Available online at www.sciencedirect.com

ScienceDirect

Journal of Approximation Theory 262 (2021) 105508

Journal of Approximation Theory

www.elsevier.com/locate/jat

Full Length Article

# A novel class of stabilized greedy kernel approximation algorithms: Convergence, stability and uniform point distribution

Tizian Wenzel[a,*], Gabriele Santin[a,b], Bernard Haasdonk[a]

[a] *Institute for Applied Analysis and Numerical Simulation, University of Stuttgart, Germany*
[b] *Center for Information and Communication Technology, Fondazione Bruno Kessler, Trento, Italy*

Received 27 March 2020; received in revised form 21 September 2020; accepted 14 November 2020
Available online 2 December 2020

Communicated by M.J. Johnson

**Abstract**

Kernel based methods provide a way to reconstruct potentially high-dimensional functions from meshfree samples, i.e., sampling points and corresponding target values. A crucial ingredient for this to be successful is the distribution of the sampling points. Since the computation of an optimal selection of sampling points may be an infeasible task, one promising option is to use greedy methods.

Although these methods may be very effective, depending on the specific greedy criterion the chosen points might quickly lead to instabilities in the computation. To circumvent this problem, we introduce and investigate a new class of *stabilized* greedy kernel algorithms, which can be used to create a scale of new selection strategies.

We analyze these algorithms, and in particular we prove convergence results and quantify in a precise way the distribution of the selected points. These results allow to prove, in the case of certain Sobolev kernels, that the algorithms have optimal stability and optimal convergence rates, including for functions outside the native space of the kernel. The results also apply to the case of the usual $P$-greedy algorithm, significantly improving state-of-the-art results available in the literature. Illustrative experiments are presented that support the theoretical findings and show improvements of the stabilized algorithms in terms of accuracy due to improved stability.


*Keywords:* Kernel methods; Greedy algorithms; Convergence rates

* Corresponding author.
*E-mail addresses:* tizian.wenzel@mathematik.uni-stuttgart.de (T. Wenzel), gsantin@fbk.eu (G. Santin), haasdonk@mathematik.uni-stuttgart.de (B. Haasdonk).

https://doi.org/10.1016/j.jat.2020.105508

## 1. Introduction

We start with introducing our central terminology and notation, which is well-known in kernel approximation literature, cf. [6,8,30]. For a nonempty set $\Omega \subset \mathbb{R}^d$ a real valued kernel is defined as a symmetric function $k : \Omega \times \Omega \to \mathbb{R}$. The *kernel matrix* $A \in \mathbb{R}^{N\times N}$ associated to this kernel and to an arbitrary set of points $X_N = \{x_1, .., x_N\} \subset \Omega$ is given by entries $A_{ij} = k(x_i, x_j)$. In the following the sets $X_N$ will always be assumed to consist of pairwise distinct points. A kernel $k$ is called *strictly positive definite* if the kernel matrix is positive definite for any such $X_N \subset \Omega$ and any $N \in \mathbb{N}$.

For every strictly positive definite kernel on $\Omega$ there is a unique *native space* $\mathcal{H}_k(\Omega)$ with inner product $(\cdot, \cdot)_{\mathcal{H}_k(\Omega)}$, i.e., a Hilbert space of functions 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. $(f, k(\cdot, x))_{\mathcal{H}_k(\Omega)} = f(x)\ \forall f \in \mathcal{H}_k(\Omega), \forall x \in \Omega$.

Strictly positive definite kernels can be used to interpolate functions. For this, consider a function $f \in \mathcal{H}_k(\Omega)$. Then there is a unique interpolant $s_N$ of $f$ on $X_N$, which is given by the orthogonal projection $\Pi_{V(X_N)}(f)$ of $f$ to the subspace $V(X_N) := \operatorname{span}\{k(\cdot, x_i), x_i \in X_N\}$. Additionally this interpolant is the minimum norm interpolant among all interpolants from $\mathcal{H}_k(\Omega)$. It can be expressed as

$$s_N(\cdot) = \Pi_{V(X_N)}(f) = \sum_{i=1}^{N} \alpha_i k(\cdot, x_i),$$

where the coefficients $\alpha = (\alpha_i)_{i=1}^N \in \mathbb{R}^N$ can be computed by solving the linear equation system

$$A\alpha = b, \tag{1}$$

with the kernel matrix $A$ of $k$ on $X_N$, and the vector $b = (f(x_i))_{i=1}^N$ of function values. Since the kernel is assumed to be strictly positive definite, the kernel matrix is positive definite and thus Eq. (1) possesses a unique solution.

A priori it is usually unclear how to choose an appropriate set of sampling points, although it is evident that this choice strongly influences the results of the interpolation process. To overcome this problem, often greedy methods are applied [4,21,25,28,33].

Greedy algorithms are largely studied and applied in the fields of approximation and computational mathematics. We refer for example to [29] for a general treatment of greedy methods. Convergence analysis for several greedy algorithms, also in reduced basis methods, can be found in [1,2,15,19].

In the present setting of kernel interpolation, the greedy algorithms start with an empty set $X_0 := \emptyset$ and then they iteratively add another interpolation point as $X_{N+1} := X_N \cup \{x_{N+1}\}$ at each step which is optimal with regard to some selection criterion $\eta^{(N)}$:

$$x_{N+1} = \operatorname*{arg\,max}_{x\in\Omega} \eta^{(N)}(x). \tag{2}$$

These greedy algorithms stop if the interpolant is exact or if some predefined stopping criterion is met, and they are practicable and fast.

In the greedy kernel literature there are three main selection criteria, namely $f$-greedy, $P$-greedy and $f/P$-greedy [4,21,28], that we will recall in the following. Depending on the

chosen selection criterion different interpolation point distributions may result, and they lead to different effects concerning convergence speed or stability of the numerical scheme.

The convergence is closely linked to the *fill distance* $h_{X_N,\Omega}$ which describes the largest ball which can be centered in $\Omega$ without intersecting any interpolation point $x_i \in X_N$. The *separation distance* $q_{X_N}$ measures instead the distance of the two closest interpolation points and it can be related to the stability of the interpolation procedure via the condition number of the kernel matrix. Thus we define

$$\begin{aligned} h_N &:= \quad h_{X_N,\Omega} := \sup_{x\in\Omega} \min_{x_i\in X_N} \|x - x_i\|_2, \\ q_N &:= \qquad q_{X_N} := \min_{x_i\neq x_j\in X_N} \|x_i - x_j\|_2. \end{aligned} \tag{3}$$

In the following, when no ambiguity is possible, we will use the shorthand notations $h_N$ and $q_N$.

Roughly speaking, a small fill distance implies a small error, while a large separation distance guarantees stability. If a sequence of sets of points satisfies $h_N \leq C \cdot q_N$ for some $C > 0$ and all $N \in \mathbb{N}$, then the *uniformity constant* $\rho_N := h_N/q_N \leq C$ is bounded and the sequence is called *asymptotically uniformly distributed* and, intuitively, it possesses both the desirable properties and it is thus particularly suitable for interpolation.

In this paper we introduce a new class of greedy kernel approximation algorithms, the so called *stabilized* or *restricted greedy algorithms*. They are obtained as a modification of existing selection strategies by introducing a restriction on the set of admissible points, which is steered by a parameter $\gamma \in (0, 1]$. This modification creates, for each selection rule, a scale of methods depending on $\gamma$.

We prove convergence results for the new scale of algorithms by controlling the decay of the power function associated with the selected points. This result is an extension of a result in [25], and proves that the power function of the new algorithms decays algebraically or exponentially, depending on known convergence rates for interpolation by uniform points.

In the case of certain kernels associated to Sobolev spaces, we also prove that the known decay rates are optimal up to constants, i.e., no better rate of decay of the power function can be obtained, even for points that are globally optimized instead of being selected iteratively.

We then use these upper and lower bounds on the decay of the power function to derive three groups of results. First, we obtain a refined version of a bound of [4] and use it in combination with an idea of [25] to obtain a precise decay of the fill distance of the points selected by the algorithms. This result is a strict improvement of the one of [25], and in particular it allows one to obtain error bounds for the interpolation error w.r.t. a generic $L^p$ norm (and not necessarily the $L^\infty$ norm) and for the approximation of the derivatives of the target function. These general rates of convergence coincide with the worst-case optimal ones.

Second, we modify an idea introduced in [4] to deduce precise lower bounds for the decay of the separation distance. Thanks to this result, we can guarantee a control on the condition number of the interpolation matrices and on the Lebesgue constant, and thus on the stability of the algorithms.

Third, combining the first two results we prove that the selected points are asymptotically uniform, whereby the degree of uniformity (as described in Eq. (24)) depends on the stabilization parameter $\gamma$. Furthermore, although point generation is not a primary issue of this paper, since these algorithms can be run very efficiently on very general geometries, as a side product they can be used to generate uniform point distributions for other tasks such as meshless PDE approximation.

Note that the asymptotic uniformity might suggest that these new methods do not have an advantage over other more simple uniform point generation procedures, e.g. by using geometric properties. We clearly want to emphasize the advantage of the new stabilized $f$ and $f/P$ methods in the preasymptotic range for function approximation, as they balance between function adaptivity and stability in terms of uniform point distribution, practically leading to better approximation than non-adaptive and non-stabilized algorithm, as exemplified in Section 7.2.

Moreover, a result of [23] applies to the interpolation by these new algorithms, and in particular we prove that they can be used to approximate with optimal rates of convergence functions which are outside the native space of the kernel.

All these results apply to any of the $\gamma$-stabilized greedy algorithms, with constants depending on $\gamma$. In particular, they cover the case of $P$-greedy, for which the results of this paper provide new or refined results. The $f$- and $f/P$-greedy selections are also covered. In this case the modified algorithms allow to obtain for the first time convergence rates that are at least worst-case optimal.

We remark in particular that this paper provides significant improvements over the results of [25] also in the case of the standard $P$-greedy algorithm. Indeed, we provide a refined result on the decay of the fill distance (Theorem 15), a completely new result on the decay of the separation distance (Theorem 19), about which no result was proven in [25], a new proof technique that applies under weaker conditions (Section 5.2.1), and finally several consequences of these new results (Section 6), which are impossible to prove using the results of [25] alone.

The paper is organized as follows. We start by recalling some additional details on kernel theory in Section 2, while in Section 3 the stabilized greedy algorithms are introduced. The following Section 4 gives precise bounds on the decay of the maximal power function values, and these bounds allow to prove in Section 5 that the resulting sampling points are distributed asymptotically uniformly by providing upper bounds on the fill distance and lower bounds on the separation distance. Using this uniformity, further results are drawn on the stability and convergence of the algorithms in Section 6. Section 7 concludes with some numerical experiments which complement the analytic results.

## 2. Kernel interpolation and greedy algorithms

To begin with we start with some additional background about kernel based approximation. A thorough introduction and more details can be found e.g. in the monographs [6,8,30] already mentioned in the previous section.

A way to measure the interpolation error $\|f - \Pi_{V(X_N)}(f)\|_{L^\infty}$ is given by the power function $P_N := P_{X_N} : \Omega \to \mathbb{R}$, which can be defined as the norm of the pointwise error functional, i.e.,

$$P_N(x) = \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{|f(x) - \Pi_{V(X_N)}(f)(x)|}{\|f\|_{\mathcal{H}_k(\Omega)}}. \tag{4}$$

Again we stick to the notation $P_N$ as long as there is no possible ambiguity concerning the set $X_N$.

It can be proven that it can be expressed in the following way:

$$P_N(x) = \|k(\cdot, x) - \Pi_{V(X_N)}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)}. \tag{5}$$

Reshaping Eq. (4) defining $P_N$ and taking the supremum norm directly yields the standard estimate on the interpolation error in the $L^\infty$ norm, namely

$$\|f - \Pi_{V(X_N)}(f)\|_{L^\infty(\Omega)} \leq \|P_N\|_{L^\infty(\Omega)} \cdot \|f\|_{\mathcal{H}_k(\Omega)}, \quad f \in \mathcal{H}_k(\Omega). \tag{6}$$

Instead of expressing the interpolant in terms of the basis of so called kernel translates $k(\cdot, x_i)$, it is sometimes useful to consider the Lagrange basis. The following proposition collects some classical results in this direction (see e.g. [30]).

**Proposition 1.** *Let $X_N \subset \Omega$ be pairwise distinct points. Then there exists a unique Lagrange basis $\{l_j\}_{j=1}^N$ of $V(X_N)$, i.e., $V(X_N) = \operatorname{span}\{l_j, 1 \leq j \leq N\}$ and $l_j(x_i) = \delta_{ij}$ for all $1 \leq i, j \leq N$. Furthermore, for any $f \in \mathcal{H}$ the unique interpolant of $f$ on $X_N$ can be expressed as*

$$\Pi_{V(X_N)}(f)(x) = \sum_{j=1}^{N} f(x_j) l_j(x) \quad \text{for all } x \in \Omega,$$

*and the square of the power function is given by*

$$P_N(x)^2 = k(x,x) - \sum_{j=1}^{N} k(x, x_j) l_j(x) \quad \text{for all } x \in \Omega.$$

*Moreover, the Lebesgue constant*

$$\Lambda_{X_N} := \max_{x \in \Omega} \sum_{j=1}^{N} |l_j(x)|$$

*gives an upper bound on the sampling stability of the interpolation process, i.e.*

$$\left\| \Pi_{V(X_N)}(f) \right\|_{L^\infty(\Omega)} \leq \Lambda_{X_N} \|f|_X\|_\infty \quad \text{for all } f \in \mathcal{H}.$$

As mentioned in the introduction, the success of this interpolation process depends crucially on the distribution of the points $X_N$ inside $\Omega$, which we will quantify by means of the fill distance and the separation distance defined in (3). Using simple geometric arguments, which relate the volume of the domain $\Omega$ or a surrounding ball to the sum of the volumes of balls around the points $x_i \in X_N$, one can conclude some basic estimates on the fill distance and on the separation distance (see e.g. [21]).

