Convergence analysis of the Generalized Empirical Interpolation Method

Yvon Maday, Olga Mula, Gabriel Turinici

Convergence analysis of the Generalized Empirical Interpolation Method

Y. Maday[1,2,3], O. Mula[4] and G. Turinici[2,4]

[1] UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France.

[2] Institut Universitaire de France.

[3] Brown Univ, Division of Applied Maths, Providence, RI, USA.

[4] Université Paris-Dauphine, PSL Research University, CNRS, UMR 7534, CEREMADE, 75016 Paris, France

Abstract: Let F be a compact set of a Banach space $\mathcal{X}$. This paper analyses the "Generalized Empirical Interpolation Method" (GEIM) which, given a function $f \in F$, builds an interpolant $\mathcal{J}_n[f]$ in an n-dimensional subspace $X_n \subset \mathcal{X}$ with the knowledge of n outputs $(\sigma_i(f))_{i=1}^n$, where $\sigma_i \in \mathcal{X}'$ and $\mathcal{X}'$ is the dual space of $\mathcal{X}$. The space X_n is built with a greedy algorithm that is *adapted* to F in the sense that it is generated by elements of F itself. The algorithm also selects the linear functionals $(\sigma_i)_{i=1}^n$ from a dictionary $\Sigma \subset \mathcal{X}'$. In this paper, we study the interpolation error $\max_{f\in F} \|f - \mathcal{J}_n[f]\|_{\mathcal{X}}$ by comparing it with the best possible performance on an n-dimensional space, i.e., the Kolmogorov n-width of F in $\mathcal{X}$, $d_n(F, \mathcal{X})$. For polynomial or exponential decay rates of $d_n(F, \mathcal{X})$, we prove that the interpolation error has the same behavior modulo the norm of the interpolation operator. Sharper results are obtained in the case where $\mathcal{X}$ is a Hilbert space.

1 Introduction

Let $\mathcal{X}$ be a Banach space of functions defined over a domain $\overline{\Omega} \subset \mathbb{R}^d$ or $\mathbb{C}^d$, $d \geq 1$, and let F be a compact set of $\mathcal{X}$. Without loss of generality, we assume that the functions $\varphi \in F$ satisfy $\|\varphi\|_{\mathcal{X}} \leq 1$, where $\|\cdot\|_{\mathcal{X}}$ is the norm of $\mathcal{X}$. In this paper, we investigate the approximation quality of functions in F with the "Generalized Empirical Interpolation Method" (GEIM, see [10, 11]). For a given $f \in F$, the method builds an interpolant $\mathcal{J}_n[f]$ in an n-dimensional subspace $X_n \subset \mathcal{X}$ with the knowledge of n outputs of $(\sigma_i(f))_{i=1}^n$ where the σ_i are bounded linear functionals of $\mathcal{X}'$. The approximation space X_n is built with a greedy algorithm that is *adapted* to F in the sense that it is generated by some elements of F itself. The algorithm also selects the linear functionals to be used from a dictionary $\Sigma \subset \mathcal{X}'$. This procedure is a generalization of the Empirical Interpolation Method (see [1, 9, 13]) that was originally defined for $(\mathcal{C}(\bar{\Omega}), \|\cdot\|_\infty)$ and with Dirac masses as linear functionals.

The current setting falls into in the framework of optimal recovery (see, e.g. [17, 16]) in the sense that, for a given $f \in F$, we want to approximate it in an appropriate basis and using the knowledge of certain outputs $\sigma_i(f)$, $1 \leq i \leq n$. However note that there is an important difference with respect to the classical setting of this field which assumes that F is the unit ball of a smoothness space in $\mathcal{X}$. In the current setting, we allow F to have more general types of geometry/regularity. For this reason, our approximation space X_n and the linear functionals are chosen depending on F. This is in contrast to other methods like polynomial/spline/radial interpolation (see, e.g., [5]) or even meshless methods [2] since there the basis functions are fixed in advance. This idea of adaptivity makes GEIM be also different from kriging [19] where an underlying stochastic process is a priori given. Here we only assume that F is compact, which enlarges the range of potential applications. For instance, it is possible to apply the methodology to the case where F is the solution manifold of a parametric PDE. This case is relevant for at least two types of applications. The first concerns the approximation of this manifold with *reduced basis* methods and the treatment of nonlinearities in the PDE (see [9]). The second concerns the coupling of measurement data (represented by the $\sigma_i(f)$, $1 \leq i \leq n$) with parametrized models (represented by the space X_n) in a systematic way and on the basis of suitable functional spaces. For the particular setting of GEIM, numerical examples can be found in [10, 11] where it was also explained how the method could assist in the placement of sensors in real physical experiments and the minimization

of their number. After the submission of this paper, this line of research has been further developed in several relevant works. For Hilbert spaces, an extension has been proposed in [14] which consists in a particular least squares approximation with $m \geq n$ measurement data. In addition, it has been proven in [4] that, when $\mathcal{X}$ is a Hilbert space, both GEIM and the least-squares method are optimal in a sense that will be clarified later in this paper. Last but not least, we would like to cite the even more recent work [7] on data assimilation in Banach spaces which can be seen as a further abstraction of GEIM and [14, 4].

Since the interpolation operator $\mathcal{J}_n : \mathcal{X} \to X_n$ built by the GEIM targets the elements of F, it is important to quantify the error

$$\max_{u \in F} \|u - \mathcal{J}_n[u]\|_{\mathcal{X}}. \tag{1}$$

The aim of this paper is to compare (1) with the best possible performance in an n dimensional subspace of $\mathcal{X}$, which is given by the Kolmogorov n-width of F in $\mathcal{X}$,

$$d_n(F, \mathcal{X}) \coloneqq \inf_{\substack{X_n \subset \mathcal{X} \\ \dim(X_n) \leq n}} \max_{u \in F} \inf_{v \in X_n} \|u - v\|_{\mathcal{X}}. \tag{2}$$

In particular, we show that when $d_n(F, \mathcal{X})$ decays polynomially or exponentially, the interpolation error has the same behavior modulo a coefficient depending on the norm of the interpolation operator

$$\Lambda_n \coloneqq \sup_{\varphi \in \mathcal{X}} \frac{\|\mathcal{J}_n[\varphi]\|_{\mathcal{X}}}{\|\varphi\|_{\mathcal{X}}}, \tag{3}$$

called Lebesgue constant. The conditions required on F to guarantee a certain decay rate in the sequence $(d_n(F, \mathcal{X}))_{n \geq 1}$ are far from trivial and require some regularity of the functions of F (see [18] for a general reference). For the particular case when F is the solution manifold of a parametric PDE, we refer to [6] in which the decay of $(d_n(F, \mathcal{X}))_{n \geq 1}$ is connected with the regularity of the manifold with respect to the parameters.

