the original publication [6].

In addition to the selection criteria, we remark that for a practical numerical implementation the greedy algorithms stop if a predefined bound (either on, e.g., the accuracy or the numerical stability) is reached, or if the interpolant is exact.

Finally, to analyze and implement these algorithms it is useful to consider the *Newton basis* $\{v_j\}_{j=1}^n$ of V_n (see [15, 18]), which is obtained by applying the Gram-Schmidt orthonormalization process to $\{k(\cdot, x_j), j = 1, \dots, n\}$ whereby $\{x_j, j = 1, \dots, n\}$ are the pairwise distinct points that are incrementally selected by the greedy procedure. We recall that we have

$$s_n(x) = \sum_{j=1}^n \langle f, v_j \rangle_{\mathcal{H}_k(\Omega)} v_j(x),$$

and it can be shown that it holds $\langle f, v_j \rangle_{\mathcal{H}_k(\Omega)} = |(f - s_{j-1})(x_j)| / P_{X_{j-1}}(x_j)$. If $s_n \xrightarrow{n \rightarrow \infty} f$ in $\mathcal{H}_k(\Omega)$, we have

$$\|f\|_{\mathcal{H}_k(\Omega)}^2 = \sum_{j=1}^{\infty} \left(\frac{|(f - s_j)(x_{j+1})|}{P_{X_j}(x_{j+1})} \right)^2. \quad (7)$$

3 Analysis of Greedy Algorithms in an Abstract Setting

This section extends the abstract analysis of greedy algorithms in Hilbert spaces introduced in [4]. For this, let \mathcal{H} be a Hilbert space with norm $\|\cdot\| = \|\cdot\|_{\mathcal{H}}$. Let $\mathcal{F} \subset \mathcal{H}$ be a compact subset and assume for notational convenience only that it holds $\|f\|_{\mathcal{H}} \leq 1$ for all $f \in \mathcal{F}$.

We consider algorithms that select elements f_0, f_1, \dots , without yet specifying any particular selection criterion. We define $V_n := \text{span}\{f_0, \dots, f_{n-1}\}$ and the following quantities, whereby Y_n is any n -dimensional subspace of \mathcal{H} :

$$\begin{aligned} d_n &:= d_n(\mathcal{F})_{\mathcal{H}} := \inf_{Y_n \subset \mathcal{H}} \sup_{f \in \mathcal{F}} \text{dist}(f, Y_n)_{\mathcal{H}} \\ \sigma_n &:= \sigma_n(\mathcal{F})_{\mathcal{H}} := \sup_{f \in \mathcal{F}} \text{dist}(f, V_n)_{\mathcal{H}} \\ v_n &:= \text{dist}(f_n, V_n)_{\mathcal{H}}. \end{aligned} \quad (8)$$

The quantities d_n and σ_n have already been used in [4], where d_n is the Kolmogorov n -width of \mathcal{F} , and we recall that the compactness of \mathcal{F} is equivalent to require that $\lim_n d_n = 0$ (see [19]). On the other hand, the newly introduced quantity v_n does not seem in itself to be an interesting quantity for the abstract setting, and it was only denoted as $a_{n,n}$ within [4] before. However, it will be the key quantity for our new analysis in the kernel setting in Sects. 4 and 5.

As we focus on Hilbert spaces, expressions like (\cdot, V_n) can be computed via the orthogonal projector in \mathcal{H} onto V_n , that we denote as Π_{V_n} . We have the following elementary properties:

1. Estimates: $d_n \leq \sigma_n$ and $v_n \leq \sigma_n$ for all $n \in \mathbb{N}$.
2. Monotonicity: $(\sigma_n)_{n \in \mathbb{N}}$ and $(d_n)_{n \in \mathbb{N}}$ are monotonically decreasing.
3. Initial value: $d_0 \leq \sigma_0 \leq 1$.

The paper [4] considers *weak* greedy algorithms that choose, for some fixed $0 < \gamma \leq 1$, the elements f_n such that

$$v_n \equiv \text{dist}(f_n, V_n)_{\mathcal{H}} \equiv \sigma_n(\{f_n\})_{\mathcal{H}} \geq \gamma \cdot \sup_{f \in \mathcal{F}} \sigma_n(\{f\})_{\mathcal{H}} = \gamma \cdot \sigma_n(\mathcal{F}), \tag{9}$$

and shows that, roughly speaking, an asymptotic polynomial or exponential decay of d_n yields a polynomial or exponential decay of σ_n , i.e., the weak greedy algorithms essentially realize the Kolmogorov widths up to multiplicative constants. We remark that this analysis includes the *strong* greedy algorithm, i.e., $\gamma = 1$.

In the following, we show in Sect. 3.1 that even without using the selection of Eq. (9)—i.e., the elements f_0, f_1, \dots may even be randomly chosen within \mathcal{F} —comparable statements hold for v_n .

3.1 Greedy Approximation with Arbitrary Selection Rules

We start by stating a simple modification of [4, Theorem 3.2.] and a subsequent corollary. The theorem is actually valid for any sequence $\{f_i\}_i \subset \mathcal{F}$, but since we are interested in greedy algorithms we phrase the result by assuming that the f_i are selected in terms of an arbitrary selection rule.

Theorem 1 *Consider a compact set \mathcal{F} in a Hilbert space \mathcal{H} , and a greedy algorithm that selects elements from \mathcal{F} according to any arbitrary selection rule.*

We have the following inequalities between v_n, σ_n and d_n for any $N \geq 0, K \geq 1, 1 \leq m < K$:

$$\prod_{i=1}^K v_{N+i}^2 \leq \left(\frac{K}{m}\right)^m \left(\frac{K}{K-m}\right)^{K-m} \sigma_{N+1}^{2m} d_m^{2K-2m}.$$

Proof The result is obtained by simply omitting the last step in the proof of Theorem 3.2 in [4]. Namely, any element in the sequence of selected functions can be represented by its coefficients representation on a certain orthonormal basis, obtained by a Gram–Schmidt orthonormalization process on the previously selected functions. These coefficients are collected into an infinite dimensional matrix (see Section 3 in [4]). It is possible to apply Lemma 2.1 in [4], in order to obtain the two bounds (3.2) and (3.3) in [4]. We follow the original proof up to this point, i.e., right before Eq. (3.4), which is the bound on the quantity $a_{N+i, N+i}^2$. Using the second-to-last equation on p. 459 in [4] and our definition of v_n , in our notation we have

$$a_{N+i, N+i}^2 = \|f_{N+i} - \Pi_{V_{N+i}} f_{N+i}\|_{\mathcal{H}}^2 = \text{dist}(f_{N+i}, V_{N+i})_{\mathcal{H}}^2 = v_{N+i}^2,$$

and this gives the result. In the original paper, an additional step in Eq. (3.4) is used to obtain a bound on σ_n instead of v_n . □

Similarly to the approach used in [4], in the following corollary we make suitable choices of N, K, m to specialize the result to the case of algebraically or exponentially decaying Kolmogorov widths.

Corollary 2 *Under the assumptions of Theorem 1 the following holds.*

(i) *If $d_n(\mathcal{F}) \leq C_0 n^{-\alpha}$, $n \geq 1$, then it holds*

$$\left(\prod_{i=n+1}^{2n} v_i \right)^{1/n} \leq 2^{\alpha+1/2} \tilde{C}_0 e^\alpha \log(n)^\alpha n^{-\alpha}, \quad n \geq 3, \tag{10}$$

with $\tilde{C}_0 := \max\{1, C_0\}$.

(ii) *If $d_n(\mathcal{F}) \leq C_0 e^{-c_0 n^\alpha}$, $n = 1, 2, \dots$, then it holds*

$$\left(\prod_{i=n+1}^{2n} v_i \right)^{1/n} \leq \sqrt{2\tilde{C}_0} \cdot e^{-c_1 n^\alpha}, \quad n \geq 2, \tag{11}$$

with $\tilde{C}_0 := \max\{1, C_0\}$ and $c_1 = 2^{-(2+\alpha)} c_0 < c_0$.

Proof First of all we observe that for $1 \leq m < n$, we have $0 < x := m/n < 1$. Using $x^{-x}(1-x)^{x-1} \leq 2$ for $x \in (0, 1)$ we obtain

$$\left[\binom{n}{m}^m \left(\frac{n}{n-m} \right)^{n-m} \right]^{1/n} = x^{-x}(1-x)^{x-1} \leq 2.$$

We use Theorem 1 for $N = K = n$ and any $1 \leq m < n$, i.e. we have

$$\begin{aligned} \prod_{i=1}^n v_{n+i}^2 &\leq \binom{n}{m}^m \left(\frac{n}{n-m} \right)^{n-m} \sigma_{n+1}^{2m} d_m^{2n-2m} \\ \Rightarrow \left(\prod_{i=1}^n v_{n+i} \right)^{1/n} &\leq \left[\binom{n}{m}^m \left(\frac{n}{n-m} \right)^{n-m} \right]^{1/2n} \sigma_{n+1}^{m/n} d_m^{(n-m)/n} \\ &\leq \sqrt{2} \sigma_{n+1}^{m/n} d_m^{(n-m)/n} \leq \sqrt{2} \cdot d_m^{(n-m)/n}, \end{aligned} \tag{12}$$

where we took the $2n$ -th root for the second line and used the monotonicity and boundedness of $(\sigma_n)_{n \in \mathbb{N}}$ in the last step, i.e. $\sigma_{n+1}^{m/n} \leq \sigma_1^{m/n} \leq 1$.