**Theorem 2.** *Let $\Omega \subset \mathbb{R}^d$ be bounded, $(X_N)_{N \in \mathbb{N}}$ be a sequence of sets of points within $\Omega$. Then there are constants $c_\Omega > 0, C'_\Omega > 0$ such that*

$$\begin{aligned} h_N &\geq c_\Omega \cdot N^{-1/d}, \\ q_N &\leq C'_\Omega \cdot N^{-1/d}. \end{aligned} \tag{7}$$

We recall that whenever for a given sequence of point sets there exists $C > 0$ such that additionally $h_N \leq C \cdot q_N$ holds for all $N \in \mathbb{N}$, then $(X_N)_{N \in \mathbb{N}}$ is said to be *asymptotically uniformly distributed*.

For the selection of these sampling points we are interested in greedy algorithms which, as mentioned, construct a *nested* sequence of points $(X_N)_{N \in \mathbb{N}}$ starting from $X_0 := \emptyset$ and updating it at each iteration as $X_{N+1} := X_N \cup \{x_{N+1}\}$ by selecting a point that maximizes a given selection criterion $\eta^{(N)}$ over $\Omega$. In particular, the commonly used $f$-, $P$-, and $f/P$-greedy selection rules use the following error indicators:

i. $f$-greedy: $\eta_f^{(N)}(x) = |f(x) - \Pi_{V(X_N)}(f)(x)|$
ii. $P$-greedy: $\eta_P^{(N)}(x) = P_N(x)$
iii. $f/P$-greedy: $\eta_{f/P}^{(N)}(x) = |f(x) - \Pi_{V(X_N)}(f)(x)| / P_N(x)$.

The first two criteria are clearly aiming at reducing the pointwise interpolation error either directly, or via (6) by reducing the power function. The $f/P$-greedy ("$f$ over $P$ greedy") selection, instead, combines the two and it can be proven to be locally optimal, i.e., it provides the best possible reduction of the interpolation error in the native space norm, at each iteration. However, for $f/P$-greedy it is necessary to perform the maximization over $\Omega \setminus X_N$ as the fraction is not well-defined in already selected points $X_N$. Indeed, both the numerator and denominator of $\eta^{(N)}_{f/P}$ vanish, and it has been proven in Theorem 1 in [33] that, for all $x_i \in X_N$, $\lim_{x\to x_i} \eta^{(N)}_{f/P}$ depends on the direction along which the limit is taken. In particular, $\eta^{(N)}_{f/P}$ cannot be extended by continuity to $X_N$.

We remark that here and in the following if the maximum in a selection rule is not unique then any of the points realizing the maximal value can be chosen arbitrarily. Whenever we will prove or recall results about the distribution of *the* sequence of points selected by a greedy algorithm we refer to *any* possible sequence.

So far our discussion applies to any given strictly positive definite kernel, but most of the analytical results of this paper are specialized to the remarkable case of kernels generating Sobolev spaces. To be more precise, we consider the class of translational invariant kernels, i.e., there exists a function $\Phi : \mathbb{R}^d \to \mathbb{R}$ such that the kernel can be expressed as $k(x, y) = \Phi(x - y)$. A special case is given by radial basis function kernels which can be expressed as $k(x, y) = \phi(\|x-y\|_2)$ with a *radial basis function* $\phi : \mathbb{R} \to \mathbb{R}$, i.e., $\Phi(x) := \phi(\|x\|)$. Depending on the Fourier transform of this function $\Phi$, the native space $\mathcal{H}_k(\Omega)$ can be characterized in terms of Sobolev spaces. That means that if there exist constants $c_\Phi, C_\Phi > 0$ and $\tau > d/2$ such that

$$c_\Phi(1 + \|\omega\|_2^2)^{-\tau} \leq \hat{\Phi}(\omega) \leq C_\Phi(1 + \|\omega\|_2^2)^{-\tau} \quad \forall \omega \in \mathbb{R}^d, \tag{8}$$

then the native space $\mathcal{H}_k(\mathbb{R}^d)$ is norm equivalent to the Sobolev space $W_2^\tau(\mathbb{R}^d)$. Under some mild conditions on the boundary this result also holds for domains $\Omega \subset \mathbb{R}^d$, which will be assumed in the following. Kernels whose native space is norm equivalent to such a Sobolev space $W_2^\tau$ will be called *kernels of finite smoothness* $\tau$.

Generally, on the Sobolev spaces $W_p^k(\Omega)$, we use the following notation to denote the usual (semi-)norms for $k \in \mathbb{N}$,

$$|u|_{W_p^k(\Omega)} := \left( \sum_{|\alpha|=k} \|D^\alpha u\|^p_{L^p(\Omega)} \right)^{1/p} \quad \text{for } 1 \leq p < \infty,$$
$$|u|_{W_\infty^k(\Omega)} := \sup_{|\alpha|=k} \|D^\alpha u\|_{L^\infty(\Omega)},$$

where $\alpha \in \mathbb{N}^d$ is a multiindex and $|\alpha| := \alpha_1 + \cdots + \alpha_d$ is its length.

Moreover, for this class of kernels error estimates are available. A very general way to introduce them is provided by the following Theorem 3. A first version of this theorem was first proved in [22, Theorem 2.12] and [32, Theorem 2.6] with slightly different assumptions on the indices. The improvements on these indices were justified in [23]. The final form stated here in Theorem 3 was taken from [14, Theorem 2.2], with notation adapted to this paper.

**Theorem 3.** *Suppose $\Omega \subset \mathbb{R}^d$ is a bounded domain satisfying an interior cone condition and having a Lipschitz boundary. Let $X \subset \Omega$ be a discrete set with sufficiently small fill distance $h = h_{X,\Omega}$. Let $\tau = k + s$ with $k \in \mathbb{N}, 0 \leq s < 1, 1 \leq p < \infty, 1 \leq q \leq \infty, m \in \mathbb{N}_0$ with*

*$k > m + d/p$ if $p > 1$ or $k \geq m + d/p$ if $p = 1$. Then for each $u \in W_p^\tau(\Omega)$ we have that*

$$|u|_{W_q^m(\Omega)} \leq C \left( h^{(\tau - m - d(1/p - 1/q)_+)} \cdot |u|_{W_p^\tau(\Omega)} + h^{-m} \|u|_X\|_\infty \right)$$

*where $C > 0$ is a constant independent of $u$ and $h$ and $(x)_+ = \max\{x, 0\}$.*

Replacing $u$ with the residual $f - \Pi_{V(X_N)}$ makes the second term on the right hand side vanish, and this idea is used to obtain error bounds for the error of the interpolation of $f$ and its derivatives, w.r.t. a suitable $L^q$ norm.

We recall that for these kernels of finite smoothness, several convergence results also for the three greedy algorithms are available. In particular, quasi-optimal rates have been proven in [25] for the $P$-greedy algorithm, while the results for $f$-greedy (see [21]) and for $f/P$-greedy (see [33]) are optimal only under some restrictive assumptions, and when these are not satisfied there is a significant gap between these proved rates and the one observed in numerical experiments. In particular, several applications of these algorithms (see e.g. [16,21,24,26,34]) have shown numerical evidence that the $f$- and $f/P$-greedy algorithms provide usually a faster convergence than $P$-greedy, and the commonly accepted explanation is that the first two are able to select points that are tailored to the specific target function. Nevertheless, there are no theoretical results that prove this heuristic fact, not even for special function classes. In this view, the gap in the theory is even larger. As mentioned in the introduction, one of the goals of this paper is to show that under small modifications, also these algorithms can be proven to have worst-case optimal convergence rates.

**Remark 4.** Although this paper is focused on the interpolation of scalar-valued functions, we would like to mention that most of the results can be extended to deal with vector-valued functions. Indeed, it is possible to define a notion of strictly positive definite matrix-valued kernel (see [20]), which allows to generalize the theory of kernels and native spaces to the case of vectorial functions. Among a vast variety of matrix-valued kernels, a particularly simple class is given by the so-called separable kernels, which are obtained by linear combinations of scalar kernels with positive definite matrices [34].

In particular, a very effective and common choice to approximate functions with values in $\mathbb{R}^q$ for some $q \geq 1$ is to consider a matrix valued kernel $K(x, y) := k(x, y)I$, where $I$ is the $q \times q$ identity matrix and $k$ is a standard scalar-valued kernel. This choice is one of the fundamental tools of the Vectorial Kernel Orthogonal Greedy Algorithm (VKOGA) of [33], which extends the greedy algorithms to vectorial functions by selecting a set of points which is shared over the $q$ components, thus resulting in an interpolant with fewer centers, which is hence faster to evaluate.

It has been proven in [34] (see Remark 1 and Lemma 3.8 in that paper) that the native space of this matrix-valued kernel $K$ is given by the tensor product of $q$ copies of the standard native space of the scalar-valued kernel $k$, and in particular that the power function of $K$ is defined as a vector-valued function whose components are identical and equal to the usual power function of $k$.

Thus, the upper and lower bounds on the scalar-valued power function and the related results obtained in this paper can immediately be translated to the vectorial case by component-wise application, if this matrix valued kernel $K := kI$ is used.

**Remark 5.** We remark that RBF kernels are usually defined using an additional shape or scale parameter $\varepsilon > 0$ that shrinks or flattens the kernel around its center, i.e., $K(x, y) := \phi(\varepsilon\|x -$

$y\|$). The appropriate choice of this parameter is often a crucial ingredient to obtain an accurate approximation, but we do not address this problem in this paper.

We are justified to do so because in the case of the kernels of finite smoothness, that are the main focus of the following theory, the change of $\varepsilon$ has the only effect of changing the norm of $\mathcal{H}_k(\Omega)$ to an equivalent one (see e.g. [31] for a discussion of this fact in the case of the Wendland kernels). This means in particular that in the error bounds that we are using (e.g. Theorem 3) a change of $\varepsilon$ would modify the multiplicative constant, but not the rate of convergence.

It is anyhow still worth noting that the greedy algorithm applied to the same kernel with two different shape parameters will very likely produce two different sequences of points. Nevertheless, in the following we will prove results on the asymptotic distribution of these sequences, and these are in any case quasi-uniform independently of $\varepsilon$.

## 3. Stabilized greedy algorithms

As a motivation for the introduction of stabilized greedy algorithms one can observe that the $f/P$-greedy algorithm does not need to be well defined. The reason is that there might not exist a maximum of $|f - \Pi_{V(X_N)}(f)|/P_N$ within $\Omega \setminus X_N$. This set is open, so the supremum of $|f - \Pi_{V(X_N)}(f)|/P_N$ needs not to be attained, and thus the greedy selection rule is not able to choose a point. This situation is presented in the following example.

**Example 6.** Consider the interval $[0, 1]$ and the unscaled linear Matérn kernel $k(x, y) = (1 + |x - y|) \cdot \exp(-|x - y|)$, and take

$$f(x) := -x + x^2 + k(x, 0) \overset{x \geq 0}{=} -x + x^2 + (1 + x) \cdot \exp(-x).$$

The maximum of $f$ is attained in 0, thus the residual $r_1$ is given by

$$r_1(x) := f(x) - s_1(x) = f(x) - f(0) \cdot k(x, 0) = -x + x^2.$$

The power function $P_1$ is given by $P_1(x) = \sqrt{1 - (1 + x)^2 \cdot \exp(-2x)}$. The calculation of $\lim_{x \searrow 0} |r_1|^2/P_1^2$ with l'Hôpital's rule shows that $\lim_{x \searrow 0} \frac{r_1(x)^2}{P_1(x)^2} = 1$. Furthermore for $0 < x \leq 1$ it can be estimated that $r_1(x)^2/P_1(x)^2 < 1$. Thus $|r_1|/P_1$ approaches its supremum for $x \to 0$, but $0 \notin \Omega \setminus X_1$ since $X_1 = \{x_1\} = \{0\}$. Thus the $f/P$ greedy procedure is not well defined.

A plot of the function $f$, the residual $r_1$, the power functions $P_0$, $P_1$ and the ratio $|r_1|/P_1$ as well as the first interpolation point $x_1$ is given in Fig. 1.

In numerical implementations the point selection is performed on a discretization of the set $\Omega$, such that the algorithm will still work. Nevertheless it might still happen that the next point is chosen close to an old interpolation point such that the separation distance drops significantly, which worsens the stability of the interpolation procedure.

One way to circumvent this limitation is given by the *$\gamma$-stabilized* or *$\gamma$-restricted* greedy algorithms, which will be introduced in the following definition. The notation *restricted* will be clear after the definition, while the notation *stabilized* will be explained in Section 5 based on derived stability results.

**Definition 7.** Let $X_0 := \emptyset$ and $0 < \gamma \leq 1$. For every $N \in \mathbb{N} \cup \{0\}$, define $\Omega_\gamma^{(N)} := \{x \in \Omega : P_N(x) \geq \gamma \cdot \|P_N\|_\infty\}$.

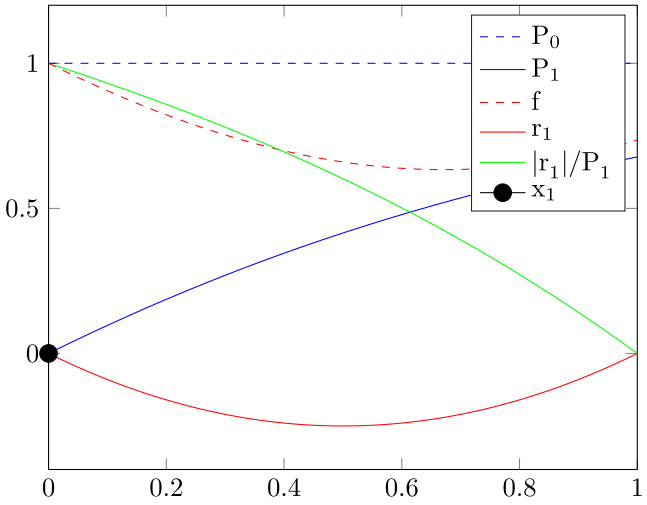


**Fig. 1.** Motivating example for the $\gamma$-stabilized algorithm based on the linear Matérn kernel: The maximum of $|r_1|/P_1$ is attained in $x_1$, which is not element of $\Omega \setminus X_1$. The horizontal axis describes the domain $\Omega = [0, 1]$ whereas the vertical axis describes the function values.