The paper is organized as follows: in section 2 we explain the approximation strategy and the greedy algorithm of the GEIM. We also show that, under appropriate hypothesis, the greedy algorithm is of a weak type in the sense of [3]. In section 3, we make several remarks on its practical implementation. Then we present the main results of the paper and derive convergence rates of the interpolation error in Banach spaces in section 4. In section 5, improved results are given for the particular case of Hilbert spaces. Note that results on convergence rates of GEIM have already been given in [12] for $\mathcal{X} = L^2(\Omega)$ and the main contribution of this work is to extend the results to general Banach spaces. Our methodology is based on the works [3, 8] on convergence rates of reduced basis and should be seen as a generalization of them in a sense that will be specified in 4.2.

2 The Generalized Empirical Interpolation Method

In what follows, we assume that the dimension of the vector space spanned by F is larger than $\mathcal{N}$, where $\mathcal{N}$ is some given large number depending on F ($\mathcal{N}$ could potentially be infinite). This hypothesis is made since the asymptotic decay of the interpolation error is trivial otherwise. We also suppose given a dictionary $\Sigma \subset \mathcal{X}'$ of bounded linear functionals with the following properties:

P1: $\forall \sigma \in \Sigma$, $\|\sigma\|_{\mathcal{X}'} = 1$.

P2: *Unisolvence*: If $\varphi \in \text{span}\{F\}$ is such that $\sigma(\varphi) = 0$, $\forall \sigma \in \Sigma$, then $\varphi = 0$.

P3: $\forall \varphi \in \text{span}\{F\}$ the set $\{\sigma(\varphi), \sigma \in \Sigma\} \subset \mathbb{R}$ is closed.

For all $n < \mathcal{N}$, our goal is to build a sequence of n-dimensional subspaces of $\mathcal{X}$: $X_n = \text{span}\{\varphi_i \in \mathcal{X}\}_{i=0}^{n-1}$, that approximate well enough the elements of F. As already mentioned in the introduction, one of the key features that makes the approximation adapted to F is that the basis functions φ_i, $0 \leq i \leq n-1$, are chosen in F itself. The approximation in X_n of any $\varphi \in \mathcal{X}$ is done with its so-called generalized interpolant $\mathcal{J}_n[\varphi] \in X_n$ which satisfies the following interpolation property for a set of "well-chosen" bounded linear functionals $\{\sigma_0, \sigma_1, \ldots, \sigma_{n-1}\}$ of Σ:

$$\mathcal{J}_n[\varphi] = \sum_{j=0}^{n-1} \beta_j \varphi_j, \quad \text{such that} \quad \sigma_i(\mathcal{J}_n[\varphi]) = \sigma_i(\varphi), \ \forall i = 0, \ldots, n-1. \tag{4}$$

The construction of the spaces X_n and the selection of the suitable bounded linear functionals of Σ is recursively carried out by a greedy algorithm where the basis functions are selected one after another in F and tuned to improve the accuracy of the interpolation process. In general, from a practical point of view, since F is an infinite compact set, the n^{th} function is chosen in a finite, large enough set F_n, where $\{F_n\}_{n\in\mathbb{N}}$ is an embedded sequence of subsets of F. Similarly $\{\widetilde{\Sigma}_n\}_{n\in\mathbb{N}}$ is an embedded sequence of finite subsets of Σ. For any $n \in \mathbb{N}$, $\widetilde{\Sigma}_n$ satisfies P2 and P3 for the set F_n. In section 3, we discuss how F_n can be chosen in practice. At the initialization of the greedy algorithm, the first interpolating function φ_0 is chosen as

$$\varphi_0 = \arg\max_{\varphi\in F_0} \|\varphi\|_{\mathcal{X}}, \tag{5}$$

and we define $X_1 \coloneqq \text{span}\{\varphi_0\}$. The first interpolating bounded linear functional is

$$\sigma_0 = \arg\max_{\sigma\in\widetilde{\Sigma}_0} |\sigma(\varphi_0)|.$$

The interpolation operator $\mathcal{J}_1 : \mathcal{X} \mapsto X_1$ is defined such that (4) is true for $n = 1$, i.e., $\sigma_0(\mathcal{J}_1[\varphi]) = \sigma_0(\varphi)$, for any $\varphi \in \mathcal{X}$. To facilitate the computation of the generalized interpolant, we express it in terms of $q_0 \coloneqq \varphi_0/\|\varphi_0\|_{\mathcal{X}}$. In this basis, the interpolant reads $\mathcal{J}_1[\varphi] = \sigma_0(\varphi)q_0$, for any $\varphi \in \mathcal{X}$. We then proceed by induction. Assume that, for a given $1 \leq n < \mathcal{N}$, we built the set of interpolating functions $\{q_0, q_1, \ldots, q_{n-1}\}$ and the set of associated interpolating bounded linear functionals $\{\sigma_0, \sigma_1, \ldots, \sigma_{n-1}\}$ such that the operator $\mathcal{J}_n[\varphi] = \sum_{j=0}^{n-1} \alpha_j^n(\varphi)q_j$, is well defined for any $\varphi \in \mathcal{X}$ and the coefficients $\{\alpha_j^n(\varphi)\}_{j=0,\ldots,n-1}$, are given by the interpolation problem

$$\begin{cases} \text{Find } \left(\alpha_j^n(\varphi)\right)_{j=0}^{n-1} \text{ such that:} \\ \sum_{j=0}^{n-1} \alpha_j^n(\varphi)\sigma_i(q_j) = \sigma_i(\varphi), \quad \forall i = 0, \ldots, n-1. \end{cases}$$