In order to prove the statements (i) and (ii), we conclude now in two different ways:

(i) For n fixed we choose a fixed $0 < \omega \ll 1$ and define $m^* := \lceil \omega n \rceil \in \mathbb{N}$, i.e. $\omega n \leq m^* < \omega n + 1$. Using $d_n \leq 1$, $d_n \leq \tilde{C}_0 n^{-\alpha}$ with $\tilde{C}_0 := \max\{1, C_0\}$, and since d_n is non-increasing, we can estimate:

$$\begin{aligned} \left(\prod_{i=1}^n v_{n+i} \right)^{1/n} &\leq \sqrt{2} \cdot d_{m^*}^{(n-m^*)/n} \leq \sqrt{2} \cdot d_{\lceil \omega n \rceil}^{(n-\omega n-1)/n} \\ &\leq \sqrt{2} \tilde{C}_0^{(1-\omega)-1/n} \lceil \omega n \rceil^{-\alpha(1-\omega)+\alpha/n} \end{aligned}$$

$$\begin{aligned} &\leq \sqrt{2}\tilde{C}_0^{(1-\omega)-1/n}(\omega n)^{-\alpha(1-\omega)+\alpha/n} \\ &\leq \sqrt{2}\tilde{C}_0^{(1-\omega)}\omega^{-\alpha(1-\omega)}n^{-\alpha(1-\omega)}(n^{1/n})^\alpha \\ &\leq \sqrt{2}\tilde{C}_0^{(1-\omega)}\omega^{-\alpha(1-\omega)}n^{-\alpha(1-\omega)}2^\alpha. \end{aligned}$$

It follows that for each $\omega \in (0, 1)$ it holds that

$$\left(\prod_{i=n+1}^{2n} v_i\right)^{1/n} \leq 2^{\alpha+1/2}\tilde{C}_0 n^{-\alpha} C(\omega, n), \tag{13}$$

with $C(\omega, n) = \tilde{C}_0^{-\omega}\omega^{-\alpha(1-\omega)}n^{\alpha\omega}$. For each n , the inequality holds in particular for an optimally chosen $\bar{\omega} := \bar{\omega}(n)$ in $(0, 1)$. To find a good candidate $\bar{\omega}$ we minimize the upper bound $\tilde{C}(\omega, n) := \omega^{-\alpha}n^{\alpha\omega}$, which satisfies $C(\omega, n) \leq \tilde{C}(\omega, n)$ since $\omega, \alpha \geq 0$ and $\tilde{C}_0 \geq 1$. It holds

$$\partial_\omega \tilde{C}(\omega, n) = n^{\alpha\omega}\alpha\omega^{-1-\alpha}(-1 + \omega \log(n)),$$

which vanishes in $\bar{\omega} = 1/\log(n)$, and it is negative on the left of this value and positive on the right. It follows that if $\bar{\omega} \in (0, 1)$, i.e., $n \geq 3$, then we can choose the constant $C(\bar{\omega}, n)$ in Eq. (13), which gives the statement since

$$C(\bar{\omega}, n) \leq \tilde{C}(\bar{\omega}, n) = \log(n)^\alpha n^{\alpha/\log(n)} = e^\alpha \log(n)^\alpha.$$

- (i) We pick $m = \lceil n/2 \rceil$ and make use of the assumed decay $d_n(\mathcal{F}) \leq \tilde{C}_0 e^{-c_0 n^\alpha}$ to estimate

$$\begin{aligned} \left(\prod_{i=1}^n v_{n+i}\right)^{1/n} &\leq \sqrt{2} \cdot d_m^{(n-m)/n} = \sqrt{2} \cdot d_{\lceil n/2 \rceil}^{(n-\lceil n/2 \rceil)/n} \\ &\leq \sqrt{2} \cdot C_0^{1/2} e^{-c_0/2(n/2)^\alpha \cdot (1-1/n)} \\ &= \sqrt{2} \cdot C_0^{1/2} e^{-2^{-1-\alpha}c_0 n^\alpha \cdot (1-1/n)} \\ &\stackrel{n \geq 2}{\leq} \sqrt{2} \cdot C_0^{1/2} e^{-2^{-2-\alpha}c_0 n^\alpha} \\ &= \sqrt{2C_0} e^{-c_1 n^\alpha}, \end{aligned}$$

where $c_1 := 2^{-(2+\alpha)}c_0$, and this concludes the proof. □

Remark 3 Observe that the constant $\tilde{C}_0 2^{\alpha+1/2} e^\alpha = \tilde{C}_0 \sqrt{2} (2e)^\alpha$ in (10) is significantly smaller than the one obtained in [4] for the algebraic rate, which is $C_0 2^{5\alpha+1}$. However, we have here instead the logarithmic factor in n , even if we presume that it may be possible to remove it with a finer analysis. This conjecture is supported by the fact that we found neither an analytical nor numerical example which required the additional logarithmic factor in n .

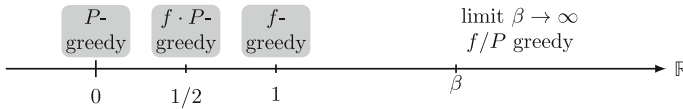


Fig. 1 Visualization of the scale of the β -greedy algorithms on the real line. Several important cases for $\beta \in \{0, 1/2, 1\}$ and $\beta \rightarrow \infty$ are marked

4 Analysis of Greedy Algorithms in the Kernel Setting

This section introduces and analyses β -greedy algorithms that are a scale of greedy algorithms which generalize the P -, $f \cdot P$ -, f - and f/P -greedy algorithms.

We work under the assumption

$$\|P_0\|_{L^\infty(\Omega)} = \sup_{x \in \Omega} \|k(\cdot, x)\|_{\mathcal{H}_k(\Omega)} = \sup_{x \in \Omega} \sqrt{k(x, x)} \leq 1. \tag{14}$$

4.1 A Scale of Greedy Algorithms: β -Greedy

We start with the definition of β -greedy algorithms.

Definition 4 A greedy kernel algorithm is called β -greedy algorithm with $\beta \in [0, \infty]$, if the next interpolation point is chosen as follows.

1. For $\beta \in [0, \infty)$ according to

$$x_{n+1} = \operatorname{argmax}_{x \in \Omega \setminus X_n} |(f - s_n)(x)|^\beta \cdot P_n(x)^{1-\beta}. \tag{15}$$

2. For $\beta = \infty$ according to the f/P -greedy algorithm.

As depicted in Fig. 1, for $\beta = 0$ this is the P -greedy algorithm, for $\beta = 1/2$ it is the $f \cdot P$ -algorithm, and for $\beta = 1$ it is the f -greedy algorithm. In the limit $\beta \rightarrow \infty$ it makes sense to define the algorithm to be the f/P -greedy algorithm.²

Observe that the β -greedy algorithms are well defined also for $1 < \beta < \infty$. Indeed, in this case $1 - \beta < 0$ and thus the power function part occurs as a divisor, and this may potentially be a problem since $P_n(x_i) = 0$ for all $1 \leq i \leq n$. Nevertheless, the standard power function estimate gives

$$|(f - s_n)(x)|^\beta \cdot P_n(x)^{1-\beta} = \frac{P_n(x)^\beta \cdot \|f - s_n\|_{\mathcal{H}_k(\Omega)}^\beta}{P_n(x)^{\beta-1}} \leq \|f - s_n\|_{\mathcal{H}_k(\Omega)}^\beta P_n(x),$$

i.e. it holds $\lim_{x \rightarrow x_j} |(f - s_n)(x)|^\beta \cdot P_n(x)^{1-\beta} = 0$ for all $x_j \in X_n$.

Remark 5 [Generalizations of the β -selection rule] We remark that it is sufficient to consider only one parameter $\beta > 0$ for the weighting of $|(f - s_n)(x)|$ and $P_n(x)$ as it was done in Eq. (15), in the sense that using two different parameters would be

² In this framework, the f/P -greedy algorithm is a limit case. This can be seen as an explanation why the f/P -greedy selection rule is sometimes not well defined, as discussed in Example 6 in [32].

useless. Indeed, due to the strict monotonicity of the function $x \mapsto x^{1/\alpha}$, for some $\alpha > 0$ and for $\gamma \in \mathbb{R}$ it holds

$$\operatorname{argmax}_{x \in \Omega \setminus X_n} |(f - s_n)(x)|^\alpha \cdot P_n(x)^\gamma = \operatorname{argmax}_{x \in \Omega \setminus X_n} |(f - s_n)(x)| \cdot P_n(x)^{\gamma/\alpha},$$

which shows that only the ratio γ/α is decisive. The specific parametrization via β and $1 - \beta$ in Eq. (15) was chosen in order to obtain f/P -greedy as the limit case $\beta \rightarrow \infty$.

4.2 Analysis of β -Greedy Algorithms

We can now prove the convergence of these algorithms. So far, analysis of greedy kernel algorithms mainly focused on estimates on $\|f - s_i\|_{L^\infty(\Omega)}$. Here and in the following, different quantities will be analyzed with the goal of bounding instead $\min_{i=n+1, \dots, 2n} \|f - s_i\|_{L^\infty(\Omega)}$. We remark that no requirements on the kernel k or the set Ω are needed for the results of this section, and especially for Theorem 8, as the proofs are based solely on RKHS theory.

We start by proving a key technical statement for greedy kernel interpolation that provides a bound on the product of the residual terms $r_i := f - s_i$. This result holds independently of the strategy that is used to select the points, greedy or not.