Then the $\gamma$-stabilized or $\gamma$-restricted greedy selection criterion is defined as

$$x_{N+1} = \arg\max_{x \in \Omega_\gamma^{(N)}} \eta^{(N)}(x),$$

where $\eta^{(N)} : \Omega \to \mathbb{R}_{\geq 0}$ is some given selection criterion.

The naming *restricted* of the selection criterion is clearly based on the restriction of the admissible set of points from the set $\Omega$ to $\Omega_\gamma^{(N)}$. Moreover in the following, when we want to focus on the restricted set $\Omega_\gamma^{(N)}$ of admissible points while the concrete error indicator $\eta^{(N)}$ is not of decisive importance, we will denote this as *any $\gamma$-stabilized algorithm.*

We stress that the definition can be applied to any already existing selection criterion. This yields for example the $\gamma$-stabilized $f/P$-greedy algorithm, which selects the next points according to

$$x_{N+1} = \arg\max_{x \in \Omega_\gamma^{(N)}} \frac{|(f - \Pi_{V(X_N)}(f))(x)|}{P_N(x)}.$$

Moreover, for $\gamma = 0$ it holds $\Omega_\gamma^{(N)} = \Omega$, thus there is no restriction at all and the algorithm coincides with the standard greedy algorithm with selection criterion $\eta^{(N)}$. For example, the $\gamma$-stabilized $f/P$-greedy reduces in this case to the usual $f/P$-greedy algorithm. That is the reason why the case $\gamma = 0$ is excluded from the definition.

For $\gamma = 1$, instead, it holds $\Omega_1^{(N)} = \{x \in \Omega : P_N(x) = \|P_N\|_\infty\}$, i.e., the selection criterion reduces to a special realization of the $P$-greedy criterion: In case of multiple points realizing the maximal power function value, a point with the maximal $\eta^{(N)}$ value is picked.

All together, these stabilized algorithms provide all the intermediate cases between $\gamma = 0$ and $\gamma = 1$, i.e., they interpolate between standard, known methods to produce a new scale of

selection criteria with a potentially significantly increased flexibility due to a combination of the properties of two selection criteria.

We remark that a $\gamma$-stabilized $P$-greedy selection rule is still the same as the standard $P$-greedy selection rule, since one of the points realizing the maximal power function value is chosen. That means the $P$-greedy selection rule is not affected by the stabilization.

## 4. Convergence rate of the power function

In this section we give two results. First we prove an upper bound on the decay of the maximal value of the power function by generalizing results of [25]. This, using the bound (6), immediately leads to convergence of the interpolation procedure with the help of any $\gamma$-stabilized greedy algorithm, independently of the chosen selection rule. Then we prove a lower bound on the decay of the maximal value of the Power function, which implies it cannot drop arbitrarily fast. Since the lower and upper bounds coincide up to constants, we can conclude that they are quasi-optimal.

Note that in the next theorem, and in most of the results of this paper, the bounds contain constants that depend on negative powers of $\gamma$, i.e., a smaller restriction parameter yields a larger constant. For the limit $\gamma \to 0$ this factor tends to infinity, thus no statements on the unstabilized algorithms are possible. In all other cases the restriction parameter is positive and a fixed parameter of the problem.

The proof of the following Theorem 8 is along the lines of the analysis within [25], with some modifications required to include the effect of the parameter $\gamma$.

**Theorem 8.** *Assume that $\Omega \subset \mathbb{R}^d$ is a compact domain which satisfies an interior cone condition and has a Lipschitz boundary. Suppose that $k$ is a translational invariant kernel of finite smoothness $\tau > d/2$. Then any $\gamma$-stabilized algorithm applied to a function in $f \in \mathcal{H}_k(\Omega)$ gives a sequence of point sets $X_N \subset \Omega$ such that*

$$\|P_N\|_{L^\infty(\Omega)} \leq C_P \cdot \gamma^{-2} \cdot N^{\frac{1}{2}-\frac{\tau}{d}} \tag{9}$$

*holds with $C_P = 2^{5\tau/d-1/2} \cdot c_1$. The constant $c_1$ is independent of $\gamma$ and $N$. In particular, it holds*

$$\|f - \Pi_{V(X_N)}(f)\|_{L^\infty(\Omega)} \leq C_P \cdot \gamma^{-2} \cdot N^{\frac{1}{2}-\frac{\tau}{d}} \|f\|_{\mathcal{H}_k(\Omega)}. \tag{10}$$

**Proof.** To prove the decay rate, one does not need the concrete selection criterion, it is only necessary to focus on the constraint given by the $\gamma$-restriction to $\Omega_\gamma^{(N)} \equiv \{x \in \Omega \mid P_N(x) \geq \gamma \cdot \|P_N\|_\infty\}$, i.e., the case where in any step a point is chosen arbitrarily from the restricted set. this case includes the proof for the decay rates of the $\gamma$-stabilized algorithms.

Since the proof is based on a Corollary from [5] and a linkage which was used in [25], an overview will be given in the following:

i. To begin, we describe the setting within [5] which will be called *abstract setting*. Although the cited paper deals with general Banach spaces, we report the results already adapted to the Hilbert space setting.
Consider a Hilbert space $(\mathcal{H}, \|\cdot\|)$ and a compact subset $\mathcal{F} \subset \mathcal{H}$ which is assumed to satisfy $\|f\| \leq 1 \ \forall f \in \mathcal{F}$. For elements $f_1, .., f_N \in \mathcal{F}$ define $V_N := \text{span}\{f_1, .., f_N\} \subset \mathcal{H}$. The question is how well does the subspace $V_N$ approximate $\mathcal{F}$. For this reason the

following two definitions are introduced:

$$\sigma_N := \sigma_N(\mathcal{F})_{\mathcal{H}} := \sup_{f \in \mathcal{F}} \|f - \Pi_{V_N}(f)\|_{\mathcal{H}},$$

$$d_N := d_N(\mathcal{F})_{\mathcal{H}} := \inf_{\substack{Y \subset \mathcal{H} \\ \dim(Y)=N}} \sup_{f \in \mathcal{F}} \|f - \Pi_Y(f)\|_{\mathcal{H}}.$$

In the literature, $d_N(\mathcal{F}_{\mathcal{H}})$ is usually called the *Kolmogorov N-width*. For $N = 0$ we set $Y = \{0\}$ respective $V_N = \{0\}$. Moreover, the subscript $\mathcal{H}$ will be mostly dropped, and the quantity $\sigma_N$ is further defined for single elements $f \in \mathcal{H}$ via $\sigma_N(f) := \sigma_N(\{f\})$.
The following algorithm, which was called weak greedy algorithm with constant $\gamma$ in a Hilbert space $\mathcal{H}$, was investigated in [5]:

- First step: Choose $f_1 \in \mathcal{F}$ such that $\sigma_0(f_1) = \|f_1\| \overset{!}{\geq} \gamma \cdot \sigma_0(\mathcal{F})_{\mathcal{H}}$.
- Iterative step: Given $\{f_1, .., f_N\}$ already chosen, and defining $V_N := \text{span}\{f_1, .., f_N\}$, choose $f_{N+1} \in \mathcal{F}$ such that

$$\sigma_N(f_{N+1}) = \|f_{N+1} - \Pi_{V_N}(f_{N+1})\| \overset{!}{\geq} \gamma \cdot \sigma_N(\mathcal{F}).$$

The following statement is a result within [5], namely Corollary 3.3, (ii). It will be used for this proof:

($\star$) For the weak greedy algorithm with constant $\gamma$ in a Hilbert space $\mathcal{H}$ we have the following:

(ii) If $d_N(\mathcal{F}) \leq C_0 \cdot N^{-\alpha}\ \forall N \in \mathbb{N}$, then it holds $\sigma_N(\mathcal{F}) \leq C_1 \cdot N^{-\alpha}\ \forall N \in \mathbb{N}$ with $C_1 = 2^{5\alpha+1} \cdot \gamma^{-2} \cdot C_0$.

ii. The paper [25] used this previously stated result and applied it to the setting of kernel interpolation, which will be called *kernel setting* in the following.
For this purpose choose $\mathcal{H} = \mathcal{H}_k(\Omega)$ and $\mathcal{F} = \{k(\cdot, x) : x \in \Omega\}$. Without loss of generality, we assume $\|k(\cdot, x)\|_{\mathcal{H}_k(\Omega)} \leq 1$, otherwise consider the kernel normalized with $\|k(\cdot, x)\|_{\mathcal{H}_k(\Omega)} = k(x, x)^{1/2} \equiv \text{const}$. The fact that the choice of $\mathcal{F}$ satisfies the requirements can be seen in Lemma 3.1. of [25]. The choice of $\mathcal{F} = \{k(\cdot, x) : x \in \Omega\}$ means that any $f = k(\cdot, x) \in \mathcal{F}$ can be uniquely associated with an $x \in \Omega$ and vice versa. This is the key ingredient to link the abstract setting with the kernel setting. Thus it holds $V_N \equiv \text{span}\{f_1, .., f_N\} = \text{span}\{k(\cdot, x_i), x_i \in X_N\}$ where $X_N$ is the set of points which correspond to the functions $f_1, .., f_N$ of the abstract setting. Furthermore for $f \in \mathcal{F}$ we get

$$\begin{aligned}\sigma_N(f) \equiv &\|f - \Pi_{V_N}(f)\|_{\mathcal{H}_k(\Omega)} = \|k(\cdot, x) - \Pi_{V_N}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)} \\ = &P_N(x), \end{aligned} \tag{11}$$

$$\begin{aligned}\sigma_N(\mathcal{F}_h) \equiv &\sup_{f \in \mathcal{F}_h} \|f - \Pi_{V_N}(f)\|_{\mathcal{H}_k(\Omega)} = \sup_{x \in \Omega} \|k(\cdot, x) - \Pi_{V_N}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)} \\ = &\|P_N\|_{L^\infty(\Omega)}, \end{aligned} \tag{12}$$

where the representation of the power function from Eq. (5) was used.

iii. Now we turn to the current algorithm:

- First step: Choose $f_1 = k(\cdot, x_1) \in \mathcal{F} \equiv \{k(\cdot, x) : x \in \Omega\}$ such that $\|f_1\| = \|k(\cdot, x_1)\| \geq \gamma \cdot \sigma_0(\mathcal{F}) = \gamma \cdot \|P_0\|_{L^\infty(\Omega)}$.
- Iterative step: $\{f_1, .., f_N\} = \{k(\cdot, x_1), .., k(\cdot, x_N)\} \subset \mathcal{F}$ is already chosen, and $V_N = \text{span}\{f_1, .., f_N\} = \text{span}\{k(\cdot, x_1), .., k(\cdot, x_N)\}$. This means that we have already chosen points $\{x_1, .., x_N\} =: X_N$. Now choose $f_{N+1} \in \mathcal{F}$ (associated with $x_{N+1}$) such that

$$\sigma_N(f_{N+1}) \geq \gamma \cdot \max_{f \in \mathcal{F}} \sigma_N(f) = \gamma \cdot \sigma_N(\mathcal{F}),$$

$$\overset{(11)}{\Leftrightarrow} P_N(x_{N+1}) \geq \gamma \cdot \max_{x \in \Omega} P_N(x) = \|P_N\|_{L^\infty(\Omega)},$$

  where Eq. (11) was used. Thus weak greedy algorithms with constant $\gamma$ in the Hilbert space $\mathcal{H}_k(\Omega)$ exactly correspond to $\gamma$–stabilized greedy kernel algorithms which select any point within $\Omega_\gamma^{(N)}$.

Thus the results from the abstract setting can be used, especially the Corollary ($\star$) which was stated in the end of the first point. The requirements on $d_N(\mathcal{F})$ are satisfied due to Lemma 3.1. in [25], which states $d_N(\mathcal{F}) \leq c_1 \cdot N^{-\alpha_1}$. Thus the same decay can be concluded for any $\gamma$-stabilized greedy algorithm. In detail, using the stated decay rates for $d_N(\mathcal{F})$ from Lemma 3.1. in [25] it holds for kernels $k$ with finite smoothness $\tau > d/2$:

$$\|P_N\|_{L^\infty(\Omega)} \leq C_P \cdot \gamma^{-2} \cdot N^{\frac{1}{2} - \frac{\tau}{d}}$$

with $C_P = 2^{5\tau/d - 1/2} \cdot c_1$ (since $\alpha = \tau/d - 1/2$).

Thus Inequality (9) is proven and an application of (6) yields the bound (10). □

**Remark 9.** Observe that Theorem 8 is formulated in terms of a fixed function $f \in \mathcal{H}_k(\Omega)$. This is done since a general selection rule $\eta^{(N)}$ may be function dependent, and thus a target function needs to be specified. Nevertheless, the error bound (10) is based on the Power function only, and thus it applies also to any $g \in \mathcal{H}$ with $g \neq f$.

Moreover, if a function independent selection rule is used, there is no need to specify target values to run the corresponding $\gamma$-stabilized version of the algorithm.

Theorem 8 is the only result of this paper that applies not only to the class of kernels of finite smoothness, and it can indeed be extended to any kernel for which there exist known algebraic or exponential convergence results for fixed point distributions, which can be transferred to the ones of the $\gamma$-greedy power function. We refer to [25] for more details on this idea, and we mention here only the following additional result, which covers the case of the Gaussian kernel.