We now define $\varepsilon_n(\varphi) \coloneqq \|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}}$, for any $\varphi \in \mathcal{X}$, and choose φ_n such that

$$\varphi_n = \arg\max_{\varphi\in F_n} \varepsilon_n(\varphi) \tag{6}$$

and then σ_n such that

$$\sigma_n = \arg\max_{\sigma\in\widetilde{\Sigma}_n} |\sigma(\varphi_n - \mathcal{J}_n[\varphi_n])|.$$

Note that the existence of σ_n is ensured by the property P3 of the dictionary $\widetilde{\Sigma}_n$. The next basis function is then $q_n = (\varphi_n - \mathcal{J}_n[\varphi_n])/\sigma_n(\varphi_n - \mathcal{J}_n[\varphi_n])$. We finally set $X_{n+1} \coloneqq \text{span}\{q_j,\ j \in [0, n]\} = \text{span}\{\varphi_j,\ j \in [0, n]\}$. The interpolation operator $\mathcal{J}_{n+1} : \mathcal{X} \mapsto X_{n+1}$ is given by

$$\forall \varphi \in \mathcal{X}, \quad \mathcal{J}_{n+1}[\varphi] = \sum_{j=0}^{n} \alpha_j^{n+1}(\varphi)q_j$$

and the coefficients $\alpha_j^{n+1}(\varphi)$, $j = 0, \ldots, n$, are given by the interpolation problem

$$\begin{cases} \text{Find } \left(\alpha_j^{n+1}(\varphi)\right)_{j=0}^{n} \text{ such that:} \\ \sum_{j=0}^{n} \alpha_j^{n+1}(\varphi)\sigma_i(q_j) = \sigma_i(\varphi), \quad \forall i = 0, \ldots, n. \end{cases}$$

It has been proven in [13] (for EIM) and [11] (for GEIM) that for any $1 \leq n \leq \mathcal{N}$, the set $\{q_j,\ j \in [0, n-1]\}$ is linearly independent and that this interpolation procedure is well-posed in $\mathcal{X}$. This follows from the fact that the matrix $(\sigma_i(q_j))_{0\leq i,j\leq n-1}$ is lower triangular with diagonal entries equal to 1. Using the triangular inequality it is standard to derive the following inequality on the interpolation error

$$\forall \varphi \in \mathcal{X}, \quad \|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}} \leq (1 + \Lambda_n) \inf_{\psi_n\in X_n} \|\varphi - \psi_n\|_{\mathcal{X}}, \tag{7}$$

where Λ_n is the Lebesgue constant, i.e., the norm of the interpolation operator $\mathcal{J}_n : \mathcal{X} \to X_n$ defined in (3). Note that nothing at this point ensures that $\sup_{\varphi\in F} \|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}}$ is close to $d_n(F,\mathcal{X})$. One of the main reasons is that the basis set of X_n has been derived in a hierarchical manner with the greedy algorithm. It is actually the purpose of this paper to connect these two quantities. As we will see in sections 4 and 5, the Lebesgue constant will be involved in the bounds and it is therefore important to discuss its behavior as n increases. First of all, Λ_n depends both on F and Σ. In a practical implementation, it also depends on the choice of the subsets F_i and $\widetilde{\Sigma}_i$, $0 \le i \le n-1$. This observation is particularly clear in the case of Hilbert spaces where $\Lambda_n = 1/\beta_n$, with

$$\beta_n \coloneqq \inf_{\varphi\in X_n} \sup_{\sigma\in \mathrm{span}\{\sigma_0,\dots,\sigma_{n-1}\}} \frac{\langle \varphi,\sigma\rangle_{\mathcal{X},\mathcal{X}'}}{\|\varphi\|_{\mathcal{X}}\|\sigma\|_{\mathcal{X}'}}$$

and the bound (7) holds without the 1 on the right hand side (see [11]). In addition, it has recently been proven in [4] that, in this setting, $1/\beta_n$ is the smallest constant to relate $\max_{\varphi\in F} \|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}}$ and $\max_{\varphi\in F} \inf_{\psi_n\in X_n} \|\varphi - \psi_n\|_{\mathcal{X}}$. A (generally very pessimistic) bound for Λ_n that does not take into account the dependence on F and Σ was derived in [10] and reads

$$\Lambda_n \le 2^{n-1} \max_{i\in[0,n-1]} \|q_i\|_{\mathcal{X}}. \tag{8}$$

In [13], an example that achieves the bound (8) was built but it involves a set F with large n-width in $\mathcal{X}$. This does not correspond to the current setting since we assume that the sequence $(d_n(F,\mathcal{X}))$ has a relatively fast decay rate. In this context, numerical evaluations of Λ_n indicate that it increases polynomialy (with small degree) in the worst case scenario (see, e.g., [1], [9], [13] for examples in the case of the EIM and [11] for GEIM). For this reason, the conjecture that Λ_n depends mildly on the dimension when the n-width of F is small and/or decays with n seems reasonable.