Lemma 6 *For any sequence $\{x_i\}_{i \in \mathbb{N}} \subset \Omega$ and any $f \in \mathcal{H}_k(\Omega)$ it holds for all $n = 1, 2, \dots$ that*

$$\left[\prod_{i=n+1}^{2n} |r_i(x_{i+1})| \right]^{1/n} \leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1}) \right]^{1/n}. \tag{16}$$

Proof Let

$$R_n^2 := \left[\prod_{i=n+1}^{2n} \left(\frac{r_i(x_{i+1})}{P_i(x_{i+1})} \right)^2 \right]^{1/n}.$$

The geometric arithmetic mean inequality gives

$$R_n^2 \leq \frac{1}{n} \sum_{i=n+1}^{2n} \left(\frac{r_i(x_{i+1})}{P_i(x_{i+1})} \right)^2 = \frac{1}{n} \left(\sum_{i=0}^{2n} \left(\frac{r_i(x_{i+1})}{P_i(x_{i+1})} \right)^2 - \sum_{i=0}^n \left(\frac{r_i(x_{i+1})}{P_i(x_{i+1})} \right)^2 \right).$$

We now use Eq. (7) applied to s_{2n+1} and s_{n+1} , and the properties of orthogonal projections to obtain

$$\begin{aligned} R_n^2 &\leq \frac{1}{n} \left(\|s_{2n+1}\|_{\mathcal{H}_k(\Omega)}^2 - \|s_{n+1}\|_{\mathcal{H}_k(\Omega)}^2 \right) \leq \frac{1}{n} \left(\|f\|_{\mathcal{H}_k(\Omega)}^2 - \|s_{n+1}\|_{\mathcal{H}_k(\Omega)}^2 \right) \\ &= \frac{1}{n} \|f - s_{n+1}\|_{\mathcal{H}_k(\Omega)}^2 = \frac{1}{n} \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}^2. \end{aligned}$$

It follows that $R_n \leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}$, and thus

$$\left[\prod_{i=n+1}^{2n} |r_i(x_{i+1})| \right]^{1/n} \leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1}) \right]^{1/n}.$$

□

In order to derive convergence statements in the $L^\infty(\Omega)$ norm based on Lemma 6, it is now required to find a relationship between $|r_i(x_{i+1})|$ and $\|r_i\|_{L^\infty(\Omega)}$. To this end, we have the following lemma for β -greedy algorithms. Observe that the sequence of points depends on the value of β , i.e. $x_n \equiv x_n^{(\beta)}$, but for notational convenience we drop the superscript.

Lemma 7 Any β -greedy algorithm with $\beta \in [0, \infty]$ applied to a function $f \in \mathcal{H}_k(\Omega)$ satisfies for $i = 0, 1, \dots$:

(a) In the case of $\beta \in [0, 1]$:

$$\|r_i\|_{L^\infty(\Omega)} \leq |r_i(x_{i+1})|^\beta \cdot P_i(x_{i+1})^{1-\beta} \cdot \|f - s_i\|_{\mathcal{H}_k(\Omega)}^{1-\beta}. \tag{17}$$

(b) In the case of $\beta \in (1, \infty]$ with $1/\infty := 0$:

$$\|r_i\|_{L^\infty(\Omega)} \leq \frac{|r_i(x_{i+1})|}{P_i(x_{i+1})^{1-1/\beta}} \cdot \|P_i\|_{L^\infty(\Omega)}^{1-1/\beta}. \tag{18}$$

Proof We prove the two cases separately:

(a) For $\beta = 0$, i.e. the P -greedy algorithm, this is the standard power function estimate in conjunction with the P -greedy selection criterion $P_n(x_{n+1}) = \|P_n\|_{L^\infty(\Omega)}$. For $\beta = 1$ this holds with equality as it is simply the selection criterion of f -greedy since we have here $r_n(x_{n+1}) = \|r_n\|_{L^\infty(\Omega)}$. We thus consider $\beta \in (0, 1)$ and let $\tilde{x}_{i+1} \in \Omega$ be such that $|r_i(\tilde{x}_{i+1})| = \|r_i\|_{L^\infty(\Omega)}$. Then, the selection criterion from Eq. (15) gives

$$|r_i(x)|^\beta \cdot P_i(x)^{1-\beta} \leq |r_i(x_{i+1})|^\beta \cdot P_i(x_{i+1})^{1-\beta} \quad \forall x \in \Omega,$$

and in particular

$$P_i(\tilde{x}_{i+1}) \leq \frac{|r_i(x_{i+1})|^{\frac{\beta}{1-\beta}}}{|r_i(\tilde{x}_{i+1})|^{\frac{\beta}{1-\beta}}} \cdot P_i(x_{i+1}).$$

Using this bound with the standard power function estimate gives

$$\begin{aligned} \|r_i\|_{L^\infty(\Omega)} &= |r_i(\tilde{x}_{i+1})| \leq P_i(\tilde{x}_{i+1}) \cdot \|f - s_i\|_{\mathcal{H}_k(\Omega)} \\ &\leq \frac{|r_i(x_{i+1})|^{\frac{\beta}{1-\beta}}}{|r_i(\tilde{x}_{i+1})|^{\frac{\beta}{1-\beta}}} \cdot P_i(x_{i+1}) \cdot \|f - s_i\|_{\mathcal{H}_k(\Omega)} \end{aligned}$$

$$= \frac{|r_i(x_{i+1})|^{\frac{\beta}{1-\beta}}}{\|r_i\|_{L^\infty(\Omega)}^{\frac{\beta}{1-\beta}}} \cdot P_i(x_{i+1}) \cdot \|f - s_i\|_{\mathcal{H}_k(\Omega)}.$$

This can be rearranged for $\|r_i\|_{L^\infty(\Omega)}$ to yield the final result.

(a) For $\beta \in (1, \infty)$, the selection criterion from Eq. (15) can be rearranged to

$$|r_i(x)|^\beta \leq \frac{|r_i(x_{i+1})|^\beta}{P_i(x_{i+1})^{\beta-1}} \cdot P_i(x)^{\beta-1} \quad \forall x \in \Omega \setminus X_i,$$

and taking the supremum $\sup_{x \in \Omega \setminus X_i}$ gives

$$\|r_i\|_{L^\infty(\Omega)} \leq \frac{|r_i(x_{i+1})|}{P_i(x_{i+1})^{\frac{\beta-1}{\beta}}} \cdot \|P_i\|_{L^\infty(\Omega)}^{\frac{\beta-1}{\beta}}.$$

For $\beta = \infty$, the selection criterion of the f/P -greedy algorithm can be directly rearranged to yield the statement (when using the notation $1/\infty = 0$). □

Using the results of Lemma 7 as lower bounds on $|r_i(x_{i+1})|$, it is now possible to control the left hand side of Inequality (16). This gives the main theorem of this section:

Theorem 8 *Any β -greedy algorithm with $\beta \in [0, \infty]$ applied to a function $f \in \mathcal{H}_k(\Omega)$ satisfies the following error bound for $n = 1, 2, \dots$:*

(a) *In the case of $\beta \in [0, 1]$:*

$$\left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)} \right]^{1/n} \leq n^{-\beta/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1}) \right]^{1/n}. \tag{19}$$

(b) *In the case of $\beta \in (1, \infty]$ with $1/\infty := 0$:*

$$\left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)} \right]^{1/n} \leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1})^{1/\beta} \right]^{1/n}. \tag{20}$$

Proof We prove the two cases separately:

(a) For $\beta = 0$, i.e. P -greedy, Eq. (17) gives $\|r_i\|_{L^\infty(\Omega)} \leq P_i(x_{i+1}) \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}$. Taking the product $\prod_{i=n+1}^{2n}$ and the n -th root in conjunction with the estimate $\|r_i\|_{\mathcal{H}_k(\Omega)} \leq \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}$ for $i = n + 1, \dots, 2n$ gives the result. For $\beta \in (0, 1]$, we start by reorganizing the estimate (17) of Lemma 7 to get

$$|r_i(x_{i+1})| \geq \left(\|r_i\|_{L^\infty(\Omega)}^{1/\beta} \right) / \left(P_i(x_{i+1})^{\frac{1-\beta}{\beta}} \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}^{\frac{1-\beta}{\beta}} \right),$$

and we use this to bound the left hand side of Eq. (16) as

$$\begin{aligned} n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1}) \right]^{1/n} &\geq \left[\prod_{i=n+1}^{2n} |r_i(x_{i+1})| \right]^{1/n} \\ &\geq \left[\prod_{i=n+1}^{2n} \left(\|r_i\|_{L^\infty(\Omega)}^{1/\beta} \right) / \left(P_i(x_{i+1})^{1-\beta} \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}^{1-\beta} \right) \right]^{1/n} \\ &= \left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)}^{1/\beta} \right]^{1/n} \left[\prod_{i=n+1}^{2n} P_i(x_{i+1})^{1-\beta} \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}^{1-\beta} \right]^{-1/n}. \end{aligned}$$

Rearranging the factors, and using again the fact that $\|r_i\|_{\mathcal{H}_k(\Omega)} \leq \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}$ for $i = n + 1, \dots, 2n$, gives

$$\begin{aligned} &\left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)}^{1/\beta} \right]^{1/n} \\ &\leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1})^{1/\beta} \right]^{1/n} \cdot \left[\prod_{i=n+1}^{2n} \|r_i\|_{\mathcal{H}_k(\Omega)}^{1-\beta} \right]^{1/n} \\ &\leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1})^{1/\beta} \right]^{1/n} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}^{1-\beta} \\ &\leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}^{1/\beta} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1})^{1/\beta} \right]^{1/n}. \end{aligned}$$

Now, the inequality can be raised to the exponent β to give the final statement.