**Theorem 10.** *Assume that $\Omega \subset \mathbb{R}^d$ is a compact domain which satisfies an interior cone condition and has a Lipschitz boundary. Suppose that $k$ is a translational invariant kernel of infinite smoothness like the Gaussian kernel. Then any $\gamma$-stabilized algorithm applied to a*

*function in $f \in \mathcal{H}_k(\Omega)$ gives a sequence of point sets $X_N \subset \Omega$ such that*

$$\|P_N\|_{L^\infty(\Omega)} \leq c_2 \cdot \gamma^{-1} \cdot e^{-c_3 \cdot N^{-1/d}}$$

*holds. The constants $c_2$ and $c_3$ are independent of $\gamma$ and $N$. In particular, we obtain*

$$\|f - \Pi_{V(X_N)}(f)\|_{L^\infty(\Omega)} \leq c_2 \cdot \gamma^{-1} \cdot e^{-c_3 \cdot N^{-1/d}} \|f\|_{\mathcal{H}_k(\Omega)}. \tag{13}$$

The next Theorem 12 gives a lower bound on the decay of the maximal value of the power function. The proof requires the following preliminary Lemma 11, which estimates the native space norm of a bump function. It is part of a proof within [3, Theorem 1], where we corrected a minor mistake in an exponent.

**Lemma 11.** *Let $\Psi \in C^\infty(\mathbb{R}^d)$ be a bump function, i.e., having support within the unit ball and satisfying $\Psi(0) = 1$, $\|\Psi\|_{L^\infty(\mathbb{R}^d)} = 1$. Consider a bounded domain $\Omega \subset \mathbb{R}^d$ satisfying an interior cone condition and a native space which is norm equivalent to $W_2^\tau(\Omega)$ with $\tau > \frac{d}{2}$. Then there exists a value $0 < p_0 < 1$ such that the native space norm of a rescaled version of $\Psi$ with $0 < p \leq p_0$ can be estimated as*

$$\left\| \Psi\left(\frac{\cdot}{p}\right) \right\|_{\mathcal{H}_k(\Omega)} \leq C \cdot p^{\frac{d}{2}-\tau} \cdot \|\Psi\|_{\mathcal{H}_k(\Omega)}.$$

Now the following Theorem 12 states a lower bound on the decay of the power function. The result holds in general for any distribution of the points and hence especially for the point distribution created by any greedy algorithm.

**Theorem 12.** *Assume that $k$ is a translational invariant kernel with finite smoothness $\tau > d/2$. Let $\Omega \subset \mathbb{R}^d$ be bounded, $(X_N)_{N\in\mathbb{N}}$ be a sequence of sets of points within $\Omega$. Then there exist a value $h_0 > 0$ and constants $\tilde{c}_P, c_P > 0$ such that if $h_N \leq h_0$ then*

$$\|P_N\|_{L^\infty(\Omega)} \geq \tilde{c}_P \cdot h_N^{\tau-\frac{d}{2}}$$

*and moreover*

$$\|P_N\|_{L^\infty(\Omega)} \geq c_P \cdot N^{\frac{1}{2}-\frac{\tau}{d}}.$$

**Proof.** The setting of Lemma 11 with $p := h_N$ is considered. Thus the following inequality on the native space norm of a bump function rescaled and shifted to some point $\tilde{x} \in \Omega$ holds:

$$\frac{1}{\left\| \Psi\left(\frac{\cdot-\tilde{x}}{h_N}\right) \right\|_{\mathcal{H}_k(\Omega)}} \geq \frac{1}{C} \cdot \frac{h_N^{\tau-d/2}}{\|\Psi\|_{\mathcal{H}_k(\Omega)}}. \tag{14}$$

Since the bump function is smooth and the native space is norm equivalent to a Sobolev space, the bump function and any rescaled and shifted version of it are in the native space. Moreover, if we fix a set $X_N \subset \Omega$ of points, since the rescaled function $\Psi_{h_N,\tilde{x}} := \Psi((\cdot-\tilde{x})/h_N)$ has only support within a ball of radius $h_N$, it can be placed among the sampling points such that no sampling point lies in the interior of the support of $\Psi_{h_N,\tilde{x}}$.[1] Thus the interpolant to

[1] In fact it is needed to rescale with $(1+\epsilon)h_N$ with $\epsilon > 0$ fixed, e.g. $\epsilon = 0.01$. The reason is that the supremum within the definition of the fill distance needs not to be attained. This means we choose $\tilde{x} \in \Omega$ such that $h_N \equiv \sup_{x\in\Omega} \min_{x_i\in X_N} \|x - x_i\|_2 \leq (1+\epsilon) \cdot \min_{x_i\in X_N} \|\tilde{x} - x_i\|_2$. For the sake of simplicity we drop these technical details here.

$\Psi_{h_N,\tilde{x}}$ is the zero function. Therewith one can conclude:

$$\begin{aligned} P_N(x) &= \sup_{0\neq f\in\mathcal{H}_k(\Omega)} \frac{|(f-\Pi_{V(X_N)}(f))(x)|}{\|f\|_{\mathcal{H}_k(\Omega)}} \geq \frac{\Psi_{h_N,\tilde{x}}(x)-\Pi_{V(X_N)}(\Psi_{h_N,\tilde{x}})(x)}{\|\Psi_{h_N,\tilde{x}}\|_{\mathcal{H}_k(\Omega)}} \\ &= \frac{\Psi_{h_N,\tilde{x}}(x)}{\|\Psi_{h_N,\tilde{x}}\|_{\mathcal{H}_k(\Omega)}} \\ \Rightarrow \|P_N\|_\infty &= \sup_{x\in\Omega} \frac{\Psi_{h_N,\tilde{x}}(x)}{\|\Psi_{h_N,\tilde{x}}\|_{\mathcal{H}_k(\Omega)}} = \frac{1}{\|\Psi_{h_N,\tilde{x}}\|_{\mathcal{H}_k(\Omega)}} \overset{(14)}{\geq} \frac{1}{C}\cdot\frac{h_N^{\tau-d/2}}{\|\Psi\|_{\mathcal{H}_k(\Omega)}}. \end{aligned}$$

Setting $\tilde{c}_P := 1/\left(C\cdot\|\Psi\|_{\mathcal{H}_k(\Omega)}\right)$ proves the first inequality.

Moreover, using the lower bound on the fill distance $h_N \geq c_\Omega \cdot N^{-1/d}$ from Theorem 2 and remembering that $\tau > d/2 \Leftrightarrow \tau - d/2 > 0$ finally yields

$$\|P_N\|_\infty \geq \tilde{c}_P\cdot c_\Omega^{\tau-d/2}\cdot\left(N^{-1/d}\right)^{\tau-d/2} = \tilde{c}_P\cdot c_\Omega^{\tau-d/2}\cdot N^{\frac{1}{2}-\frac{\tau}{d}}.$$

Setting $c_P := \tilde{c}_P\cdot c_\Omega^{\tau-d/2}$ gives the desired result. □

This result can be applied to the case of the $\gamma$-stabilized greedy algorithms, i.e., to the sequence of points obtained by such an algorithm. Thus it gives a lower bound on the decay of the power function. If one compares this lower bound with the upper bound within Theorem 8, it is clear that the decay rates coincide. Thus, as mentioned in the beginning of this section, both results give exact rates and cannot be improved up to constants. Especially, this means that the optimal decay rate for the maximal value of the power function can be achieved with greedy methods, for example with the proposed $\gamma$-stabilized greedy method.

## 5. Uniformity of the selected points

Based on the previous results some statements about the point distribution of the selected points using kernels of finite smoothness $\tau$ can be derived. The first subsection gives an upper estimate on the fill distance, whereas the second subsection derives lower estimates on the separation distance. These results prove that the points selected by any $\gamma$-stabilized greedy algorithm are asymptotically uniformly distributed.

For the $P$–greedy algorithm, it was supposed in [4] that this property also holds true, but it was not proven up to now.

### 5.1. Estimate on the fill distance

For a first result on the fill distance we use the following Lemma 13, which is a refinement of Theorem 3.1 in [3].

**Lemma 13.** *Let $\Omega$ be a compact domain in $\mathbb{R}^d$ satisfying an interior cone condition. Suppose that the kernel $k$ is a translational invariant kernel with finite smoothness $\tau$. Then there exists a constant $M > 0$ with the following property: If $\epsilon > 0$ and $X_N = \{x_1, .., x_N\} \subset \Omega$ are given such that*

$$\|f - \Pi_{V(X_N)}(f)\|_{L^\infty} \leq \epsilon\cdot\|f\|_{\mathcal{H}_k(\Omega)} \text{ for all } f\in\mathcal{H}_k(\Omega), \tag{15}$$

*then the fill distance of $X_N$ satisfies*

$$h_N \leq M\cdot\epsilon^{1/(\tau-d/2)}.$$

**Proof.** The same argumentation as in the proof of Theorem 12 is used, i.e., the bump function $\Psi$ from Lemma 11 as well as its rescaled and shifted function $\Psi_{h_N} \equiv \Psi((\cdot - \tilde{x})/h_N)$ are considered. The interpolant $\Pi_{V(X_N)}(\Psi_{h_N})$ to $\Psi_{h_N}$ is again the zero function, so we can conclude

$$\begin{aligned} 1 = \|\Psi_{h_N}\|_{L^\infty} &= \|\Psi_{h_N} - \Pi_{V(X_N)}(\Psi_{h_N})\|_\infty \leq \epsilon \cdot \|\Psi_{h_N}\|_{\mathcal{H}_k(\Omega)} \\ &\leq C \cdot \epsilon \cdot h_N^{d/2-\tau} \cdot \|\Psi\|_{\mathcal{H}_k(\Omega)}. \end{aligned}$$

where in the last line Lemma 11 was used. This can be rearranged to conclude the result

$$\begin{aligned} & h_N^{\tau-d/2} \leq C \cdot \epsilon \cdot \|\Psi\|_{\mathcal{H}_k(\Omega)} \\ \Leftrightarrow \quad & h_N \leq M \cdot \epsilon^{1/(\tau-d/2)} \end{aligned}$$

with $M := (C \cdot \|\Psi\|_{\mathcal{H}_k(\Omega)})^{1/(\tau-d/2)}$. □

**Remark 14.** We remark that the original proof of Theorem 3.1 in [4] provided, under the same assumptions, a bound $h_N \leq M_\alpha \cdot \epsilon^{1/(\alpha-d/2)}$ for all $\alpha > \tau$, though not for $\alpha = \tau$.

Observe that any bound on the power function yields a way to satisfy the condition stated in Inequality (15) with $\epsilon = \|P_N\|_{L^\infty(\Omega)}$. This can be combined with the upper bound on the decay of the power function from Inequality (9) to conclude the following estimate on the fill distance.

**Theorem 15.** *Assume that $k$ is a translational invariant kernel with finite smoothness $\tau$. Then there exists a constant $c > 0$ with the following property: Any $\gamma$-stabilized algorithm applied to a function on a compact set $\Omega \subset \mathbb{R}^d$ which satisfies an interior cone condition and has a Lipschitz boundary gives a sequence of point sets $X_N \subset \Omega$ such that it holds*

$$h_N \leq c \cdot \gamma^{-2/(\tau-d/2)} \cdot N^{-1/d}.$$

This upper bound has a decay of $N^{-1/d}$. Recalling Theorem 2 which gives a lower bound with the same decay of $N^{-1/d}$ shows that these decay rates are exact and cannot be improved further.

### 5.2. Estimate on the separation distance

Based on the lower bound of the decay of the power function, even statements on the separation distance are possible.

In the following we prove two different results. First, in Theorem 12 we obtain an optimal rate of decay of the separation distance. Since the proof is mainly based on the application of the mean value theorem, further smoothness is needed, namely $\tau > d/2 + 1$.

To circumvent this limitation we show another proof strategy in Theorem 20, which gives, however, rates which are optimal only under certain conditions, which are discussed at the end of this section.

First of all, the following simple Lemma 16 is stated. Since the proof easily follows by induction on $N$, we omit it here.

**Lemma 16.** *If there exist constants $c, \alpha > 0$, such that for all $N \in \mathbb{N}$ it holds $\operatorname{dist}(\Omega_\gamma^{(N)}, X_N) \geq c \cdot N^{-\alpha}$, then it follows $q_{N+1} \geq c \cdot N^{-\alpha}$ $\forall N \in \mathbb{N}$.*

To prove the first main result Theorem 19, an estimate on the derivative of the residual is needed. This is achieved by using Theorem 3 with $X = X_N$ for $u \in W_p^\tau(\Omega)$ satisfying $u|_{X_N} = 0$ and setting, for $\tau > d/2$, the values $m = 1, q = \infty, s = 0, k = \tau, p = 2$. Thus the following lemma holds.

**Lemma 17.** *Suppose $\Omega \subset \mathbb{R}^d$ is a bounded domain satisfying an interior cone condition and having a Lipschitz boundary. Let $X_N \subset \Omega$ be a discrete set with sufficiently small fill distance $h = h_{X_N,\Omega}$. Let $\tau > 1 + d/2$. Then for each $u \in W_2^\tau(\Omega)$ with $u|_{X_N} = 0$ we have that*

$$|u|_{W_\infty^1(\Omega)} \equiv \sup_{|\alpha|=1} \|D^\alpha u\|_\infty \leq C h^{\tau-1-d/2} \cdot |u|_{W_2^\tau(\Omega)}, \tag{16}$$

*where $C > 0$ is a constant independent of $u$ and $h$.*

Moreover, in the proof of Theorem 19 we need a bound on the derivative of the power function. This will turn out to involve a generalized power function, which can be used to bound the error in the interpolation of derivatives of $f$, exactly as the standard power function bounds the error in the interpolation of pointwise values of $f$. We recall a construction of this bound in the following theorem. Observe that the condition on the functionals is in particular realized if $k$ is a translational invariant kernel with smoothness $\tau > d/2 + 1$ (see e.g. Chapter 16 in [30]).