3 Some comments on the practical implementation of the greedy algorithm

Let us now discuss the choice of the finite sets F_n for $n \in \mathbb{N}$. These are to be chosen large enough such that the search over F_n of $\arg\max_{\varphi\in F_n} \varepsilon_n(\varphi)$ (in (6)) is "close enough" to the same search over F. The following lemma quantifies this idea:

Lemma 3.1 *For any $0 < \eta < 1$, there exits a family $\{F_n \equiv F_n^\eta\}_{n=0}^{\mathcal{N}}$ of finite subsets of F such that the greedy algorithm satisfies*

$$\begin{cases} \max\limits_{\varphi\in F_0^\eta} \|\varphi\|_{\mathcal{X}} \ge \eta \max\limits_{\varphi\in F} \|\varphi\|_{\mathcal{X}}, \\ \max\limits_{\varphi\in F_n^\eta} \|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}} \ge \eta \max\limits_{\varphi\in F} \|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}}, \quad \forall\ n \in \{1,\dots,\mathcal{N}\}. \end{cases} \tag{9}$$

Proof. We first address the case $n = 0$ as follows. Let $\eta_0 = \eta$. By compactness of F, there exists a finite subset $F_0^\eta \subset F$ and a function $\tilde{\varphi}_0 \in F$ such that

$$\max_{\varphi\in F} \min_{\psi\in F_0^\eta} \|\varphi - \psi\|_{\mathcal{X}} \le (1-\eta_0)\|\tilde{\varphi}_0\|_{\mathcal{X}}. \tag{10}$$

Let $\varphi_0 = \arg\max\limits_{\psi\in F_0^\eta} \|\psi\|_{\mathcal{X}}$ and $\varphi_0^{\max} = \arg\max\limits_{\varphi\in F} \|\varphi\|_{\mathcal{X}}$. For any $\psi \in F_0^\eta$,

$$\|\varphi_0\|_{\mathcal{X}} \ge \|\psi\|_{\mathcal{X}} \ge -\|\psi - \varphi_0^{\max}\|_{\mathcal{X}} + \|\varphi_0^{\max}\|_{\mathcal{X}},$$

from which we infer that

$$\|\varphi_0\|_{\mathcal{X}} \ge -\min_{\psi\in F_0^\eta} \|\psi - \varphi_0^{\max}\|_{\mathcal{X}} + \|\varphi_0^{\max}\|_{\mathcal{X}},$$

which, from (10), yields

$$\|\varphi_0\|_{\mathcal{X}} \ge -(1-\eta_0)\|\tilde{\varphi}_0\|_{\mathcal{X}} + \|\varphi_0^{\max}\|_{\mathcal{X}} \ge \eta_0 \|\varphi_0^{\max}\|_{\mathcal{X}}.$$

This completes the proof of the first inequality of (9). For any $1 \leq n \leq \mathcal{N}$, let

$$\eta_n = 1 - (1-\eta)/(1+\Lambda_n). \tag{11}$$

We define $F_n^\eta \coloneqq F_{n-1}^\eta \cup \Xi_n^{\eta_n}$, where $\Xi_n^{\eta_n}$ is a finite subset of F such that

$$\max_{\varphi \in F} \min_{\psi \in \Xi_n^{\eta_n}} \|\varphi - \psi\|_{\mathcal{X}} \leq (1-\eta_n)\|r_n[\tilde{\varphi}_n]\|_{\mathcal{X}}, \tag{12}$$

for some $\tilde{\varphi}_n \in F$ and where $r_n[\varphi] \coloneqq \varphi - \mathcal{J}_n[\varphi],\ \forall \varphi \in \mathcal{X}$. The existence of $\Xi_n^{\eta_n}$ and $\tilde{\varphi}_n$ follows from the compactness of F, the fact that $r_n : \mathcal{X} \to \mathcal{X}$ is continuous (with a norm that is upper-bounded by $1+\Lambda_n$) and that $r_n(F)$ is a compact subset of $\mathcal{X}$. Using that $\Xi_n^{\eta_n} \subset F_n^\eta$ and inequality (7), we derive

$$\max_{\varphi \in F} \min_{\psi \in F_n^\eta} \|r_n[\varphi - \psi]\|_{\mathcal{X}} \leq \max_{\varphi \in F} \min_{\psi \in \Xi_n^{\eta_n}} \|r_n[\varphi - \psi]\|_{\mathcal{X}} \leq (1+\Lambda_n) \max_{\varphi \in F} \min_{\psi \in \Xi_n^{\eta_n}} \|\varphi - \psi\|_{\mathcal{X}}.$$

Thus, by using (12) in the previous inequality, we have

$$\max_{\varphi \in F} \min_{\psi \in F_n^\eta} \|r_n[\varphi - \psi]\|_{\mathcal{X}} \leq (1+\Lambda_n)(1-\eta_n)\|r_n[\tilde{\varphi}_n]\|_{\mathcal{X}} = (1-\eta)\|r_n[\tilde{\varphi}_n]\|_{\mathcal{X}}. \tag{13}$$

Let $\varphi_n = \arg\max_{\psi \in F_n^\eta} \|r_n[\psi]\|_{\mathcal{X}}$ and $\varphi_n^{\max} = \arg\max_{\varphi \in F} \|r_n[\varphi]\|_{\mathcal{X}}$. For any $\psi \in F_n^\eta$,

$$\|r_n[\varphi_n]\|_{\mathcal{X}} \geq \|r_n[\psi]\|_{\mathcal{X}} \geq -\|r_n[\psi - \varphi_n^{\max}]\|_{\mathcal{X}} + \|r_n[\varphi_n^{\max}]\|_{\mathcal{X}},$$

which, by using inequality (13), finally yields

$$\|\varphi_n - \mathcal{J}_n[\varphi_n]\|_{\mathcal{X}} \geq -(1-\eta)\|r_n[\tilde{\varphi}_n]\|_{\mathcal{X}} + \|r_n[\varphi_n^{\max}]\|_{\mathcal{X}} \geq \eta\ \|\varphi_n^{\max} - \mathcal{J}_n[\varphi_n^{\max}]\|_{\mathcal{X}}.$$

which ends the proof. ■

The parameter η above quantifies to what extent the search over $F_n = F_n^\eta$ differs from the search over F in the greedy algorithm. This relaxation expressed in the form of (9) is known as the weak greedy algorithm (in the sense defined in section 1.3 of [3]). We set $\eta = 1$ as an extreme case where $F_n^\eta = F$ (this is the strong greedy algorithm). Then, the smaller the η, the coarser the search over F_n^η will be in comparison with a search over F.