(b) For $\beta \in (1, \infty]$ we proceed similarly by first rewriting Eq. (18) of Lemma 7 as

$$|r_i(x_{i+1})| \geq \left(\|r_i\|_{L^\infty(\Omega)} \cdot P_i(x_{i+1})^{1-1/\beta} \right) / \left(\|P_i\|_{L^\infty(\Omega)}^{1-1/\beta} \right),$$

and we lower bound the left hand side of Eq. (16) as

$$\begin{aligned} n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1}) \right]^{1/n} &\geq \left[\prod_{i=n+1}^{2n} |r_i(x_{i+1})| \right]^{1/n} \\ &\geq \left[\prod_{i=n+1}^{2n} \left(\|r_i\|_{L^\infty(\Omega)} \cdot P_i(x_{i+1})^{1-1/\beta} \right) / \left(\|P_i\|_{L^\infty(\Omega)}^{1-1/\beta} \right) \right]^{1/n}. \end{aligned}$$

Rearranging for $\left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)} \right]^{1/n}$ yields

$$\left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)} \right]^{1/n} \leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} \|P_i\|_{L^\infty(\Omega)}^{1-1/\beta} \right]^{1/n} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1})^{1/\beta} \right]^{1/n},$$

which gives the final result due to $\|P_i\|_{L^\infty(\Omega)} \leq 1$ for all $i = 0, 1, \dots$ □

4.3 An Improvement of the Standard Estimate

As an additional consequence of Lemma 7, Corollary 9 gives a new inequality that can be seen as an *improved standard power function estimate*, i.e. an improvement compared to the standard power function estimate from Eq. (3), that holds for any β -greedy algorithm.

Corollary 9 [Improved standard power function estimate] *Any β -greedy algorithm with $\beta \in [0, \infty]$ applied to a function $f \in \mathcal{H}_k(\Omega)$ satisfies for $i = 0, 1, \dots$ the following improved standard power function estimate (with $1/\infty := 0$):*

$$\|r_i\|_{L^\infty(\Omega)} \leq \|r_i\|_{\mathcal{H}_k(\Omega)} \cdot \begin{cases} P_i(x_{i+1}) & \beta \in [0, 1] \\ P_i(x_{i+1})^{1/\beta} & \beta \in (1, \infty] \end{cases} \tag{21}$$

Proof For both $\beta \in [0, 1]$ and $\beta \in (1, \infty]$ we use the upper bounds on $\|r_i\|_{L^\infty(\Omega)}$ as stated in Lemma 7 and further estimate the quantity $|r_i(x_{i+1})|$ via the standard power function estimate from Eq. (3) to get

$$\|r_i\|_{L^\infty(\Omega)} \leq |r_i(x_{i+1})|^\beta \cdot P_i(x_{i+1})^{1-\beta} \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}^{1-\beta} \leq P_i(x_{i+1}) \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}$$

for $\beta \in [0, 1]$, and

$$\|r_i\|_{L^\infty(\Omega)} \leq \frac{|r_i(x_{i+1})|}{P_i(x_{i+1})^{1-1/\beta}} \cdot \|P_i\|_{L^\infty(\Omega)}^{1-1/\beta} \leq P_i(x_{i+1})^{1/\beta} \cdot \|r_i\|_{\mathcal{H}_k(\Omega)}$$

for $\beta \in (1, \infty]$ by using $\|P_i\|_{L^\infty(\Omega)} \leq \|P_0\|_{L^\infty(\Omega)} \leq 1$ for all $i = 0, 1, \dots$ (see Eq. (14)). □

The estimate from Eq. (21) is an improved estimate in comparison with Eq. (3), in that it provides a bound on $\|r_i\|_{L^\infty(\Omega)}$ instead of $|r_i(x_{i+1})|$, and this is a strictly larger quantity except that in the case of the f -greedy algorithm (i.e. $\beta = 0$), where they coincide. Moreover, for $\beta \in [0, 1]$ the right hand side of the estimates of Eq. (3) and (21) coincide, while for $\beta > 1$ this improvement comes at the price of a smaller exponent on the power function term, since $1/\beta < 1$.

Table 1 Connection between the abstract setting and the kernel setting

Abstract setting	\mathcal{H}	$f \in \mathcal{F}$	$\text{span}\{f_0, \dots, f_{n-1}\}$
Kernel setting	$\mathcal{H}_k(\Omega)$	$k(\cdot, x), x \in \Omega$	$\text{span}\{k(\cdot, x_i), x_i \in X_n\}$

Remark 10 We will see in the following how to obtain convergence rates of the term $\min_{n+1 \leq i \leq 2n} \|r_i\|_{L^\infty(\Omega)}$. From a practitioner point of view this kind of result might be unsatisfactory, as it is unclear which interpolant s_i gives the best approximation. In this case it is possible to resort to the improved standard power function estimate of Corollary 9: This inequality suggests to pick s_{i^*} with $i^* := \operatorname{argmin}_{n+1 \leq i \leq 2n} P_i(x_{i+1})$.

5 Convergence Rates for Greedy Kernel Interpolation

We can finally combine the abstract Hilbert space analysis from Sect. 3 and the greedy kernel interpolation analysis from Sect. 4 and apply them to concrete classes of kernels.

First of all, we recall a convenient connection that was established in [20] between the abstract analysis of [4] and kernel interpolation. We repeat it as we need to include also the extension of Sect. 3, i.e., the new quantity ν_n . The goal is to frame the β -greedy algorithms as particular instances of the general greedy algorithm of Sect. 3. In this view we choose $\mathcal{H} = \mathcal{H}_k(\Omega)$ and $\mathcal{F} = \{k(\cdot, x), x \in \Omega\}$. The fact that this set is compact is implied by the decay to zero of its Kolmogorov width, that is equivalent to the existence of a sequence of points such that the associated power function converges to zero (see Eq. (23)). This choice means that $f = k(\cdot, x) \in \mathcal{F}$ can be uniquely associated with an $x \in \Omega$ and vice versa. This yields a realization of the abstract greedy algorithm that produces an approximation set

$$V_n = \text{span}\{f_0, \dots, f_{n-1}\} = \text{span}\{k(\cdot, x_i) \mid i = 1, \dots, n\} = V(X_n),$$

and thus this is a greedy kernel algorithm, with an appropriate selection rule. Table 1 summarizes these assignments.

With these choices, as can be seen from the definition in Eq. (8), σ_n is simply the maximal power function value and ν_n is the power function value at the selected point.

$$\begin{aligned} \sigma_n &\equiv \sup_{f \in \mathcal{F}} \operatorname{dist}(f, V_n)_{\mathcal{H}} = \sup_{f \in \mathcal{F}} \|f - \Pi_{V_n}(f)\|_{\mathcal{H}} \\ &= \sup_{x \in \Omega} \|k(\cdot, x) - \Pi_{V(X_n)}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)} = \|P_n\|_{L^\infty(\Omega)}, \\ \nu_n &\equiv \operatorname{dist}(f_n, V_n)_{\mathcal{H}} = \|f_n - \Pi_{V_n}(f)\|_{\mathcal{H}} \\ &= \|k(\cdot, x_{n+1}) - \Pi_{V_n}(k(\cdot, x_{n+1}))\|_{\mathcal{H}_k(\Omega)} = P_n(x_{n+1}). \end{aligned} \quad (22)$$

Moreover, d_n can be similarly bounded as

$$\begin{aligned}
 d_n &\equiv \inf_{Y_n \subset \mathcal{H}} \sup_{f \in \mathcal{F}} \text{dist}(f, Y_n)_{\mathcal{H}} = \inf_{Y_n \subset \mathcal{H}} \sup_{f \in \mathcal{F}} \|f - \Pi_{Y_n}(f)\|_{\mathcal{H}} \\
 &\leq \inf_{Y_n \subset \mathcal{F}} \sup_{f \in \mathcal{F}} \|f - \Pi_{Y_n}(f)\|_{\mathcal{H}_k(\Omega)} = \inf_{X_n \subset \Omega} \|P_{X_n}\|_{L^\infty(\Omega)}, \tag{23}
 \end{aligned}$$

and thus any convergence statement on $\|P_{X_n}\|_{L^\infty(\Omega)}$ for a given set of points $X_n \subset \Omega$ gives via Eq. (23) a bound on d_n .

Additionally, observe that the assumption $\|f\|_{\langle} \leq 1$ for $f \in \mathcal{F}$ implies in the kernel setting that

$$\|P_0\|_{L^\infty(\Omega)} = \sup_{x \in \Omega} \sqrt{k(x, x)} = \sup_{x \in \Omega} \|k(\cdot, x)\|_{\mathcal{H}_k(\Omega)} \leq 1. \tag{24}$$

5.1 Convergence Rates for β -Greedy Algorithms

From Theorem 8, it is now easily possible to derive convergence statements and decay rates for the kernel greedy algorithms, by bounding the right-hand side by Inequality (2) and using the interpretations of v_i and d_n from Eq. (22) and Eq. (23).