**Lemma 18.** *Suppose $\Omega \subset \mathbb{R}^d$ and let $k$ be a strictly positive definite kernel. Let $\delta_x \circ \partial_i : \mathcal{H}_k(\Omega) \to \mathbb{R}$ be the linear functional given by the evaluation in $x \in \Omega$ of the derivative in the $i$th coordinate direction.*

*If $\delta_x \circ \partial_i \in \mathcal{H}_k(\Omega)'$ for all $x \in \Omega$ and for all $i = 1, .., d$, then it holds that*

$$\begin{aligned} P_N(\delta_x \circ \partial_i) &:= \left\| \partial_i^1 k(x, \cdot) - \sum_{j=1}^N \partial_i^1 k(x, x_j) l_j \right\|_{\mathcal{H}_k(\Omega)} \\ &= \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{|\partial_i f(x) - \partial_i \Pi_{V(X_N)} f(x)|}{\|f\|_{\mathcal{H}_k(\Omega)}}, \end{aligned} \tag{17}$$

*where $\{l_j\}_{j=1}^N$ denotes the Lagrange basis of Proposition 1.*

**Proof.** We have that the composition of the point evaluation with the partial derivative is a continuous functional, and the corresponding Riesz representer is $v_\lambda = \partial_i^1 k(x, \cdot)$, i.e., it holds $\lambda(f) = (\delta_x \circ \partial_i)(f) = (f, v_\lambda)_{\mathcal{H}_k(\Omega)} = (\partial_i f)(x)$. Now the proof is quite similar to the proof which shows the equivalence of the two expressions in (17) above for the ordinary power function, i.e., without the partial derivative.

Recall from Proposition 1 that the interpolant of a function $g$ can be expressed in terms of the Lagrange basis as $\Pi_{V(X_n)} g(\cdot) = \sum_{j=1}^N g(x_j) l_j(\cdot)$. Set $g = v_\lambda$ and calculate

$$\begin{aligned} \lambda(f) - \lambda(\Pi_{V(X_N)}(f)) &= (v_\lambda, (\mathrm{Id} - \Pi_{V(X_N)})(f))_{\mathcal{H}_k(\Omega)} \\ &= ((\mathrm{Id} - \Pi_{V(X_N)})(v_\lambda), f)_{\mathcal{H}_k(\Omega)} \\ &= (v_\lambda - \sum_{j=1}^N v_\lambda(x_j) l_j(\cdot), f)_{\mathcal{H}_k(\Omega)}. \end{aligned}$$

Thus it holds

$$\begin{aligned}\sup_{0\neq f\in\mathcal{H}_k(\Omega)} \frac{|\partial_i f(x) - \partial_i \Pi_{V(X_N)} f(x)|}{\|f\|_{\mathcal{H}_k(\Omega)}} &= \sup_{0\neq f\in\mathcal{H}_k(\Omega)} \frac{|\lambda(f) - \lambda(\Pi_{V(X_N)}(f))|}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &= \sup_{0\neq f\in\mathcal{H}_k(\Omega)} \frac{|(v_\lambda - \sum_{j=1}^N v_\lambda(x_j) l_j(\cdot), f)_{\mathcal{H}_k(\Omega)}|}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &\leq \sup_{0\neq f\in\mathcal{H}_k(\Omega)} \frac{\|v_\lambda - \sum_{j=1}^N v_\lambda(x_j) l_j(\cdot)\|_{\mathcal{H}_k(\Omega)} \cdot \|f\|_{\mathcal{H}_k(\Omega)}}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &= \left\| v_\lambda - \sum_{j=1}^N v_\lambda(x_j) l_j(\cdot) \right\|_{\mathcal{H}_k(\Omega)} \\ &= \left\| \partial_i k(x,\cdot) - \sum_{j=1}^N \partial_i^1 k(x, x_j) l_j(\cdot) \right\|_{\mathcal{H}_k(\Omega)}.\end{aligned}$$

Equality is attained by choosing $f = v_\lambda - \sum_{j=1}^N v_\lambda(x_j) l_j(\cdot)$, hence both expressions are equal and the proof is finished. □

Now the first main theorem concerning the separation distance can be formulated. The proof follows the lines of the ones of Lemma 4.2 and Theorem 4.3 of [4], although we include a refined estimate of the derivative of the power function. Moreover, while the original theorem was used to obtain a decay of the power function, we use exactly this decay here to derive a bound on the separation distance.

**Theorem 19.** *Let $\Omega \subset \mathbb{R}^d$ be a compact domain satisfying an interior cone condition and having a Lipschitz boundary. Suppose that $k$ is a kernel of finite smoothness $\tau$ such that $\tau > d/2 + 1$.*

*Then any $\gamma$-stabilized algorithm yields a sequence of set of points such that*

$$q_{N+1} \geq \tilde{C} \cdot \gamma^2 \cdot h_N \quad \forall N \in \mathbb{N}$$

*and moreover*

$$q_{N+1} \geq C \cdot \gamma^2 \cdot N^{-1/d} \quad \forall N \in \mathbb{N},$$

*for suitable constants $\tilde{C}, C > 0$ independent of $N$.*

**Proof.** For structural reasons the proof is split up into two parts:

i. The representation of the power function in terms of the Lagrange basis (see Proposition 1) yields a possibility to evaluate the partial derivative in the $i$th direction, $i \in \{1, .., d\}$. Here $\partial_i^1$ denotes the partial derivative in the $i$th direction with respect to the first argument:

$$\begin{aligned}\partial_i P_N^2(x) =& \partial_i \left( k(x,\cdot) - \sum_{j=1}^N k(x, x_j) l_j, k(x,\cdot) - \sum_{j=1}^N l_j k(x, x_j) \right)_{\mathcal{H}_k(\Omega)} \\ =& 2 \cdot \left( \partial_i^1 k(x,\cdot) - \sum_{j=1}^N \partial_i^1 k(x, x_j) l_j, k(x,\cdot) - \sum_{j=1}^N l_j k(x, x_j) \right)_{\mathcal{H}_k(\Omega)}\end{aligned}$$

$$\Rightarrow\ |\partial_i P_N^2(x)| \leq 2 \cdot \underbrace{\left\| \partial_i^1 k(x,\cdot) - \sum_{j=1}^{N} \partial_i^1 k(x,x_j) l_j \right\|_{\mathcal{H}_k(\Omega)}}_{=: P_N(\delta_x \circ \partial_i)} \cdot P_N(x)$$
$$= 2 \cdot P_N(\delta_x \circ \partial_i) \cdot P_N(x). \tag{18}$$

Lemma 18 can be used together with Eq. (16) to estimate $P_N(\delta_x \circ \partial_i)$:

$$\begin{aligned} P_N(\delta_x \circ \partial_i) &= \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{|\partial_i f(x) - \partial_i \Pi_{V(X_N)} f(x)|}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &\leq \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{\sup_{|\alpha|=1} \|D^\alpha (f - \Pi_{V(X_N)} f)\|_\infty}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &\overset{(16)}{\leq} \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{C \cdot h_N^{\tau - d/2 - 1} \cdot |f - \Pi_{V(X_N)} f|_{W_2^\tau(\Omega)}}{\|f\|_{\mathcal{H}_k(\Omega)}}. \end{aligned}$$

The application of Eq. (16) was possible since due to Theorem 15 the fill distance $h_N$ will be sufficiently small for $N$ sufficiently large. Now the semi-norm $|\cdot|_{W_2^\tau}$ can be estimated from above by the full norm $\|\cdot\|_{W_2^\tau}$ and then the norm equivalence $\|\cdot\|_{W_2^\tau(\Omega)} \asymp \|\cdot\|_{\mathcal{H}_k(\Omega)}$ can be used as well as $\|\mathrm{Id} - \Pi_{V(X_N)}\|_{\mathcal{L}(\mathcal{H}_k(\Omega))} \leq 1$ due to orthogonality of the projection:

$$\begin{aligned} P_N(\delta_x \circ \partial_i) &\leq C \cdot h_N^{\tau - d/2 - 1} \cdot \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{\|f - \Pi_{V(X_N)} f\|_{W_2^\tau(\Omega)}}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &\leq C' \cdot h_N^{\tau - d/2 - 1} \cdot \sup_{0 \neq f \in \mathcal{H}_k(\Omega)} \frac{\|f - \Pi_{V(X_N)} f\|_{\mathcal{H}_k(\Omega)}}{\|f\|_{\mathcal{H}_k(\Omega)}} \\ &\leq C' \cdot h_N^{\tau - d/2 - 1}. \end{aligned}$$

This estimate for $P_N(\delta_x \circ \partial_i)$ can be plugged into Inequality (18) to obtain

$$|\partial_i P_N^2(x)| \leq 2C' \cdot h_N^{\tau - d/2 - 1} \cdot P_N(x). \tag{19}$$

ii. Now the mean value theorem can be used to estimate $\mathrm{dist}(\Omega_\gamma^{(N)}, X_N)$. Thus, we consider $\tilde{x} \in \Omega_\gamma^{(N)}$ and apply the mean value theorem to $P_N^2$ on the line segment between $\tilde{x}$ and $x_j \in X_N$. This gives a point $\eta = (1-t)\tilde{x} + t x_j$ with $t \in [0,1]$. Since it holds $P_N^2(x_j) = 0$, we have

$$\begin{aligned} \gamma^2 \cdot \|P_N\|_\infty^2 &\leq P_N^2(\tilde{x}) = (\nabla P_N^2)(\eta) \cdot (\tilde{x} - x_j) \\ &\leq \|(\nabla P_N^2)(\eta)\|_2 \cdot \|\tilde{x} - x_j\|_2. \end{aligned} \tag{20}$$

The first factor on the right hand side can be estimated further with help of Eq. (19) from above:

$$\begin{aligned} \|(\nabla P_N^2)(\eta)\|_2 &= \left( \sum_{i=1}^{d} ((\partial_i P_N^2)(\eta))^2 \right)^{1/2} \\ &\overset{(19)}{\leq} \left( \sum_{i=1}^{d} (2C' \cdot h_N^{\tau - d/2 - 1} \cdot P_N(\eta))^2 \right)^{1/2} \\ &\leq 2C' \cdot h_N^{\tau - d/2 - 1} \cdot d \cdot \|P_N\|_{L^\infty}. \end{aligned}$$

Plugging that estimate for $\|(\nabla P_N^2)(\eta)\|_2$ into the Estimate (20) yields

$$\begin{aligned} \gamma^2 \cdot \|P_N\|_\infty^2 \leq & 2C' \cdot h_N^{\tau-d/2-1} \cdot d \cdot \|P_N\|_\infty \cdot \|\tilde{x} - x_j\|_2 \\ \Leftrightarrow \quad \|\tilde{x} - x_j\|_2 \geq & \frac{\gamma^2}{2C' \cdot d} \cdot \frac{\|P_N\|_\infty}{h_N^{\tau-d/2-1}}. \end{aligned}$$

Now the first lower bound of Theorem 12 gives

$$\|\tilde{x} - x_j\|_2 \geq \frac{\gamma^2 \tilde{c}_P}{2C' \cdot d} \cdot \frac{h_N^{\tau-d/2}}{h_N^{\tau-d/2-1}} = \gamma^2 \cdot \tilde{C} \cdot h_N, \tag{21}$$

where we set $\tilde{C} := \tilde{c}_P / \left(2dC'\right)$. Further applying the lower bound (7) gives

$$\|\tilde{x} - x_j\|_2 \geq \gamma^2 \cdot \tilde{C} \cdot c_\Omega \cdot N^{-1/d} = \gamma^2 \cdot C \cdot N^{-1/d}, \tag{22}$$

with $C := \tilde{C} \cdot c_\Omega$.

Since $\tilde{x} \in \Omega_\gamma^{(N)}$ was arbitrary, using (21) and (22) it holds $\text{dist}(\Omega_\gamma^{(N)}, X_N) \geq \tilde{C} \cdot \gamma^2 \cdot h_N$ and $\text{dist}(\Omega_\gamma^{(N)}, X_N) \geq \tilde{C} \cdot \gamma^2 \cdot N^{-1/d}$. Lemma 16 can finally be applied and it directly yields $q_{N+1} \geq \tilde{C} \cdot \gamma^2 \cdot h_N$ and $q_{N+1} \geq \tilde{C} \cdot \gamma^2 \cdot N^{-1/d}$.

This argumentation holds for $N$ sufficiently large. For small $N$ we can simply adjust the constants $\tilde{C}, C$, and this finishes the proof. □

Altogether Theorem 19 states a lower bound on the separation distance which decays as $N^{-1/d}$. Remind that Theorem 15 gives an upper bound on the fill distance which decays also as $N^{-1/d}$. Thus this shows that these decay rates are exact and cannot be improved further.

*5.2.1. A weaker result under weaker conditions*

Theorem 19 needed a further restriction on the smoothness of the kernel, namely $\tau > d/2 + 1$. To circumvent this limitation another way to estimate the distance is proposed in the following, which does not need this assumption.