An important point to note is that the construction above depends on the application of the finite covering property of compact sets. Hence the question: how to obtain the sets $\{F_n\}_{n=0}^{\mathcal{N}}$ in practice? In full generality, this task is not entirely possible since it requires optimizations over the whole set F. However, the problem becomes feasible if we have some additional knowledge of the manifold F like, e.g., information about its geometry or regularity. As an example, let us consider the case where F is a set of parameter dependent functions $F = \{u(.;\mu); \mu \in \mathcal{D}\}$ where $\mathcal{D}$ is a compact set of parameters[1]. Then the derivability of the mapping $\mu \in \mathcal{D} \mapsto u(.;\mu)$ and a known uniform bound on this derivative with respect to μ (no regularity in the spacial direction is assumed here) allows to build a finite covering from a finite set in the compact set $\mathcal{D}$ in a completely constructive way.

Instead of working with such certified a priori adaptive subsets, another a posteriori adaptative option can be proposed following the arguments presented in [15] where a knowledge of the geometry of F is learnt on the fly as the greedy algorithm is implemented.

Note finally that, in many actual implementations, a less ideal approach is used where a fixed, unique, large enough, subset $F_{finite} \subset F$ (and a fixed subset $\widetilde{\Sigma}$) is chosen. In frequent cases where the Kolmogorov dimension is rapidly decaying to zero the greedy algorithm ends after very few iterations and this crude procedure actually works well in practice.

4 Convergence rates of the GEIM in a Banach space

In order to have consistent notations in what follows, we define $\varphi_n = 0$ and $X_n = X_{\mathcal{N}}$ for $n > \mathcal{N}$. In this section, $\mathcal{X}$ is a Banach space.

[1] Note that in the application of GEIM discussed in section 1, the compact set F is of this form.

4.1 Preliminary notations and properties

To fix some notations, let K be a nonempty subset of $\mathcal{X}$. For every $\varphi \in \mathcal{X}$, the distance between φ and the set K is

$$\text{dist}(\varphi, K) := \inf_{y \in K} \|\varphi - y\|_{\mathcal{X}}.$$

For any $\varphi \in \mathcal{X}$, the metric projection of φ onto K is the set

$$P_K(\varphi) = \{z \in K : \|\varphi - z\|_{\mathcal{X}} = \text{dist}(\varphi, K)\}.$$

In general, this set can be empty or composed of one or more than one element. However, in the particular case where K is a finite dimensional vector space, $P_K(\varphi)$ is not empty. For any $n \geq 1$, the non empty set

$$P_n(\varphi) = \{z \in X_n : \|\varphi - z\|_{\mathcal{X}} = \text{dist}(\varphi, X_n)\} \tag{14}$$

will denote the metric projection of $\varphi \in \mathcal{X}$ onto X_n. Since the uniqueness of the metric projection onto X_n is not necessarily ensured, in the following, $P_n(\varphi)$ will denote one of the elements of the set (14). We now define

$$\tau_n(F)_{\mathcal{X}} := \max_{f \in F} \|f - P_n(f)\|_{\mathcal{X}}, \quad n \geq 1. \tag{15}$$

Note that $\tau_n(F)_{\mathcal{X}} \leq \max_{f \in F} \|f\|_{\mathcal{X}} \leq 1$ given that the elements of F have norm less than 1. We will use the abbreviation τ_n and d_n for $\tau_n(F)_{\mathcal{X}}$ and $d_n(F, \mathcal{X})$. Likewise, $(\tau_n)_{n=1}^{\infty}$ and $(d_n)_{n=1}^{\infty}$ will denote the sequences $(\tau_n(F)_{\mathcal{X}})_{n=1}^{\infty}$ and $(d_n(F, \mathcal{X}))_{n=1}^{\infty}$ respectively. Finally we introduce the parameter

$$\gamma_n = \frac{\eta}{1 + \Lambda_n}, \quad \forall\ 1 \leq n \leq \mathcal{N}. \tag{16}$$

where η was introduced in (9) and Λ_n is the Lebesgue constant.

4.2 Main strategy to derive convergence rates

We start this section by proving the following lemma.

Lemma 4.1 *For any $n \geq 1$, the function φ_n defined in* (6) *verifies*

$$\|\varphi_n - P_n(\varphi_n)\|_{\mathcal{X}} \geq \gamma_n \tau_n. \tag{17}$$

Proof. From equation (7) applied to $\varphi = \varphi_n$ we have $\|\varphi_n - P_n(\varphi_n)\|_{\mathcal{X}} \geq \dfrac{1}{1 + \Lambda_n} \|\varphi_n - \mathcal{J}_n(\varphi_n)\|_{\mathcal{X}}$. But $\|\varphi_n - \mathcal{J}_n(\varphi_n)\|_{\mathcal{X}} \geq \eta \|\varphi - \mathcal{J}_n(\varphi)\|_{\mathcal{X}}$ for any $\varphi \in F$ according to the definition of φ_n. Thus $\|\varphi_n - P_n(\varphi_n)\|_{\mathcal{X}} \geq \gamma_n \|\varphi - \mathcal{J}_n(\varphi)\|_{\mathcal{X}} \geq \gamma_n \|\varphi - P_n(\varphi)\|_{\mathcal{X}}$. ■

Lemma 4.1 shows that the weak greedy algorithm of GEIM has very similar properties as the one in [8]. The difference is that, in [8], inequality (17) involves a constant parameter γ independent of n. We take this into account in our analysis: in section 4.3 we analyze the convergence rates for $(\tau_n)_{n=1}^{\infty}$ by extending the proofs of [8] to the case where γ depends on n. For the sake of comparison, we first recall here their main results in lemmas 4.2 and 4.3 below:

Lemma 4.2 (Corollary $4.2 - (ii)$ **of [8])** *If, for $\alpha > 0$, we have $d_n \leq C_0 n^{-\alpha}$, $n = 1, 2, \ldots$, then for any $0 < \beta < \min\{\alpha, 1/2\}$, we have $\tau_n \leq C_1 n^{-\alpha + 1/2 + \beta}$, $n = 1, 2, \ldots$, with*

$$C_1 := \max\left\{ C_0 4^{4\alpha+1} \gamma^{-4} \left(\frac{2\beta + 1}{2\beta} \right)^{\alpha} ; \max_{n=1,\ldots,7} n^{\alpha - \beta - 1/2} \right\}.$$

Lemma 4.3 (Corollary $4.2 - (iii)$ **of [8])** *If, for $\alpha > 0$, $d_n \leq C_0 e^{-c_1 n^{\alpha}}$, $n = 1, 2, \ldots$, then $\tau_n < \sqrt{2C_0} \gamma^{-1} \sqrt{n} e^{-c_2 n^{\alpha}}$, $n = 1, 2, \ldots$, where $c_2 = 2^{-1-2\alpha} c_1$. The factor $\sqrt{n}$ can be removed by reducing the constant c_2.*

Once this generalization is done, by using inequality (7), convergence rates for the interpolation error will easily follow (see section 4.4).