Corollary 11 *Assume that a β -greedy algorithm with $\beta \in [0, \infty]$ is applied to a function $f \in \mathcal{H}_k(\Omega)$. Let $\alpha, C_0, c_0 > 0$ be given constants, and set $1/\infty := 0$. Recall $r_i \equiv f - s_i$:*

1. *If there exists a sequence $(X_n)_{n \in \mathbb{N}} \subset \Omega$ of sets of points such that*

$$\left\| \tilde{f} - \Pi_{X_n} \tilde{f} \right\|_{L^\infty(\Omega)} \leq C_0 n^{-\alpha} \|\tilde{f}\|_{\mathcal{H}_k(\Omega)} \quad \forall \tilde{f} \in \mathcal{H}_k(\Omega),$$

then for all $\beta \geq 0$ and for all $n \geq 3$ it holds

$$\min_{n+1 \leq i \leq 2n} \|r_i\|_{L^\infty(\Omega)} \leq C \cdot n^{-\frac{\min\{1, \beta\}}{2}} (\log(n) \cdot n^{-1})^{\frac{\alpha}{\max\{1, \beta\}}} \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}, \tag{25}$$

with $C := (2^{\alpha+1/2} \max\{1, C_0\} e^\alpha)^{\frac{1}{\max\{1, \beta\}}}$. In particular

$$\min_{n+1 \leq i \leq 2n} \|r_i\|_{L^\infty(\Omega)} \leq C \cdot \log(n)^\alpha \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \begin{cases} n^{-\alpha-1/2} & f - \text{greedy} \\ n^{-\alpha-1/4} & f \cdot P - \text{greedy} \\ n^{-\alpha} & P - \text{greedy} \end{cases}.$$

2. *If there exists a sequence $(X_n)_{n \in \mathbb{N}} \subset \Omega$ of sets of points³ such that*

$$\left\| \tilde{f} - \Pi_{X_n} \tilde{f} \right\|_{L^\infty(\Omega)} \leq C_0 e^{-c_0 n^\alpha} \|\tilde{f}\|_{\mathcal{H}_k(\Omega)} \quad \forall \tilde{f} \in \mathcal{H}_k(\Omega),$$

then for all $\beta \geq 0$ and for all $n \geq 2$ it holds

$$\min_{n+1 \leq i \leq 2n} \|r_i\|_{L^\infty(\Omega)} \leq C \cdot n^{-\frac{\min\{1, \beta\}}{2}} e^{-c_1 n^\alpha} \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}, \tag{26}$$

³ We remark that this sequence of sets of points does not need to be nested.

with $C := (\sqrt{2 \max\{1, C_0\}})^{\frac{1}{\max\{1, \beta\}}}$ and $c_1 = 2^{-(2+\alpha)} c_0 / \max\{1, \beta\}$. In particular

$$\min_{i=n+1, \dots, 2n} \|r_i\|_{L^\infty(\Omega)} \leq C \cdot e^{-c_1 n^\alpha} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \begin{cases} n^{-1/2} & f - \text{greedy} \\ n^{-1/4} & f \cdot P - \text{greedy} \\ n^0 & P - \text{greedy} \end{cases}.$$

3. For f/P -greedy, for any kernel and for all $n \geq 1$ it holds

$$\min_{n+1 \leq i \leq 2n} \|r_i\|_{L^\infty(\Omega)} \leq n^{-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)}.$$

Proof The proof is a simple combination of Corollary 2 and Theorem 8, with the addition of the following simple steps:

First, the worst case bounds in $\mathcal{H}_k(\Omega)$ (either algebraic or exponential) imply the same bound on the power function via Eq. (2). Second, in all cases we use the results of Theorem 8 in combination with the bound

$$\min_{i=n+1, \dots, 2n} \|r_i\|_{L^\infty(\Omega)} \leq \left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)} \right]^{1/n}.$$

Then, Eq. (19) and (20) of Theorem 8 can be jointly written as

$$\left[\prod_{i=n+1}^{2n} \|r_i\|_{L^\infty(\Omega)} \right]^{1/n} \leq n^{-\frac{\min\{1, \beta\}}{2}} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \left[\prod_{i=n+1}^{2n} P_i(x_{i+1}) \right]^{\frac{1}{n \max\{1, \beta\}}}.$$

Plugging the bounds of Corollary 2 in the last inequality gives the result of the first two points. The third point directly follows from Eq. (20) for $\beta = \infty$ due to $P_i(x_{i+1}) \leq 1$ for all $i = 1, 2, \dots$. □

5.2 Translational Invariant Kernels

Strictly positive definite and translational invariant kernels are popular kernels for applications. To specialize our result to this interesting case, in this subsection we use the following assumption.

Assumption 1 Let $k(x, y) = \Phi(x - y)$ be a strictly positive definite translational invariant kernel with associated reproducing kernel Hilbert space $\mathcal{H}_k(\Omega)$, whereby the domain $\Omega \subset \mathbb{R}^d$ is assumed to be bounded with Lipschitz boundary and interior cone condition.

In this context, we have the following special case of Corollary 11. To highlight the results in the most relevant cases, we state them only for $\beta \in \{0, 1/2, 1, \infty\}$ even if similar statements hold for general $\beta > 0$.

Corollary 12 Under Assumptions 1, any β -greedy algorithm with $\beta \in \{0, 1/2, 1, \infty\}$ applied to some function $f \in \mathcal{H}_k(\Omega)$ satisfies the following error bounds for $n = 0, 1, \dots$, where the constants are defined as in Corollary 11.

1. In the case of kernels of finite smoothness $\tau > d/2$

$$\min_{i=n+1, \dots, 2n} \|r_i\|_{L^\infty(\Omega)} \leq C \cdot \log(n)^{\tau/d-1/2} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \begin{cases} n^{-\tau/d} & f - \text{greedy} \\ n^{1/4-\tau/d} & f \cdot P - \text{greedy} \\ n^{1/2-\tau/d} & P - \text{greedy} \end{cases} .$$

2. In the case of kernels of infinite smoothness:

$$\min_{i=n+1, \dots, 2n} \|r_i\|_{L^\infty(\Omega)} \leq C \cdot e^{-c_1 n^{1/d}} \cdot \|r_{n+1}\|_{\mathcal{H}_k(\Omega)} \cdot \begin{cases} n^{-1/2} & f - \text{greedy} \\ n^{-1/4} & f \cdot P - \text{greedy} \\ n^0 & P - \text{greedy} \end{cases} .$$

Observe that for any $\beta \in (0, 1]$ we have the additional convergence of order $n^{-\beta/2}$ or $n^{-1/2}$ for $\beta > 1$. The additional decay is faster for increasing $\beta \in (0, 1]$, i.e. increasing the weight of the target data-dependent term in the selection criterion gives better decay rates. Especially, the proven decay rate for f -greedy is better than the one for $f \cdot P$ -greedy which is better than the one for P -greedy.

This additional convergence proves in particular that the Kolmogorov barrier can be broken, i.e., approximation rates that are better than the ones provided by the Kolmogorov width can be obtained for any function in $\mathcal{H}_k(\Omega)$. Indeed, as discussed above any bound on d_n turns into a bound on $\|P_n\|_{L^\infty(\Omega)}$, which can then be used in Corollary 11 or Corollary 12.

This is particularly relevant for the kernels whose RKHS is norm equivalent to a Sobolev space. But also other general kernels of low smoothness are of interest, since it might happen that the power function is decaying at arbitrarily slow speed, while the adaptive points selected by a β -greedy algorithm provide an additional convergence rate.

Moreover, the additional decay for $\beta > 0$ is dimension independent and thus it does not incur in the curse of dimensionality. This is of interest in particular for translational invariant kernels of Corollary 12, as both the algebraic and the exponential decay of the power function (or Kolmogorov width) degrade with the dimension d and thus the additional term gains more importance.

Despite this notable relevance, the estimates of Corollary 11 and Corollary 12 are likely not optimal in the algebraic case. Indeed, for kernels with algebraically decaying Kolmogorov width, in the case of the P -greedy algorithm ($\beta = 0$) bounds without the additional $\log(n)^\alpha$ factor are known [20]. We thus expect that the inconvenient additional $\log(n)^\alpha$ factor is not required for any of the β -greedy algorithms. We remark that this factor is related to the additional ϵ within Corollary 2, but we did not find a way to get rid of it, with exception of $\beta = 0$, i.e. the P -greedy case. Moreover, we obtained our bounds by means of the worst-case bounds on $(\prod_{i=n+1}^{2n} P_i(x_{i+1}))^{1/n}$ from Corollary 2. Numerically, a faster decay than the worst case bound from Corollary 2 can often be observed (see the examples in Sect. 6.1). Especially, for each β value we obtain a different sequence of points and thus a different decay of the corresponding power function values.

Remark 13 [Additional convergence orders] Additional convergence orders can be obtained from the decay of $\|r_n\|_{\mathcal{H}_k(\Omega)}$. Even if this quantity is in general decaying at

arbitrarily slow speed for a general $f \in \mathcal{H}_k(\Omega)$, we mention the case of superconvergence [24, 25], which allows to bound $\|r_i\|_{\mathcal{H}_k(\Omega)} \leq C_f \cdot \|P_i\|_{L^2(\Omega)}$ for special functions $f \in \mathcal{H}_k(\Omega)$. The original superconvergence requirement $f = Tg$ (whereby T is the kernel integral operator and $g \in L^2(\Omega)$, i.e. $(Tg)(x) = \int_{\Omega} k(x, y)g(y)dy$) can be extended to functions $f \in \mathcal{H}_k(\Omega)$ such that $|\langle f, g \rangle_{\mathcal{H}_k(\Omega)}| \leq C_f \cdot \|g\|_{L^q(\Omega)}$ for all $g \in \mathcal{H}_k(\Omega)$ (see [22, Theorem 19]).

Remark 14 [Stability] The stability of the greedy interpolation, as computed here by the so-called direct method, is mainly linked to the smallest eigenvalue of the kernel interpolation matrix. A standard result [23] gives the upper estimate $\lambda_{\min}(X_n) \leq P_{n-1}(x_n)^2$. In view of the estimates of Eqs. (25) and (26), this means that a faster convergence based on a faster decay of the power function values $P_i(x_{i+1})$ directly negatively influences the stability. This holds especially for $\beta > 1$, because in this case the upper bound for the convergence in terms of the power function scales with the exponent $1/\beta < 1$.

Remark 15 [Other greedy selection rules] The analysis above shows that γ -greedy algorithms which were introduced in [32] are actually closer to the P -greedy algorithm than to target data-dependent algorithms for the case of kernels of finite smoothness $\tau > d/2$. In this case for γ -greedy algorithms the decay of $P_n(x_{n+1})$ can be both lower and upper bounded by a constant times $n^{1/2-\tau/d}$. As the point selection criteria of γ -stabilized greedy algorithms first of all look at the power function value via $P_n(x_{n+1}) \geq \gamma \cdot \|P_n\|_{L^\infty(\Omega)}$, there is no relationship as in Eq. (15) ($\beta > 0$). Thus, we cannot derive additional convergence rates.