**Theorem 20.** *Consider a continuous radial basis function kernel $k(x, y) = \phi(\|x - y\|)$ with $\phi(0) = 1$ and $\phi(r)$ monotonically decreasing for $0 \leq r \leq r_0$ for a given $r_0 > 0$. Furthermore $\phi$ is assumed to satisfy an estimate like $1 - \phi(r) \leq c \cdot r^b$ with $c > 0, b \in \mathbb{N}$ for $0 \leq r \leq r_0$.*

*If any $\gamma$-stabilized greedy algorithm is applied to a function on a compact set $\Omega \subset \mathbb{R}^d$ which satisfies an interior cone condition and has a Lipschitz boundary, then the following asymptotic estimate holds with $C > 0$:*

$$q_{N+1} \geq C \cdot \gamma^{2/b} \cdot \left(N^{\frac{1}{2}-\frac{\tau}{d}}\right)^{2/b}.$$

**Proof.** The representation of the power function from Eq. (5) will be used. Projecting to a smaller subspace yields a worse approximation, and in particular using $V(x_j) := \text{span}\{k(\cdot, x_j)\} \subset \text{span}\{k(\cdot, x_i) | x_i \in X_N\} \equiv V(X_N)$ gives the following for all $j \in \{1, .., N\}$:

$$\begin{aligned} P_N(x)^2 \equiv & \|k(\cdot, x) - \Pi_{V_N}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)}^2 \\ \leq & \min_{i=1,\dots,N} \|k(\cdot, x) - \Pi_{V(x_i)}(k(\cdot, x))\|_{\mathcal{H}_k(\Omega)}^2 \\ = & \min_{i=1,\dots,N} \|k(\cdot, x) - k(x_i, x) \cdot k(\cdot, x_i)\|_{\mathcal{H}_k(\Omega)}^2 \end{aligned}$$

where $\|k(\cdot, x_i)\|_{\mathcal{H}_k(\Omega)} = k(x_i, x_i)^{1/2} = \phi(0)^{1/2} = 1$ was used in the last step. The norm can be expressed via the dot product, then the reproducing property of the kernel can be used. Thus it holds

$$\begin{aligned} P_N(x)^2 &\leq \min_{i=1,\dots,N} \|k(\cdot, x) - k(x_i, x) \cdot k(\cdot, x_i)\|^2_{\mathcal{H}_k(\Omega)} \\ &= \min_{i=1,\dots,N} k(x,x) - 2 \cdot k(x_i, x) \cdot k(x, x_i) + k(x_i, x)^2 \cdot \underbrace{k(x_i, x_i)}_{=\phi(0)=1} \\ &= \min_{i=1,\dots,N} 1 - k(x_i, x)^2 \\ &= \min_{i=1,\dots,N} 1 - \phi(\|x - x_i\|)^2. \end{aligned}$$

Theorem 15 gives an estimate on the fill distance as $h_N \leq C\gamma^{-2/(\tau - d/2)} \cdot N^{-1/d}$. Thus for $N$ large enough it holds $h_N < r_0$, i.e., $\min_{i=1,\dots,N} \|x - x_i\| < h_N < r_0 \ \forall x \in \Omega$, hence the monotonicity of $\phi$ can be used in the following.

Theorem 12 provides a lower bound on the decay of the power function, i.e., $\|P_N\|_{L^\infty} \geq c_P \cdot N^{1/2 - \tau/d}$. Any $\gamma$-stabilized greedy algorithm restricts the point selection to $\Omega_\gamma^{(N)} \equiv \{x \in \Omega : P_N(x) \geq \gamma \cdot \|P_N\|_\infty\}$, therefore the distance from such a point $x \in \Omega_\gamma^{(N)}$ from the old points $x_i \in X_N$ can be estimated:

$$\begin{aligned} c_P^2 \cdot \left(N^{\frac{1}{2} - \frac{\tau}{d}}\right)^2 &\leq \|P_N\|_\infty^2 \leq \gamma^{-2} \cdot P_N(x)^2 \\ &\leq \gamma^{-2} \cdot \min_{i=1,\dots,N} 1 - \phi(\|x - x_i\|)^2 \\ &\leq \gamma^{-2} \cdot \min_{i=1,\dots,N} 2c\|x - x_i\|^b \\ \Rightarrow \quad \left(\frac{c_P^2 \cdot \gamma^2}{2c}\right)^{1/b} \cdot \left(N^{\frac{1}{2} - \frac{\tau}{d}}\right)^{2/b} &\leq \min_{i=1,\dots,N} \|x - x_i\| \end{aligned} \tag{23}$$

In the last step $1 - \phi(\|x - x_i\|)^2 = (1 + \phi(\|x - x_i\|)) \cdot (1 - \phi(\|x - x_i\|))$ was estimated from above with help of the assumption $1 - \phi(r) \leq c \cdot r^b$ and the estimate $\phi(\|x - x_i\|) \leq \phi(0) = 1$. Since $x \in \Omega_\gamma^{(N)}$ was arbitrary, Inequality (23) is valid for all $x \in \Omega_\gamma^{(N)}$. Thus it holds

$$\begin{aligned} \operatorname{dist}(\Omega_\gamma^{(N)}, X_N) &= \inf_{x \in \Omega_\gamma^{(N)}} \min_{i=1,\dots,N} \|x - x_i\| \\ &\geq \left(\frac{c_P^2 \cdot \gamma^2}{2c}\right)^{1/b} \cdot \left(N^{\frac{1}{2} - \frac{\tau}{d}}\right)^{2/b}. \end{aligned}$$

Now define $C = c_P^{2/b} \cdot (2c)^{-1/b}$ and apply Lemma 16 which finishes the proof. □

**Remark 21.** Theorem 20 allows one to derive statements about the separation distance for a larger class of kernels, since no further restriction $\tau > d/2 + 1$ is needed. Although the derived result is weaker, it is still optimal in certain cases, even not covered by the first theorem. For example, if one applies the theorem to the basic Matérn kernel whose radial basis function is given by $\phi(r) = e^{-r}$, the estimate yields $q_{N+1} \geq C \cdot \gamma^{-2} \cdot N^{-1/d}$ which shows uniformity also for this kernel.

## 6. Stability and refined error estimates

As anticipated, we can now use the results derived in the previous sections to draw some additional conclusions on the stability and convergence of the interpolation by any $\gamma$-stabilized algorithm.

Despite all the following results are just the combination of known results with properties of the selected points, it is worth mentioning them here to highlight the features of the new

algorithm. These features are indeed either refined versions of the ones proven in [25] for the $P$-greedy selection, or completely new results. In particular, they prove for the first time that a kernel-greedy algorithm can achieve stability, provide error bounds w.r.t. to general $L^p$ norm for the approximation of derivatives, and approximate functions which are outside of the native space.

**Corollary 22.** *Assume that $k$ is a translational invariant kernel with finite smoothness $\tau$, and let $\Omega \subset \mathbb{R}^d$ be a bounded domain which satisfies an interior cone condition. Let $(X_N) \subset \Omega$ be the sequence of sets of points selected by any $\gamma$-stabilized algorithm, and let $c, c' > 0$ denote generic constants independent of $N$ and $\gamma$.*

*i. Bounds on $\lambda_{\min}$: Assume additional smoothness $\tau > d/2 + 1$. Then the minimal eigenvalue $\lambda_{\min}(X_N)$ of the kernel matrix of $k$ on $X_N$ can be bounded as*

$$c \cdot \gamma^{4\tau-2d} \cdot N^{1-2\tau/d} \leq \lambda_{\min}(X_N) \leq c' \cdot \gamma^{-4} \cdot N^{1-2\tau/d}$$

*ii. Bound on the Lebesgue constant: Assume additional smoothness $\tau > d/2+1$. Then there exist $N_0 \in \mathbb{N}$ and a constant $c > 0$ such that for all $N \geq N_0$ the Lebesgue constant is bounded as*

$$\Lambda_{X_N} \leq c\gamma^{-2(\tau-d/2)}\sqrt{N}.$$

*iii. Bounds for derivatives: Under the conditions of Theorem 3, there exist $N_0 \in \mathbb{N}$ and a constant $c > 0$ such that for all $f \in W_p^\tau(\Omega)$ and all $N \geq N_0$ it holds*

$$\left|f - \Pi_{V(X_N)} f\right|_{W_q^m(\Omega)} \leq c\gamma^{\frac{-2\left(\tau-m-d(1/p-1/q)_+\right)}{\tau-d/2}} N^{\left(-\frac{\tau-m}{d}+\left(\frac{1}{p}-\frac{1}{q}\right)_+\right)} \|f\|_{W_p^\tau(\Omega)},$$

*where the bound holds in particular for all $f \in \mathcal{H}_k(\Omega)$, by using a different constant, $p = 2$, and the norm $\|f\|_{\mathcal{H}_k(\Omega)}$ in the right hand side.*

*iv. Asymptotic point distribution: Assume additional smoothness $\tau > d/2 + 1$. Then the sequence of points is asymptotically uniformly distributed, and in particular there exists a constant $c > 0$ such that*

$$\rho_{X_N} := \frac{h_N}{q_N} \leq c\ \gamma^{-2} \text{ for all } N \in \mathbb{N}. \tag{24}$$

*v. Escaping the native space: For all $d/2 < \beta \leq \tau$ and for all $0 \leq \mu \leq \beta$ there exist $N_0 \in \mathbb{N}$ and a constant $c > 0$ such that for all $f \in W_2^\beta(\Omega)$ and for all $N \geq N_0$ it holds*

$$\left|f - \Pi_{V(X_N)} f\right|_{W_2^\mu(\Omega)} = c\gamma^{-\frac{2(\beta-\mu)}{\tau-d/2}-2(\tau-\mu)} N^{-\frac{\beta-\mu}{d}} \|f\|_{W_2^\beta(\Omega)}.$$

**Proof.** We prove the five points separately:

i. In the case of kernel of finite smoothness it is known that the smallest eigenvalue can be bounded from below with help of the separation distance $q_N$ by $\lambda_{\min}(X_N) \geq \tilde{c} \cdot q_N^{2\tau-d}$, see e.g. [30, Chap. 12.2]. The lower estimate on the separation distance from Theorem 19 can be applied to conclude the stated lower estimate.
For the upper estimate we use $\lambda_{\min}(X_N) \leq |P_{N-1}(x_N)|^2$ [30, Chap. 12.1] and estimate the right hand side further by applying the upper bound on the maximal value of the power function from Theorem 8 which yields the result.

ii. This is simply the application of point (iv.) to Theorem 1 of [3].

iii. Here we just substitute the decay of the fill distance of Theorem 15 into the estimate of Theorem 3. The rate of decay on the right hand side is thus given by

$$h_N^{\tau-m-d(1/p-1/q)_+} \leq \left(c\cdot\gamma^{-2/(\tau-d/2)}\cdot N^{-1/d}\right)^{\tau-m-d(1/p-1/q)_+} \leq c\gamma^{\frac{-2\left(\tau-m-d(1/p-1/q)_+\right)}{\tau-d/2}} N^{-\frac{\tau-m}{d}+\left(\frac{1}{p}-\frac{1}{q}\right)_+}.$$

To obtain the bound for $f \in \mathcal{H}_k(\Omega)$, it is sufficient to recall that $\mathcal{H}_k(\Omega)$ is norm equivalent to $W_2^\tau(\Omega)$.

iv. This point is just a direct computation using the first estimate of Theorem 19.

v. Theorem 4.2 from [23] gives that for all $f \in W_2^\beta(\Omega)$ with $d/2 < \beta \leq \tau$, and for all $0 \leq \mu \leq \beta$, there exists $h_0$ such that for all $X \subset \Omega$ with $h_X \leq h_0$ it holds

$$\left|f - \Pi_{V(X)}f\right|_{W_2^\mu(\Omega)} \leq Ch_X^{\beta-\mu}\rho_X^{\tau-\mu}\|f\|_{W_2^\beta(\Omega)}.$$

We can then use the bounds of Theorem 15 and the bound on the uniformity constant proven in the previous point to obtain that

$$\begin{aligned}\left|f - \Pi_{V(X_N)}f\right|_{W_2^\mu(\Omega)} &\leq C'\gamma^{-\frac{2(\beta-\mu)}{\tau-d/2}}\gamma^{-2(\tau-\mu)}N^{-\frac{\beta-\mu}{d}}\|f\|_{W_2^\beta(\Omega)} \\ &= C'\gamma^{-\frac{2(\beta-\mu)}{\tau-d/2}-2(\tau-\mu)}N^{-\frac{\beta-\mu}{d}}\|f\|_{W_2^\beta(\Omega)}. \quad \square\end{aligned}$$

Observe that point (i.) and point (ii.) imply that using a $\gamma$-stabilized greedy algorithm yields a provable stable interpolation process, since the condition number of the interpolation matrix and the $L^\infty(\Omega)$-norm of the interpolant cannot grow arbitrary fast. Thus it is justified to name any $\gamma$-restricted algorithm also $\gamma$-*stabilized* algorithm.

Moreover, since it is now clear what stability statements we are able to prove, we are in a position to put them into perspective within the wider scope of the stability research in kernel methods. Indeed, it should be noted that there are two different notions of stability into play here.

Point ((i.)) in Corollary 22 is related to the conditioning of the kernel matrix. This means that the statement is interesting as far as the interpolant is computed by the so-called *direct method* that we are using in this paper, i.e., by inverting the kernel matrix, or equivalently by expressing the interpolant in terms of the standard basis $\{K(\cdot, x_i) : x_i \in X_N\}$. Other techniques have been extensively studied in the recent years to overcome this kind of instability by making use of different and better conditioned bases, especially for infinitely smooth RBF kernels in the flat limit, i.e., $\varepsilon \to 0$, for example RBF-QR [9,11,17], Hilbert–Schmidt SVD [7], RBF-CP [12], and RBF-GA [10].

On the other hand, point (ii.) in Corollary 22 regards what is called *stability in function space* (see [3]). Indeed, the Lebesgue constant is the norm of the interpolation operator $(\mathbb{R}^n, \|\cdot\|_\infty) \to \left(\mathcal{H}_k(\Omega), \|\cdot\|_{L^\infty(\Omega)}\right)$. This in particular means that the bound $\|\Pi_{V(X_N)}f\|_{L^\infty(\Omega)} \leq \Lambda_{X_N}\|f|_{X_N}\|_\infty$ is sharp, and that it is valid independently of the computational procedure used to compute $\Pi_{V(X_N)}f$.

Moreover, stability can be improved also by moving from interpolation to regularized algorithms, such as approximate interpolation (see e.g. [32]) and least-squares regression (see e.g. [18]).