4.3 Convergence rates for $(\tau_n)_{n=1}^{\infty}$ in the case where $(\gamma_n)_{n=1}^{\infty}$ is not constant

We start by looking for an upper bound of the sequence $(\tau_n)_{n=1}^{\infty}$ that involves the sequence of Kolmogorov n-widths $(d_n)_{n=1}^{\infty}$. The case $n = 1$ is addressed in lemma 4.4. The case $n > 1$ is addressed in theorem 4.5. From this last theorem, we infer corollaries 4.6 and 4.7 that will be useful to derive convergence rates for $(\tau_n)_{n=1}^{\infty}$.

Lemma 4.4 *For $n = 1$, $\tau_1 \leq 2(1 + \eta^{-1})d_1$.*

Proof. Given the parameter η in the GEIM greedy algorithm, let us chose $\beta > 1/\eta$. We begin by recalling and introducing some notations. First of all, φ_0 is the first interpolating function chosen in (5) by the greedy algorithm and $X_1 = \text{span}\{\varphi_0\}$. For any φ, $P_1(\varphi)$ is the metric projection of φ onto X_1. Let $\varphi_0^{\max} = \arg\max_{\varphi\in F} \|\varphi\|_{\mathcal{X}}$. From (9) in the case $n = 0$, $\|\varphi_0\|_{\mathcal{X}} \geq \eta\|\varphi_0^{\max}\|_{\mathcal{X}}$. Let X_μ be a one dimensional subspace and $E_\mu \coloneqq \max_{x\in F} \min_{y\in X_\mu} \|x-y\|_{\mathcal{X}}$. For any $\varphi \in F$, $\|\varphi - P_{X_\mu}(\varphi)\|_{\mathcal{X}} \leq E_\mu$. Let $\varphi_\mu^{\max} = \arg\max_{\varphi\in F} \|P_{X_\mu}(\varphi)\|_{\mathcal{X}}$. We now divide the proof by considering two cases of values of $\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}}$. If $\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}} \leq \frac{1+\eta}{\eta-1/\beta}\ E_\mu$, for all $\varphi \in F$:

$$\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq \|\varphi\|_{\mathcal{X}} \leq \|\varphi - P_{X_\mu}(\varphi)\|_{\mathcal{X}} + \|P_{X_\mu}(\varphi)\|_{\mathcal{X}} \leq E_\mu + \|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}},$$

which yields

$$\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq \left(1 + \frac{1+\eta}{\eta - 1/\beta}\right) E_\mu, \quad \forall \varphi \in F. \tag{18}$$

If $\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}} \geq \frac{1+\eta}{\eta-1/\beta}\ E_\mu$, we have:

$$\|P_{X_\mu}(\varphi_0)\|_{\mathcal{X}} \geq \|\varphi_0\|_{\mathcal{X}} - E_\mu \geq \eta\|\varphi_0^{\max}\|_{\mathcal{X}} - E_\mu \geq \eta\|\varphi_\mu^{\max}\|_{\mathcal{X}} - E_\mu, \tag{19}$$

and thus

$$\|P_{X_\mu}(\varphi_0)\|_{\mathcal{X}} \geq \eta\left(\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}} - E_\mu\right) - E_\mu \geq \eta\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}} - (1+\eta)E_\mu,$$

from which we infer that

$$\|P_{X_\mu}(\varphi_0)\|_{\mathcal{X}} \geq \|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}}/\beta. \tag{20}$$

From inequality (20), it follows that $\|P_{X_\mu}(\varphi_0)\|_{\mathcal{X}} > 0$ given that $\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}}$ is positive. Furthermore, for any $\varphi \in \mathcal{X}$, there exits $\lambda_\varphi \in \mathbb{R}$ such that:

$$P_{X_\mu}(\varphi) = \lambda_\varphi P_{X_\mu}(\varphi_0). \tag{21}$$

Hence the decomposition:

$$\begin{aligned} \varphi &= P_{X_\mu}(\varphi) + \varphi - P_{X_\mu}(\varphi) = \lambda_\varphi P_{X_\mu}(\varphi_0) + \varphi - P_{X_\mu}(\varphi) \\ &= \lambda_\varphi(P_{X_\mu}(\varphi_0) - \varphi_0) + \lambda_\varphi\varphi_0 + \varphi - P_{X_\mu}(\varphi). \end{aligned} \tag{22}$$

Since $\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq \|\varphi - \lambda_\varphi\varphi_0\|_{\mathcal{X}}$, we can use equation (22) to bound $\|\varphi - \lambda_\varphi\varphi_0\|_{\mathcal{X}}$ and write

$$\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq |\lambda_\varphi|\|P_{X_\mu}(\varphi_0) - \varphi_0\|_{\mathcal{X}} + \|\varphi - P_{X_\mu}(\varphi)\|_{\mathcal{X}} \leq (1 + |\lambda_\varphi|)E_\mu.$$