Remark 16 [Other norms] For kernels of finite smoothness $\tau > d/2$ on a set Ω with Lipschitz boundary satisfying an interior cone condition, the optimal rates of L^p -convergence are of order $\|r_n\|_{L^p(\Omega)} \leq c_p n^{-\tau/d+(1/2-1/p)_+}$. This rate is matched by the P -greedy algorithm (see [32, Corollary 22]), since it is proven to select asymptotically uniformly distributed points.

In the case of the f -greedy algorithm, we can use the additional factor $n^{-1/2}$ from Corollary 12 to get rid of the conversion from the L^p to the L^∞ norm, i.e. we have

$$\|r_n\|_{L^p(\Omega)} \leq \text{meas}(\Omega)^{1/p} \|r_n\|_{L^\infty(\Omega)} \leq c_\infty \log(n)^{\tau/d-1/2} n^{-\tau/d}.$$

So we have almost L^p -optimal results (up to the poly-logarithmic factor) for $p \in [1, 2]$ and even improved convergence for $p \in (2, \infty]$. Similar statements hold for general β -greedy algorithms.

6 Examples

6.1 Visualization of Results of Abstract Setting

This subsection visualizes the results from the abstract analysis in Sect. 3, especially Sect. 3.1. Again we make use of the links recalled in the beginning of Sect. 5, especially in Eqs. (22) and (23).

We consider the domain $\Omega = [0, 1]^3 \subset \mathbb{R}^3$ and the Gaussian kernel with kernel width parameter 2, i.e. $k(x, y) = \exp(-4\|x - y\|_2^2)$. Four different sequences of points are considered, with colors referring to Fig. 2:

- i. Blue: P -greedy algorithm on the whole domain Ω .
- ii. Red: P -greedy algorithm on the subdomain $\Omega_2 := \{x \in \Omega \mid (x)_3 = 1/2\}$. Like this, the dimension is effectively reduced from $d = 3$ to $d = 2$.
- iii. Yellow, violet: The points are independently randomly picked within Ω according to a uniform distribution.

The results are displayed in Fig. 2:

- The upper two figures displays the quantities $\sigma_n = \|P_n\|_{L^\infty(\Omega)}$ (left) and $\nu_n = P_n(x_{n+1})$ (right).
- The lower two figures display

$$n \mapsto \left(\prod_{j=n+1}^{2n} \sigma_j \right)^{1/n} = \left(\prod_{j=n+1}^{2n} \|P_j\|_{L^\infty(\Omega)} \right)^{1/n} \quad \text{(left),}$$

$$n \mapsto \left(\prod_{j=n+1}^{2n} \nu_j \right)^{1/n} = \left(\prod_{j=n+1}^{2n} P_j(x_{j+1}) \right)^{1/n} \quad \text{(right).}$$

For the numerical experiments, the domain Ω was discretized using $2 \cdot 10^4$ random points and Ω_2 was discretized by projecting the random points related to Ω onto Ω_2 . The algorithms run until 300 points are selected or the next selected Power function value satisfies $P_n(x_{n+1}) < 10^{-5}$.

From the top left picture, one can infer that the displayed quantity $\|P_n\|_{L^\infty(\Omega)}$ decays fastest for the P -greedy algorithm. This was expected, as the algorithms directly aims at minimizing this quantity. However, the displayed quantity $\|P_n\|_{L^\infty(\Omega)}$ does not drop at all for the P -greedy algorithm on Ω_2 , as it picks only points from Ω_2 and thus does not fill Ω .

Contrarily, the top right picture shows that the displayed quantity $P_n(x_{n+1})$ decays faster for the P -greedy algorithm on Ω_2 , while for the P -greedy algorithm on Ω we have exactly the same curve due to $P_n(x_{n+1}) = \|P_n\|_{L^\infty(\Omega)}$. The two further point choices exhibit a wiggling, noisy behavior on the displayed $P_n(x_{n+1})$ quantity, which is related to the random point choice.

The two lower figures refer to the geometric mean $(\prod_{j=n+1}^{2n} \dots)^{1/n}$ of the quantities of the upper figures. In the lower left figure, we can see that only the curve related to the P -greedy algorithm on Ω decays fast, the other curves do not decay at all or only slowly—because the points are not chosen in a way to minimize the maximal Power function value $\|P_n\|_{L^\infty(\Omega)}$. Contrarily, the P -greedy algorithm on Ω exhibits the slowest decay of the quantity $(\prod_{j=n+1}^{2n} \nu_j)^{1/n}$, which is the same curve as in the lower left figure due to $\nu_j = P_j(x_{j+1}) = \|P_j\|_{L^\infty(\Omega)} = \sigma_j$. However, all the three other choices of points provide a faster decay of the displayed quantity $(\prod_{j=n+1}^{2n} P_j(x_{j+1}))^{1/n} = (\prod_{j=n+1}^{2n} \nu_j)^{1/n}$. The theoretical reason for (at least) the same decay as the P -greedy algorithm on Ω was proven in Corollary 2.

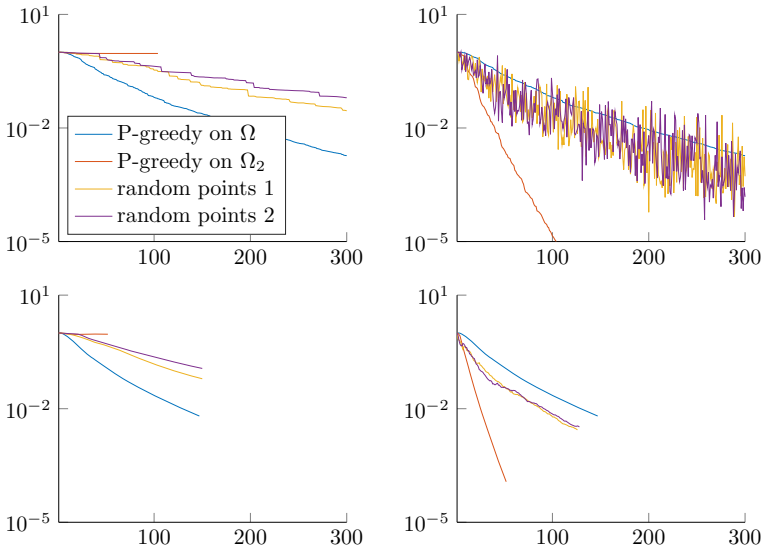


Fig. 2 Decay of several power function related quantities (y-axis) depending on their index n (x-axis) for four different choices of points: The upper two plots display the quantities $\sigma_n = \|P_n\|_{L^\infty(\Omega)}$ (left) respective $v_n = P_n(x_{n+1})$ (right). The lower two plots display the quantities $(\prod_{j=n+1}^{2n} \sigma_j)^{1/n} = (\prod_{j=n+1}^{2n} \|P_j\|_{L^\infty(\Omega)})^{1/n}$ (left) respective $(\prod_{j=n+1}^{2n} v_j)^{1/n} = (\prod_{j=n+1}^{2n} P_j(x_{j+1}))^{1/n}$ (right)

6.2 β -Greedy Algorithms Using the Wendland Kernel

We consider the application of β -greedy algorithms for the particular example of the Wendland $k = 0$ kernel on the domain $\Omega = [0, 1]$, which is defined as

$$k(x, y) = \max(1 - |x - y|, 0),$$

and thus a piecewise linear kernel. Its native space $\mathcal{H}_k(\Omega)$ is norm equivalent to the Sobolev space $W_2^1(\Omega)$. It is immediate to see that kernel interpolation using the Wendland $k = 0$ kernel on centers $X_n \subset \Omega$ boils down to piecewise linear spline interpolation on the subinterval $[\min X_n, \max X_n] \subset [0, 1]$. On $\Omega \setminus [\min X_n, \max X_n]$ the interpolant is still an affine function.

We consider the function $f : \Omega \rightarrow \mathbb{R}, x \mapsto x^\alpha$ for some $1/2 < \alpha < 1$. For $\alpha > 1/2$ it holds $f \in W_2^1(\Omega)$, thus $f \in \mathcal{H}_k(\Omega)$. It can be shown, that in the case of asymptotically uniform interpolation points—i.e. $q_n \asymp h_n \asymp n^{-1}$, whereby $q_n = \min_{x_i \neq x_j \in X_n} \|x_i - x_j\|_2$ is the so called separation distance—it is possible to lower-bound the error as (for details see “Appendix A”)

$$\|f - s_n\|_{L^\infty(\Omega)} \geq C_\alpha \cdot n^{-\alpha} \tag{27}$$

for $C_\alpha > 0$. Furthermore, independent of the way the interpolation points X_N were chosen (i.e. even for optimally chosen points), it holds

$$\|f - s_n\|_{L^\infty(\Omega)} \geq C \cdot n^{-2} \tag{28}$$

for some $C > 0$. Thus, we can infer:

- Any (greedy) algorithm that yields asymptotically uniformly distributed points cannot have a convergence rate better than $n^{-\alpha}$ for this particular example. This includes especially the P -greedy algorithm, but also any γ -stabilized greedy algorithms [32], as they are known to provide asymptotically uniform points as well, see [32, Theorem 20]. Thus, this example shows that γ -stabilized greedy algorithms cannot be expected in general to give a better approximation rate than the P -greedy algorithm (however they were motivated by their use for the preasymptotic range).

Especially for $\alpha \rightarrow 1/2$, the convergence rate can be arbitrary close to $1/2$.

- For the f -greedy and $f \cdot P$ -greedy algorithms we have a convergence of at least $\log(n)^{1/2} \cdot n^{-1}$ respective $\log(n)^{1/2} \cdot n^{-3/4}$ according to Corollary 12, which is strictly better compared to the P -greedy algorithm.