**Remark 23.** We remark that precise error bounds w.r.t. an $L^q$ norm and for the approximation of derivatives are difficult to obtain having only bounds on the $L^\infty$ norm of the power function

as the ones of Theorem 8. For example, if $k$ is smooth enough then any $f \in \mathcal{H}$ is continuous, and thus if $\Omega$ is bounded it holds

$$\begin{aligned}\left\|f - \Pi_{V(X_N)}(f)\right\|_{L^q(\Omega)} &\leq \left\|f - \Pi_{V(X_N)}(f)\right\|_{L^\infty(\Omega)} \mathrm{meas}(\Omega)^{1/q} \\ &\leq \mathrm{meas}(\Omega)^{1/q} \|P_N\|_{L^\infty(\Omega)} \left\|f - \Pi_{V(X_N)}(f)\right\|_{\mathcal{H}_k(\Omega)} \\ &\leq c\mathrm{meas}(\Omega)^{1/q} \gamma^{-2} \cdot N^{\frac{1}{2}-\frac{\tau}{d}} \left\|f - \Pi_{V(X_N)}(f)\right\|_{\mathcal{H}_k(\Omega)}.\end{aligned}$$

This technique, which was used in [25], gives an extra factor of $N^{1/2-(1/2-1/q)_+}$ which is in general suboptimal compared to the one of the point (iii.) of the corollary.

This improvement is relevant also for other scenarios that have been discussed in the literature. For example, for $q = 2$ and $m = 0$ the estimate within point (iii.) of the corollary gives a rate of convergence of order $N^{-\frac{\tau}{d}}$, and not just $N^{\frac{1}{2}-\frac{\tau}{d}}$. This improved rate of convergence was observed numerically in [27] in the setting of superconvergence with uniformly distributed points, though not proven (see the discussion at p. 21–22).

## 7. Numerical experiments

### 7.1. Decay rates of power function

To complement the analytically derived upper and lower bound on the decay of the value of the power function for any $\gamma$-stabilized algorithm, some numerical experiments are performed.

To stay as general as possible, no concrete selection criterion was chosen, and instead arbitrary points within the restricted area $\Omega_\gamma^{(N)}$ were chosen. In practice this is done by selecting a random point within the restricted set and, to minimize the effect of this non-deterministic selection, every run was repeated 10 times. This approach reduces the uncertainty of the results, but we remark that the computed values are still slightly dependent on the particular machine and random seed that we used.

The experiments were performed on $\Omega = [0, 1]^d$ for $d \in \{1, 3, 5\}$ using $\gamma \in \{0.1, 0.2, .., 0.9, 1\}$. The set $\Omega$ was discretized with $3 \cdot 10^4$ uniformly randomly sampled points.

The basic Matérn and the linear Matérn kernel were used, whose radial basis functions are given by

$$\begin{aligned}\phi_{\mathrm{bas}}(r) &= e^{-r} \\ \phi_{\mathrm{lin}}(r) &= (1 + r) \cdot e^{-r}.\end{aligned}$$

Up to $N = 800$ sampling points were chosen in each run, and they correspond to up to 800 values of the maximal value of the power function which are denoted as $\{\|P_N\|_{L^\infty(\Omega)}\}_{N\in[1:800]}$.

To estimate the numerical rate of convergence, we first computed a fit to the values $\{\log(\|P_N\|_{L^\infty(\Omega)})\}_{N\in I}$ for $I \in \{[a, b] \mid a \in \{50, 75, 100\}, b \in \{600, 700, 800\}\}$ with functions of the form $y_2(n) = \log(\alpha) + \lambda \cdot \log(n)$. The parameter $\lambda$ corresponds to the decay rate within Eq. (9) and (14) and is expected to be close to $1/2 - \tau/d$. The parameter $\alpha$ corresponds to the factors within these equations and in general it depends e.g. on the domain $\Omega$. In general there is no relation between $\alpha$ and $\lambda$.

Then, the mean value of these nine fit parameters $\alpha$ is regarded as a meaningful estimation of the numerical decay value. Using this procedure, we compensate the effect of the values of the power function for the first iterates (which are affected the most by the non-deterministic point selection) and the ones for the final iterates. This mean value and the standard deviation

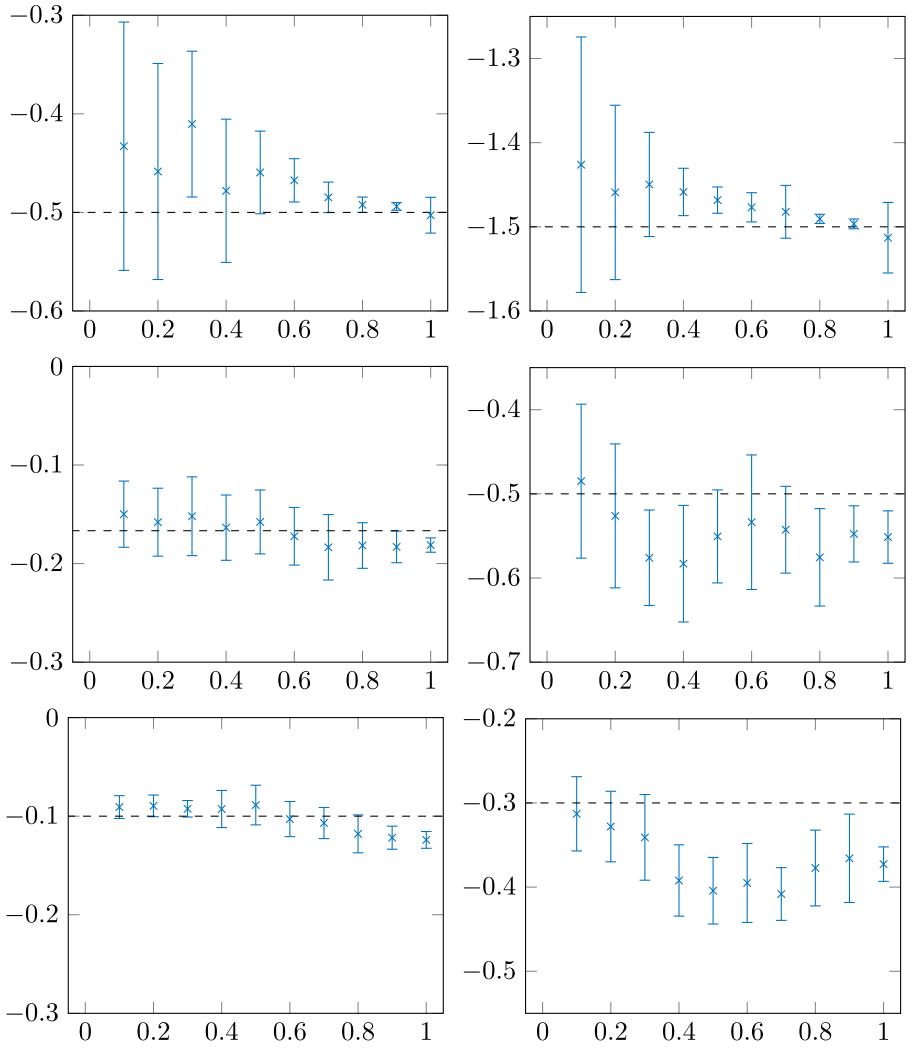


**Fig. 2.** Convergence rates for the decay of the maximal values $\|P_N\|_{L^\infty(\Omega)}$ of the power function according to stabilized greedy algorithm for varying restriction parameters $\gamma$ applied to $\Omega = [0,1]^d$.
The horizontal axis describes the restriction parameter $\gamma$ and the vertical axis describes the convergence rate $\lambda$, see Section 7.1.
From top to bottom the dimensions $d = 1, 3, 5$ were used. On the left side the unscaled basic Matérn kernel was applied, on the right side the unscaled linear Matérn kernel was used. The computed decay rates are scattered around a mean. The dashed line marks the proven decay rate of $1/2 - \tau/d$.

of these nine values is displayed in Fig. 2 for the different experimental settings. Additionally, the dotted line indicates the expected theoretical value of $1/2 - \tau/d$.

We can confirm that the computed numerical values are in accordance with the analytical result. Taking into account that the numerically computed values are fit parameters, whereas the analytically derived quantity $1/2 - \tau/d$ is the decay rate of an upper and lower bound, the results are matching.

**Table 1**
Minimal error as well as number of chosen points depending on the restriction parameter $\gamma$ for the function $f_\alpha$ with $\alpha \in \{1.51, 3.5\}$.

| $f/P$ | $\alpha = 1.51$ | | $\alpha = 3.5$ | |
|---|---|---|---|---|
| $\gamma =$ | $N_{\max}$ | $\|r_{N_{\max}}\|_\infty$ | $N_{\max}$ | $\|r_{N_{\max}}\|_\infty$ |
| 0 | 13 | $1.61 \cdot 10^{-3}$ | 29 | $4.88 \cdot 10^{31}$ |
| $10^{-4}$ | 43 | $1.80 \cdot 10^{-5}$ | 17 | $3.54 \cdot 10^{-6}$ |
| $10^{-3}$ | 74 | $9.63 \cdot 10^{-7}$ | 53 | $4.85 \cdot 10^{-7}$ |
| $10^{-2}$ | 198 | $2.31 \cdot 10^{-6}$ | 242 | $5.35 \cdot 10^{-8}$ |
| $10^{-1}$ | 815 | $2.23 \cdot 10^{-6}$ | 766 | $3.91 \cdot 10^{-8}$ |
| $10^{0}$ | 1380 | $1.47 \cdot 10^{-6}$ | 1380 | $3.53 \cdot 10^{-8}$ |

### 7.2. *Improved accuracy of* f/P*-greedy*

In the following, an example is provided which shows that the $\gamma$-stabilized $f/P$-greedy algorithm is able to yield a better interpolant than the plain $f/P$-greedy algorithm in terms of expansion size and accuracy.

For this, functions of the form

$$f_\alpha : [-0.5, 0.5] \to \mathbb{R}, x \mapsto |x|^\alpha \cdot e^{-x^2}$$

are considered which depend on a parameter $\alpha > 0$. A calculation of the Fourier transform of $f_\alpha : \mathbb{R} \to \mathbb{R}$ shows that $f_\alpha \in W_2^{\alpha+1/2-\epsilon}(\mathbb{R})$ $\forall \epsilon > 0$ and thus we also have $f_\alpha \in W_2^{\alpha+1/2-\epsilon}(\mathbb{R})$ $\forall \epsilon > 0$.

The function $f_\alpha$ is investigated with the $\gamma$-stabilized $f/P$-greedy algorithm and the unstabilized $f/P$-greedy algorithm using the linear Matérn kernel. The native space of the linear Matérn kernel on $[-0.5, 0.5]$ is norm equivalent to $W_2^2([-0.5, 0.5])$. Thus $\alpha = 1.51$ is chosen, since it yields a function with smoothness almost 2.01, which is close to the smoothness of the native space. Furthermore, $\alpha = 3.5$ is chosen, which yields a function with approximately double the smoothness of the native space.

The training and the test set consists of each $10^5$ uniformly sampled distinct points within $\Omega$. The algorithm was run until the condition number of the kernel matrix exceeded a bound of $10^{14}$. The application of the $\gamma$-stabilized $f/P$-greedy algorithms for $\gamma \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$ as well as the unstabilized $f/P$-greedy yielded different decays of the residuals. Fig. 3 displays the results and Table 1 collects some numbers, namely the number $N_{\max}$ of chosen sampling points and the remaining error, i.e., the maximal error of the final residual $\|r_{N_{\max}}\|_\infty$.

From the values in Table 1 and Fig. 3 we can draw the following conclusions:

i. In the case of $\alpha = 3.5, \gamma = 0$ the interpolant explodes due to numerical inaccuracies. Apart from this the unstabilized $f/P$-greedy algorithm (i.e., $\gamma = 0$) and the barely stabilized ones yield the fastest residual decay rates such that their lines partly hide each other in the first figure. The reason is presumably that the restriction parameter is that small, that the limitation due to $\Omega_\gamma^{(N)}$ is not an actual restriction and thus almost the same points are chosen all the time.
With increasing value of $\gamma$ the decay behaviors become worse, but they are always between the decay behaviors for the unstabilized greedy ($\gamma = 0$) and the $P$-greedy ($\gamma = 1$).

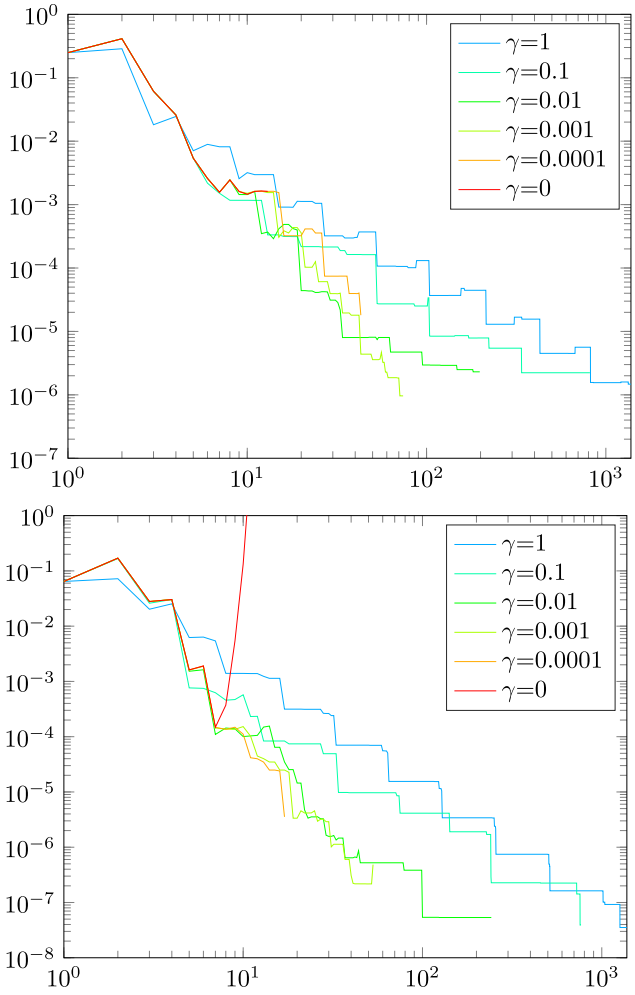


**Fig. 3.** Residual decay for functions given by $f_\alpha(x) = |x|^\alpha \cdot \exp(-x^2)$ for the linear Matérn kernel for $\alpha = 1.51$ (top) and $\alpha = 3.5$ (bottom) using the $\gamma$-stabilized $f/P$-greedy. The horizontal axis describes the number of chosen sampling points, the vertical axis describes the maximal residual value. For small values of $\gamma$ or $\gamma = 0$ the algorithm stopped early, since the condition number exceeded a predefined bound of $10^{14}$.

ii. Increasing the value of the restriction parameter $\gamma$ yields more sampling points. This can be observed best within the table.
For intermediate values of $\gamma$ similar accuracies compared to $\gamma = 1$ can be achieved but with way less sampling points. This speeds up the evaluation of the interpolant.