Furthermore, given that $\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}} \geq \|P_{X_\mu}(\varphi)\|_{\mathcal{X}}$ for any $\varphi \in F$, we have

$$\|P_{X_\mu}(\varphi_\mu^{\max})\|_{\mathcal{X}} \geq |\lambda_\varphi|\|P_{X_\mu}(\varphi_0)\|_{\mathcal{X}}, \tag{23}$$

where we have used equality (21). Inequalities (20) and (23) yield $|\lambda_\varphi| \leq \beta$ and therefore $\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq (1 + \beta)E_\mu$. As a result, we have proven that for any $\beta > 1/\eta$ and any $\varphi \in F$, $\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq g_\eta(\beta)E_\mu$, where

$$\forall \beta > 1/\eta, \quad g_\eta(\beta) := \max\left(1 + \beta; 1 + \frac{1+\eta}{\eta - 1/\beta}\right).$$

Thus $\|\varphi - P_1(\varphi)\|_{\mathcal{X}} \leq \min_{\beta>1/\eta} g_\eta(\beta)E_\mu = 2(1+\eta^{-1})E_\mu$. Since the inequality is valid for any one dimensional space X_μ, the final result follows by taking the infimum over all one dimensional spaces in $\mathcal{X}$. ∎

Theorem 4.5 *For any $N \geq 0$, consider a weak greedy algorithm for which* (17) *holds. Then, for any $K \geq 2,\ 1 \leq m < K$*

$$\prod_{i=1}^{K} \tau_{N+i}^2 \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^2} 2^K K^{K-m} \left(\sum_{i=1}^{K} \tau_{N+i}^2 \right)^m d_m^{2(K-m)}. \tag{24}$$

Proof. This result is an extension of theorem 4.1 of [8] to the case where the parameter γ depends on N, the dimension of the space X_N. The proof is a slight modification of the one in [8] but we provide it in the appendix for the self-consistency of this paper. ■

Corollary 4.6 *For any $n \geq 2$,*

$$\tau_n \leq \frac{1}{\prod_{i=1}^{n} \gamma_i^{1/n}} \sqrt{2} \min_{1 \leq m < n} \left\{ n^{\frac{n-m}{2n}} \left(\sum_{i=1}^{n} \tau_i^2 \right)^{\frac{m}{2n}} d_m^{\frac{n-m}{n}} \right\} \tag{25}$$

In particular, for any $\ell \geq 1$:

$$\tau_{2\ell} \leq 2 \frac{1}{\prod_{i=1}^{2\ell} \gamma_i^{1/2\ell}} \sqrt{\ell d_\ell}. \tag{26}$$

Proof. We take $N = 0$, $K = n$ and any $1 \leq m < n$ in (24) and use the monotonicity of $(\tau_n)_{n=1}^{\infty}$ to obtain:

$$\tau_n^{2n} \leq \prod_{i=1}^{n} \tau_i^2 \leq \frac{1}{\prod_{i=1}^{n} \gamma_i^2} 2^n n^{n-m} \left(\sum_{i=1}^{n} \tau_i^2 \right)^m d_m^{2(n-m)},$$

that yields (25). In particular, if $n = 2\ell$ and $m = \ell$, we have:

$$\tau_{2\ell} \leq \frac{1}{\prod_{i=1}^{2\ell} \gamma_i^{1/2\ell}} \sqrt{2} (2\ell)^{1/4} \left(\sum_{i=1}^{2\ell} \tau_i^2 \right)^{1/4} \sqrt{d_\ell} \leq \frac{1}{\prod_{i=1}^{2\ell} \gamma_i^{1/2\ell}} \sqrt{2} (2\ell)^{1/4} (2\ell)^{1/4} \sqrt{d_\ell} = 2 \frac{1}{\prod_{i=1}^{2\ell} \gamma_i^{1/2\ell}} \sqrt{\ell d_\ell},$$

where we used that all $\tau_i \leq 1$. ■

Corollary 4.7 *For $N \geq 0$, $K \geq 2$ and $1 \leq m < K$:*

$$\tau_{N+K} \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{1/K}} \sqrt{2K} \tau_{N+1}^{m/K} d_m^{1-m/K} \tag{27}$$

Proof. Using that $(\tau_n)_{n=1}^{\infty}$ is monotonically decreasing and following the same lines as above, we derive from inequality (24) that:

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

Therefore,

$$\tau_{N+K} \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{1/K}} \sqrt{2} K^{\frac{K-m}{2K}} \left(K \tau_{N+1}^2 \right)^{m/2K} d_m^{1-m/K} \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{1/K}} \sqrt{2K} \tau_{N+1}^{m/K} d_m^{1-m/K}.$$

■

We now derive convergence rates for the sequence $(\tau_n)_{n=1}^{\infty}$. As we will see, the convergence and its rate strongly depend on the behavior of the Lebesgue constant. As discussed in section 2, $(\Lambda_n)_{n=1}^{\infty}$ can

diverge exponentially. On the contrary and as it is often the case in practice, it can be polynomially increasing or even be bounded. We take this point into account by assuming different types of convergence rates for $(d_n)_{n=1}^{\infty}$. To begin with, Lemmas 4.8 and 4.9 below consider the case where $(d_n)_{n=1}^{\infty}$ decreases polynomially or exponentially. No assumption on the behavior of $(\Lambda_n)_{n=1}^{\infty}$ is made in these results. With the reference to γ_n introduced in (16), we have

Lemma 4.8 *For any $n \geq 1$, let $n = 4\ell + k$ (where $\ell \in \{0, 1, \dots\}$ and $k \in \{0, 1, 2, 3\}$). If, for $n \geq 1$, $d_n \leq C_0 n^{-\alpha}$, with $C_0 > 0$, then $\tau_n \leq C_0 \beta_n n^{-\alpha}$, where*

$$\beta_n = \beta_{4\ell+k} := \begin{cases} 2\left(1+\eta^{-1}\right) & \text{if } n = 1 \\ \dfrac{1}{\prod_{i=1}^{\ell_2} \gamma_{\ell_1 - \lceil \frac{k}{4} \rceil + i}^{\frac{1}{\ell_2}}} \sqrt{2\ell_2 \beta_{\ell_1}} (2\sqrt{2})^{\alpha} & \text{if } n \geq 2, \end{cases}$$

and $\ell_1 = 2\ell + \lfloor \frac{2k}{3} \rfloor$, $\ell_2 = 2\left(\ell + \lceil \frac{k}{4} \rceil\right)$, where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions respectively.