Figure 3 visualizes the convergence of several β -greedy algorithms for the described setting. One can observe that the error for the P -greedy algorithm ($\beta = 0$) decays approximately according to $n^{-1/2}$, which is in accordance with Eq. (27). For the f -greedy algorithm ($\beta = 1$) the error seems to decay according to n^{-2} , which is the fastest possible decay rate according to Eq. (28). For all intermediate β values one can observe intermediate convergence rates: For values of β closer to 1, the error decays faster. The f/P -greedy algorithm ($\beta = \infty$) seems to give a convergence in between $n^{-1/2}$ and n^{-2} .

We remark that this behavior of the error decay depending on β is not unique to the Wendland $k = 0$ kernel, but can also be observed for other kernels, domains and target functions f . This particular example was chosen, because it is analytically easily possible to derive several explicit statements on convergence rates for asymptotically uniform and adapted points.

6.3 Approximation of Franke’s Test Function

As a final example in 2D we consider the approximation of Franke’s test function which is defined on $\Omega = [0, 1]$ as

$$\begin{aligned} f(x) = & 0.75e^{-\frac{(9(x)_1-2)^2}{4} - \frac{(9(x)_2-2)^2}{4}} + 0.75e^{-\frac{(9(x)_1+1)^2}{49} - \frac{9(x)_2+1}{10}} \\ & + 0.5e^{-\frac{(9(x)_1-7)^2}{4} - \frac{(9(x)_2-3)^2}{4}} - 0.2e^{-(9(x)_1-4)^2 - (9(x)_2-7)^2}. \end{aligned}$$

Therefore, we use the linear Matérn kernel which is given as

$$k(x, y) = (1 + \|x - y\|_2) \cdot e^{-\|x - y\|_2}$$

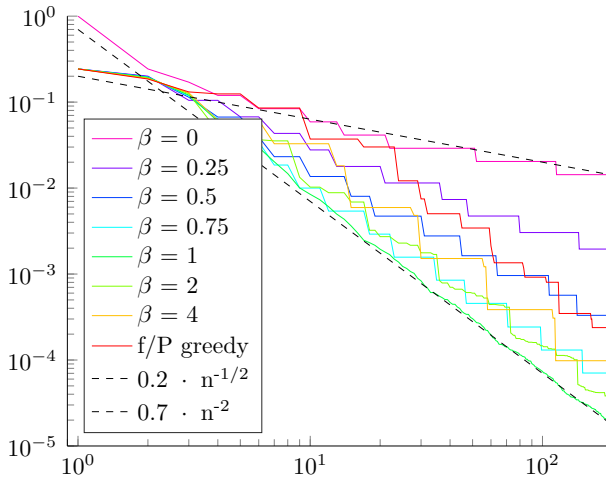


Fig. 3 Decay of the error $\|f - s_n^{(\beta)}\|_{L^\infty(\Omega)}$ (y-axis) for β -greedy algorithms in the number n of chosen interpolation points (x-axis) for $\beta \in \{0, 0.25, 0.5, 0.75, 1, 2, 4, \infty\}$ and $f(x) = x^\alpha$ with $\alpha = 0.51$. Two additional dashed lines indicate a rate of convergence of $n^{-1/2}$ and n^{-2}

and run β -greedy algorithms using $\beta \in \{0, 0.5, 1, \infty\}$. The resulting points are visualized in Fig. 4. For $\beta = 0$, i.e. P -greedy, the points are quite uniformly distributed, which is according to the theoretical results in [32]. For $\beta = \infty$, i.e. f/P -greedy, the points are quite clustered around a few spots. For $\beta = 0.5$ ($f \cdot P$ -greedy) and $\beta = 1$ (f -greedy), an intermediate behavior can be observed: The points are still slightly clustered, but also fill the whole domain.

7 Conclusion and Outlook

Using an abstract analysis of greedy algorithms in Hilbert spaces, it was shown that arbitrary point sequences—e.g., generated from arbitrary greedy kernel algorithms—yield certain decay rates for specific power function quantities. Based on these results and refined greedy kernel interpolation analysis it was possible to investigate and prove convergence statements for a range of greedy kernel algorithms including the target data-dependent f -, $f \cdot P$ - and f/P -greedy algorithms. The provided techniques and results will likely lead to further advancements, e.g., in the field of kernel quadrature.

Several points remain open, and they will be addressed in future research. First, the proven decay rate for the f/P -greedy algorithm is still not satisfactory and is likely improvable. Moreover, the results are independent of the special choice of function $f \in \mathcal{H}_k(\Omega)$. How to make use of properties of that function? It would be desirable to conclude a faster decay of the $(\prod_{i=n+1}^{2n} P_i(x_{i+1}))^{1/n}$ -quantity based on properties of the considered function $f \in \mathcal{H}_k(\Omega)$. Finally, it is still unclear if it is possible to derive general statements on the decay of $\|f - s_n\|_{\mathcal{H}_k(\Omega)}$, and what is the relationship between this fact and superconvergence.

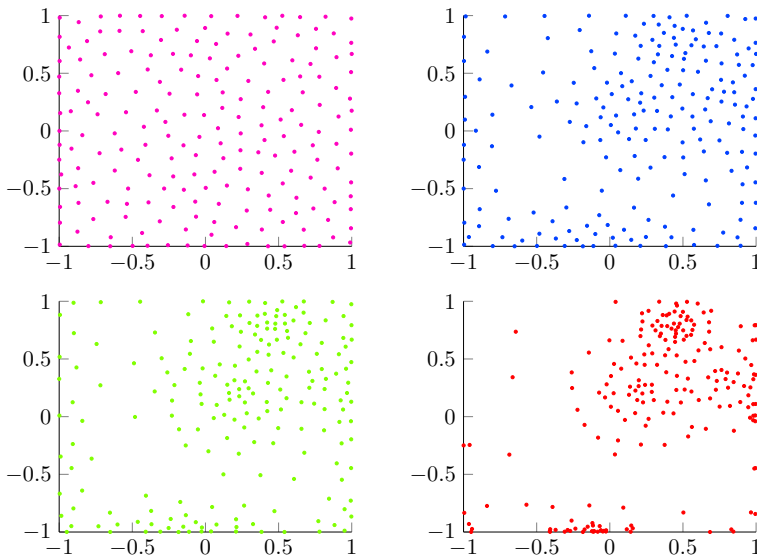


Fig. 4 Visualization of the greedily chosen points for the interpolation of Franke’s test function. From left to right, top to bottom we used $\beta \in \{0, 0.5, 1, \infty\}$. For $\beta = 0$ the points are quite uniformly distributed, for $\beta = \infty$ strong clustering can be observed. The intermediate values $\beta \in \{0.5, 1\}$ provide intermediate cases

Acknowledgements The authors acknowledge the funding of the project by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC 2075 - 390740016 and funding by the BMBF under contract 05M20VSA.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Details on Section 6.1

Consider the $k = 0$ Wendland kernel and the domain $\Omega = [0, 1]$. Then, the native space $\mathcal{H}_k(\Omega)$ is norm equivalent to the Sobolev space $W_2^1(\Omega)$. We remark that kernel interpolation using the Wendland $k = 0$ kernel on centers $X_n \subset \Omega$ boils down to piecewise linear spline interpolation on the subinterval $[\min X_n, \max X_n] \subset [0, 1]$. On $\Omega \setminus [\min X_n, \max X_n]$ the interpolant is still an affine function.

We consider the function $f : \Omega \rightarrow \mathbb{R}, x \mapsto x^\alpha$ for some $1/2 < \alpha < 1$. For $\alpha > 1/2$, it holds $f \in W_2^1(\Omega)$, thus $f \in \mathcal{H}_k(\Omega)$. We consider the interpolation using not-yet

specified points $X_n \subset \Omega$. Define $z_1 := z_1^{(n)} := \min X_n$, $z_2 := z_2^{(n)} := \min X_n \setminus \{z_1\}$. We have for $n \geq 2$:

$$s_n(x)|_{x \in [z_1, z_2]} = \frac{z_2^\alpha - z_1^\alpha}{z_2 - z_1} \cdot (x - z_1) + z_1^\alpha$$

We can estimate $\|f - s_n\|_{L^\infty(\Omega)}$ via:

$$\begin{aligned} \|f - s_n\|_{L^1([z_1, z_2])} &\leq \|f - s_n\|_{L^\infty([z_1, z_2])} \cdot |z_2 - z_1| \\ &\leq \|f - s_n\|_{L^\infty(\Omega)} \cdot |z_2 - z_1| \\ \Leftrightarrow \|f - s_n\|_{L^\infty(\Omega)} &\geq \|f - s_n\|_{L^1([z_1, z_2])} / |z_2 - z_1| \end{aligned}$$

The integral $\|f - s_n\|_{L^1([z_1, z_2])}$ can be computed as

$$\begin{aligned} \|f - s_n\|_{L^1([z_1, z_2])} &= \int_{z_1}^{z_2} x^\alpha - \left(\frac{z_2^\alpha - z_1^\alpha}{z_2 - z_1} \cdot (x - z_1) + z_1^\alpha \right) dx \\ &= \frac{1}{1 + \alpha} (z_2^{1+\alpha} - z_1^{1+\alpha}) - \int_0^{z_2 - z_1} \frac{z_2^\alpha - z_1^\alpha}{z_2 - z_1} \cdot x + z_1^\alpha dx \\ &= \frac{1}{1 + \alpha} (z_2^{1+\alpha} - z_1^{1+\alpha}) - \frac{1}{2} \cdot \frac{z_2^\alpha - z_1^\alpha}{z_2 - z_1} \cdot (z_2 - z_1)^2 - z_1^\alpha \cdot (z_2 - z_1). \end{aligned}$$