Thus there is a tradeoff between speed of the residual decay and the amount of chosen sampling points. In fact for $\alpha = 3.5$ the minimal error is attained using an intermediate restriction parameter, namely $\gamma = 10^{-3}$. The reason is that the unstabilized $f/P$-greedy algorithm and the one with a very small restriction parameter $\gamma$ stop too early due to the criterion on the

condition number. With increasing restriction parameter $\gamma$ more and more points are chosen which yield a better approximation.

As a result it can be concluded that the $\gamma$-stabilized algorithms yield also in practice a more stabilized algorithm, such that more points can be selected. Especially for the $\gamma$-stabilized $f/P$-greedy the restriction yields a better approximation, since the algorithm does not stop that early due to the exceed of the bound on the condition number. The additional sampling points help to reach a better approximation. Since the $f/P$ selection criterion incorporates the power function in the denominator, the sampling points are sometimes chosen close nearby already selected points.

Moreover, when these stability conditions are satisfied, it is evident that the use of a function-dependent selection rule improves the convergence speed with respect to a pure $P$-greedy selection. This effect, as mentioned before, is not appearing in our estimates, but it is an expected behavior that we plan to investigate further.

**Remark 24.** We remark that for any practical use of the algorithm, this tradeoff may be faced by a systematic selection of the best parameter. Namely, as with different parameters that may be involved in the kernel definition, it is possible to run for example a $K$-fold cross validation to select the parameter $\gamma$ that provides the best prediction given the knowledge of the current data. A description of this method for kernel approximation can be found e.g. in [26].

### 7.3. Distribution of sampling points

Finally we provide examples to visualize the distribution of the selected sampling points on some arbitrarily chosen domain for several values of $\gamma$. For this we consider

$$g(\varphi) = 0.35 \cdot (\cos\left(\pi \left(\frac{\varphi}{\pi}\right)^2\right) + 2) \cdot (0.15 \cdot \cos(\varphi)^2 + 0.3).$$

Let $\theta(x) \in [0, 2\pi)$ denote the standard angular coordinate of $x \in \mathbb{R}^2$ in the polar coordinate system and define $a := (0.17, 0.17)^T \in \mathbb{R}^2$. The domain $\Omega \subset \mathbb{R}^2$ is given by $\Omega = \Omega_1 \setminus \Omega_2$ with

$$\Omega_1 := \{x + (0.1, 0)^T \in \mathbb{R}^2 \mid \|x\|_2 < g(\theta(x) + \pi)\}$$
$$\Omega_2 := \{x \in \mathbb{R}^2 \mid \|x - a\|_2^2 \leq 0.003\}.$$

For the numerical implementation of the greedy selection algorithm, the domain $\Omega$ was discretized with 831 uniformly randomly distributed points. Now the linear Matérn kernel and the $\gamma$-stabilized $f/P$-greedy algorithm with $\gamma \in \{0, 0.04, 0.15, 1\}$ was applied to the function

$$f(x) = \frac{1}{\|x - a\|^2}$$

and 50 sampling points were selected for each $\gamma$-value. The different point distributions are displayed in Fig. 4. One can observe that for $\gamma = 0$ all the points are clustered close to the point $a$, which is the midpoint of the "cut out" hole $\Omega_2$. For $\gamma = 1$ the points are uniformly distributed within $\Omega$. The cases $\gamma = 0.04$ and $\gamma = 0.15$ provide intermediate distributions at the preasymptotic range, i.e. prior to the asymptotic uniform distribution proven in Section 5.

## 8. Conclusion and outlook

In this paper the class of $\gamma$-stabilized greedy kernel algorithms was introduced and investigated analytically, especially for kernels of finite smoothness $\tau > d/2$. Precise estimates

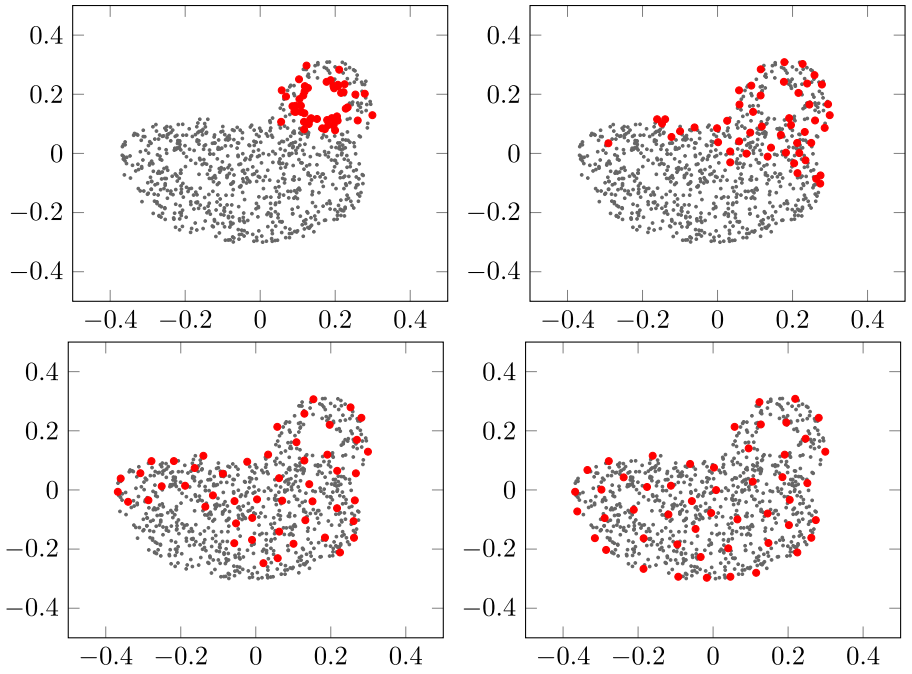


**Fig. 4.** Selected sampling points (red, large) of the $\gamma$-stabilized $f/P$-greedy algorithm using the linear Matérn kernel. From left to right, top to bottom $\gamma \in \{0, 0.04, 0.15, 1\}$ was used. For small $\gamma$ values the point distribution is better adapted to the given data function $f$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for the decay of the power function were derived, and these lead to provable convergence rates. Furthermore, the resulting point distribution was quantified and their asymptotical uniform distribution was proven. This strong result leads to improved convergence rates and stability statements. The results were illustrated with numerical examples showing improved accuracy due to a more stable distribution of the sampling points.

However, some questions remain open. Notably, though the $\gamma$-stabilized algorithm can be specialized to single functions, we derived in this paper only general convergence statements that hold for the entire set of functions of the native space. Refined statements on the adaptive behavior which could be observed in the numerical examples for specific functions will be the focus of further investigations.

Moreover, although in this paper we concentrated on input spaces $\Omega$ which satisfy an interior cone condition in $\mathbb{R}^d$, most of the theoretical results on Sobolev spaces and native spaces that were used in this paper have analogous versions that work on manifolds (see e.g. [13]). It is thus foreseeable that the current results can be extended to this setting and, one would in particular obtain a method to place uniform points on quite arbitrary geometries by requiring only a cloud of initial points and a kernel on this manifold.

## Acknowledgments

The authors acknowledge the funding of the project by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 - 390740016.

We also thank Dominik Wittwar for discussions leading to the idea of the stabilized algorithms, and the anonymous reviewers for providing constructive comments that helped improving this paper.

## References

[1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, P. Wojtaszczyk, Convergence rates for greedy algorithms in reduced basis methods, SIAM J. Math. Anal. 43 (3) (2011) 1457–1472.
[2] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme, G. Turinici, A priori convergence of the greedy algorithm for the parametrized reduced basis, Math. Modell. Numer. Anal. 46 (3) (2012) 595–603.
[3] S. De Marchi, R. Schaback, Stability of kernel-based interpolation, Adv. Comput. Math. 32 (2) (2010) 155–161.
[4] S. De Marchi, R. Schaback, H. Wendland, Near-optimal data-independent point locations for radial basis function interpolation, Adv. Comput. Math. 23 (3) (2005) 317–330.
[5] R. DeVore, G. Petrova, P. Wojtaszczyk, Greedy algorithms for reduced bases in Banach spaces, Constr. Approx. 37 (3) (2013) 455–466.
[6] G.E. Fasshauer, Meshfree Approximation Methods with MATLAB, in: Interdisciplinary Mathematical Sciences, vol. 6, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.
[7] G.E. Fasshauer, M.J. McCourt, Stable evaluation of Gaussian radial basis function interpolants, SIAM J. Sci. Comput. 34 (2) (2012) A737–A762.
[8] G.E. Fasshauer, M. McCourt, Kernel-Based Approximation Methods Using MATLAB, in: Interdisciplinary Mathematical Sciences, vol. 19, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015.
[9] B. Fornberg, E. Larsson, N. Flyer, Stable computations with Gaussian radial basis functions, SIAM J. Sci. Comput. 33 (2) (2011) 869–892.
[10] B. Fornberg, E. Lehto, C. Powell, Stable calculation of Gaussian-based RBF-FD stencils, Comput. Math. Appl. 65 (4) (2013) 627–637.
[11] B. Fornberg, C. Piret, A stable algorithm for flat radial basis functions on a sphere, SIAM J. Sci. Comput. 30 (1) (2007/08) 60–80.
[12] B. Fornberg, G. Wright, Stable computation of multiquadric interpolants for all values of the shape parameter, Comput. Math. Appl. 48 (5–6) (2004) 853–867.
[13] E. Fuselier, G.B. Wright, Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates, SIAM J. Numer. Anal. 50 (3) (2012) 1753–1776.
[14] Q.L. Gia, F. Narcowich, J. Ward, H. Wendland, Continuous and discrete least-squares approximation by radial basis functions on spheres, J. Approx. Theory 143 (1) (2006) 124–133, Special Issue on Foundations of Computational Mathematics.
[15] B. Haasdonk, Convergence rates of the POD–Greedy method, ESAIM: Math. Modell. Numer. Anal. 47 (3) (2013) 859–873.
[16] B. Haasdonk, G. Santin, Greedy Kernel approximation for sparse surrogate modeling, in: W. Keiper, A. Milde, S. Volkwein (Eds.), Reduced-Order Modeling (ROM) for Simulation and Optimization: Powerful Algorithms As Key Enablers for Scientific Computing, Springer International Publishing, Cham, 2018, pp. 21–45.
[17] E. Larsson, E. Lehto, A. Heryudono, B. Fornberg, Stable computation of differentiation matrices and scattered node stencils based on Gaussian radial basis functions, SIAM J. Sci. Comput. 35 (4) (2013) A2096–A2119.
[18] S. Li, L. Ling, K.C. Cheung, Discrete least-squares radial basis functions approximations, Appl. Math. Comput. 355 (2019) 542–552.
[19] Y. Maday, O. Mula, G. Turinici, Convergence analysis of the generalized empirical interpolation method, SIAM J. Numer. Anal. 54 (3) (2016) 1713–1731.
[20] C.A. Micchelli, M. Pontil, On learning vector-valued functions, Neural Comput. 17 (1) (2005) 177–204.
[21] S. Müller, Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden (Complexity and Stability of Kernel-based Reconstructions) (Ph.D. thesis), Fakultät für Mathematik und Informatik, Georg-August-Universität Göttingen, 2009.
[22] F.J. Narcowich, J.D. Ward, H. Wendland, Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting, Math. Comp. 74 (250) (2005) 743–763.
[23] F.J. Narcowich, J.D. Ward, H. Wendland, Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions, Constr. Approx. 24 (2) (2006) 175–186.
[24] M. Pazouki, R. Schaback, Bases for kernel-based spaces, J. Comput. Appl. Math. 236 (4) (2011) 575–588.
[25] G. Santin, B. Haasdonk, Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation, Dolomites Res. Notes Approx. 10 (2017) 68–78.

[26] G. Santin, B. Haasdonk, Kernel Methods for Surrogate Modeling, (1907.10556) 2019, https://arxiv.org/abs/1907.10556.
[27] R. Schaback, Superconvergence of kernel-based interpolation, J. Approx. Theory 235 (2018) 1–19.
[28] R. Schaback, H. Wendland, Adaptive greedy techniques for approximate solution of large RBF systems, Numer. Algorithms 24 (3) (2000) 239–254.
[29] V.N. Temlyakov, Greedy approximation, Acta Numer. 17 (2008) 235–409.
[30] H. Wendland, Scattered Data Approximation, in: Cambridge Monographs on Applied and Computational Mathematics, vol. 17, Cambridge University Press, Cambridge, 2005.
[31] H. Wendland, Multiscale radial basis functions, in: I. Pesenson, Q.T. Le Gia, A. Mayeli, H. Mhaskar, D.-X. Zhou (Eds.), Frames and Other Bases in Abstract and Function Spaces: Novel Methods in Harmonic Analysis, Vol. 1, Springer International Publishing, Cham, 2017, pp. 265–299.
[32] H. Wendland, C. Rieger, Approximate interpolation with applications to selecting smoothing parameters, Numer. Math. 101 (4) (2005) 729–748.
[33] D. Wirtz, B. Haasdonk, A vectorial kernel orthogonal greedy algorithm, Dolomites Res. Notes Approx. 6 (2013) 83–100.
[34] D. Wittwar, G. Santin, B. Haasdonk, Interpolation with uncoupled separable matrix-valued kernels, Dolomites Res. Notes Approx. 11 (2018) 23–29.