Thus, we have

$$\begin{aligned} \|f - s_n\|_{L^\infty(\Omega)} &\geq \frac{\|f - s_n\|_{L^1([z_1, z_2])}}{z_2 - z_1} \\ &= \frac{1}{1 + \alpha} \frac{z_2^{1+\alpha} - z_1^{1+\alpha}}{z_2 - z_1} - \frac{1}{2} \cdot (z_2^\alpha - z_1^\alpha) - z_1^\alpha \\ &= \frac{1}{1 + \alpha} \frac{z_2^{1+\alpha} - z_1^{1+\alpha}}{z_2 - z_1} - \frac{1}{2} \cdot (z_2^\alpha + z_1^\alpha) \\ &= z_2^\alpha \cdot \left(\frac{1}{1 + \alpha} \frac{z_2 - z_1 z_1^\alpha z_2^{-\alpha}}{z_2 - z_1} - \frac{1 + z_1^\alpha z_2^{-\alpha}}{2} \right) \\ &= z_2^\alpha \cdot \left(\frac{1}{1 + \alpha} \frac{z_2 - z_1 (1 + z_1^\alpha z_2^{-\alpha} - 1)}{z_2 - z_1} - \frac{1 + z_1^\alpha z_2^{-\alpha}}{2} \right) \\ &= z_2^\alpha \cdot \left(\frac{1}{1 + \alpha} + \frac{1}{1 + \alpha} \frac{z_1 (1 - z_1^\alpha z_2^{-\alpha})}{z_2 - z_1} - \frac{1 + z_1^\alpha z_2^{-\alpha}}{2} \right) \\ &= z_2^\alpha \cdot \left(\frac{1}{1 + \alpha} + \frac{1}{1 + \alpha} \frac{1 - z_1^\alpha z_2^{-\alpha}}{z_2 z_1^{-1} - 1} - \frac{1 + z_1^\alpha z_2^{-\alpha}}{2} \right). \end{aligned}$$

We proceed by setting $k := z_2/z_1$:

$$\|f - s_n\|_{L^\infty(\Omega)} \geq z_2^\alpha \cdot \underbrace{\left(\frac{1}{1+\alpha} + \frac{1}{1+\alpha} \frac{1-k^{-\alpha}}{k-1} - \frac{1+k^{-\alpha}}{2} \right)}_{=: h_\alpha(k)}.$$

We consider asymptotically uniform points, i.e. $\exists C > 0 \forall n \in \mathbb{N} \ h_n/q_n \leq C$. Based on the definition of the separation and fill distance we can estimate

$$k \equiv \frac{z_2}{z_1} \geq \frac{z_1 + q_n}{z_1} \geq 1 + \frac{q_n}{z_1} \geq 1 + \frac{q_n}{h_n}$$

Using the asymptotic uniformity, we have finally

$$k \geq 1 + \frac{q_n}{h_n} \geq 1 + C^{-1} > 1.$$

Finally, an analysis of the 1D function $h_\alpha : [1 + C^{-1}, \infty) \rightarrow \mathbb{R}$ shows that it holds

$$h_\alpha(k) \geq h_\alpha(1 + C^{-1}) > 0$$

for $k \in [1 + C^{-1}, \infty)$. This finally implies

$$\begin{aligned} \|f - s_n\|_{L^\infty(\Omega)} &\geq z_2^\alpha \cdot \min_{k \in [1+C^{-1}, \infty)} h_\alpha(k) \\ &\geq c \cdot n^{-\alpha} \end{aligned}$$

for some $c > 0$ due to $z_2 \geq q_n \geq \tilde{c}n^{-1}$.

References

1. Cohen, A., Dahmen, W., DeVore, R.: Orthogonal matching pursuit under the restricted isometry property. *Constr. Approx.* **45**(1), 113–127 (2017)
2. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. *Constr. Approx.* **13**(1), 57–98 (1997)
3. De Marchi, S., Schaback, R., Wendland, H.: Near-optimal data-independent point locations for radial basis function interpolation. *Adv. Comput. Math.* **23**(3), 317–330 (2005)
4. DeVore, R., Petrova, G., Wojtaszczyk, P.: Greedy algorithms for reduced bases in Banach spaces. *Constr. Approx.* **37**(3), 455–466 (2013)
5. DeVore, R.A., Temlyakov, V.N.: Some remarks on greedy algorithms. *Adv. Comput. Math.* **5**(2–3), 173–187 (1996)
6. Dutta, S., Farthing, M.W., Perracchione, E., Savant, G., Putti, M.: A greedy non-intrusive reduced order model for shallow water equations. *J. Comput. Phys.* **439**, 110378 (2021)
7. Fasshauer, G.E.: *Meshfree Approximation Methods with MATLAB*, Volume 6 Interdisciplinary Mathematical Sciences. World Scientific Publishing Co. Pte. Ltd., Hackensack (2007)
8. Fasshauer, G.E., McCourt, M.: *Kernel-Based Approximation Methods Using MATLAB*, Volume 19 of Interdisciplinary Mathematical Sciences. World Scientific Publishing Co. Pte. Ltd., Hackensack (2015)

9. Haasdonk, B., Santin, G.: Greedy kernel approximation for sparse surrogate modeling. In: Keiper, W., Milde, A., Volkwein, S. (eds.) *Reduced-Order Modeling (ROM) for Simulation and Optimization: Powerful Algorithms as Key Enablers for Scientific Computing*, pp. 21–45. Springer, Cham (2018)
10. Koepl, T., Santin, G., Haasdonk, B., Helmig, R.: Numerical modelling of a peripheral arterial stenosis using dimensionally reduced models and kernel methods. *Int. J. Numer. Methods Biomed. Eng.* **34**(8), e3095 (2018)
11. Köppel, M., Franzelin, F., Kröker, I., Oladyskhin, S., Santin, G., Wittwar, D., Barth, A., Haasdonk, B., Nowak, W., Pflüger, D., Rohde, C.: Comparison of data-driven uncertainty quantification methods for a carbon dioxide storage benchmark scenario. *Comput. Geosci.* **23**(2), 339–354 (2019)
12. Maday, Y., Nguyen, N.C., Patera, A.T., Pau, S.H.: A general multipurpose interpolation procedure: the magic points. *Commun. Pure Appl. Anal.* **8**(1), 383–404 (2009)
13. Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41**(12), 3397–3415 (1993)
14. Müller, S.: *Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden (Complexity and Stability of Kernel-based Reconstructions)*. PhD thesis, Fakultät für Mathematik und Informatik, Georg-August-Universität Göttingen (2009)
15. Müller, S., Schaback, R.: A Newton basis for kernel spaces. *J. Approx. Theory* **161**(2), 645–655 (2009)
16. Narcowich, F.J., Ward, J.D., Wendland, H.: Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. *Math. Comput.* **74**(250), 743–763 (2005)
17. Narcowich, F.J., Ward, J.D., Wendland, H.: Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. *Constr. Approx.* **24**(2), 175–186 (2006)
18. Pazouki, M., Schaback, R.: Bases for kernel-based spaces. *J. Comput. Appl. Math.* **236**(4), 575–588 (2011)
19. Pinkus, A.: *n -Widths in Approximation Theory. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, vol. 7. Springer, Berlin (1985)
20. Santin, G., Haasdonk, B.: Convergence rate of the data-independent P -greedy algorithm in kernel-based approximation. *Dolomites Res. Notes Approx.* **10**, 68–78 (2017)
21. Santin, G., Haasdonk, B.: Kernel methods for surrogate modeling. In: Benner, P., Grivet-Talocia, S., Quarteroni, A., Rozza, G., Schilders, W., Silveira, L.M. (eds.) *Model Order Reduction*, vol. 2. De Gruyter (2021)
22. Santin, G., Karvonen, T., Haasdonk, B.: Sampling based approximation of linear functionals in reproducing kernel hilbert spaces. *BIT Numer. Math.* **62**(1), 279–310 (2022)
23. Schaback, R.: Error estimates and condition numbers for radial basis function interpolation. *Adv. Comput. Math.* **3**(3), 251–264 (1995)
24. Schaback, R.: Improved error bounds for scattered data interpolation by radial basis functions. *Math. Comp.* **68**(225), 201–216 (1999)
25. Schaback, R.: Superconvergence of kernel-based interpolation. *J. Approx. Theory* **235**, 1–19 (2018)
26. Schaback, R., Wendland, H.: Adaptive greedy techniques for approximate solution of large RBF systems. *Numer. Algorithms* **24**(3), 239–254 (2000)
27. Schaback, R., Wendland, H.: Numerical techniques based on radial basis functions. In: *Curve and Surface Fitting: Saint-Malo 1999*. Vanderbilt University Press, pp. 359–374 (2000)
28. Schmidt, A., Haasdonk, B.: Data-driven surrogates of value functions and applications to feedback control for dynamical systems. *IFAC-PapersOnLine*, 51(2):307–312, 2018. 9th Vienna International Conference on Mathematical Modelling
29. Temlyakov, V.N.: Greedy approximation. *Acta Numer.* **17**, 235–409 (2008)
30. Wendland, H.: *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics, vol. 17. Cambridge University Press, Cambridge (2005)
31. Wendland, H., Rieger, C.: Approximate interpolation with applications to selecting smoothing parameters. *Numer. Math.* **101**(4), 729–748 (2005)
32. Wenzel, T., Santin, G., Haasdonk, B.: A novel class of stabilized greedy kernel approximation algorithms: convergence, stability and uniform point distribution. *J. Approx. Theory* **262**, 105508 (2021)
33. Wirtz, D., Haasdonk, B.: A vectorial kernel orthogonal greedy algorithm. *Dolomites Res. Notes Approx.* **6**, 83–100 (2013)