Proof. The proof is done by induction over n and the case $n = 1$ directly follows from lemma 4.4. In the case $n \geq 2$, we write $n = N + K$ with $N \geq 0$ and $K \geq 2$. Thanks to corollary 4.7, we have for any $1 \leq m < K$:

$$\tau_n = \tau_{N+K} \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{1/K}} \sqrt{2K} \tau_{N+1}^{m/K} d_m^{1-m/K}. \tag{28}$$

We now use that $d_m \leq C_0 m^{-\alpha}$ and the recurrence hypothesis at index $N+1$ that states $\tau_{N+1} \leq C_0 \beta_{N+1} (N+1)^{-\alpha}$ which yields:

$$\tau_{N+K} \leq C_0 \sqrt{2K} \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{\frac{1}{K}}} \beta_{N+1}^{\frac{m}{K}} \xi(N, K, m)^{\alpha} (N+K)^{-\alpha}, \tag{29}$$

where $\xi(N, K, m) = \dfrac{N+K}{m} \left(\dfrac{m}{N+1} \right)^{\frac{m}{K}}$ for any $1 \leq m < K$ and any given index $n = N + K \geq 2$, where $N \geq 0$ and $K \geq 2$. Furthermore, any $n \geq 2$ can be written as $n = 4\ell + k$ with $\ell \in \mathbb{N}$ and $k \in \{0, 1, 2, 3\}$. If $k = 1, 2$ or 3, it can easily be proven that the function ξ is bounded by $2\sqrt{2}$ by setting

$$\begin{cases} N = 2\ell - 1,\ K = 2\ell + 2,\ m = \ell + 1 \text{ and } \ell \geq 1 \text{ in the case } k = 1, \\ N = 2\ell,\ K = 2\ell + 2,\ m = \ell + 1 \text{ and } \ell \geq 0 \text{ in the case } k = 2, \\ N = 2\ell + 1,\ K = 2\ell + 2,\ m = \ell + 1 \text{ and } \ell \geq 0 \text{ in the case } k = 3. \end{cases}$$

These choices of N, K and m combined with the upper bound of ξ yield the result $\tau_n \leq C_0 \beta_n n^{-\alpha}$ in the case $k = 1, 2$ or 3. To address the case $n = 4\ell$, we come back to estimate (28) and use that $\tau_{N+1} \leq \tau_N$. It follows:

$$\tau_n \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{1/K}} \sqrt{2K} \tau_N^{m/K} d_m^{1-m/K}. \tag{30}$$

Choosing $N = K = 2\ell$ and $m = \ell$, the inequality (30) directly yields the desired result:

$$\tau_{4\ell} \leq C_0 \sqrt{2} \sqrt{2\ell \beta_{2\ell}} \frac{1}{\prod_{i=1}^{2\ell} \gamma_{2\ell+i}^{\frac{1}{2\ell}}} (2\sqrt{2})^{\alpha} (4\ell)^{-\alpha}.$$

■

Lemma 4.9 *If, for $n \geq 1$, $d_n \leq C_0 e^{-c_1 n^{\alpha}}$, with $C_0 \geq 1$ and $\alpha > 0$, then $\tau_n \leq C_0 \beta_n e^{-c_2 n^{\alpha}}$, where $c_2 := c_1 2^{-2\alpha - 1}$ and*

$$\beta_n := \begin{cases} 2\left(1+\eta^{-1}\right), & \text{if } n = 1 \\ \sqrt{2} \dfrac{1}{\prod_{i=1}^{2\lfloor \frac{n}{2} \rfloor} \gamma_i^{\frac{1}{2\lfloor \frac{n}{2} \rfloor}}} \sqrt{n}, & \text{if } n \geq 2. \end{cases}$$

Proof. The case $n = 1$ easily follows from lemma 4.4. For $n = 2\ell$ ($\ell \geq 1$), inequality (26) directly yields:

$$\tau_{2\ell} \leq 2\frac{1}{\prod_{i=1}^{2\ell}\gamma_i^{1/2\ell}}\sqrt{\ell d_\ell} \leq C_0\sqrt{2}\frac{1}{\prod_{i=1}^{2\ell}\gamma_i^{1/2\ell}}\sqrt{2\ell}e^{-\frac{c_1}{2^{1+\alpha}}(2\ell)^\alpha}, \tag{31}$$

where we used that $d_\ell \leq C_0 e^{-c_1(\ell)^\alpha}$ and that $C_0 \geq 1$. For $n = 2\ell + 1$, by using inequality (31) and $\tau_{2\ell+1} \leq \tau_{2\ell}$, we have:

$$\tau_{2\ell+1} \leq C_0\sqrt{2}\frac{1}{\prod_{i=1}^{2\ell}\gamma_i^{1/2\ell}}\sqrt{2\ell}e^{-\frac{c_1}{2^{1+\alpha}}(2\ell)^\alpha} \leq C_0\sqrt{2}\frac{1}{\prod_{i=1}^{2\ell}\gamma_i^{1/2\ell}}\sqrt{2\ell+1}e^{-\frac{c_1}{2^{1+2\alpha}}(2\ell+1)^\alpha}. \tag{32}$$

■

From lemmas 4.8 and 4.9 we observe that, if $(\Lambda_n)_{n=1}^\infty$ diverges exponentially, then an exponential decay is required for $(d_n)_{n=1}^\infty$. Let us now derive some results by adding different assumptions in the behavior of $(\Lambda_n)_{n=1}^\infty$. In corollary 4.10, we assume that $(\Lambda_n)_{n=1}^\infty$ is monotonically increasing (i.e., $(\gamma_n)_{n=1}^\infty$ monotonically decreasing).

Corollary 4.10 *Assume that $(\Lambda_n)_{n=1}^\infty$ is monotonically increasing, then:*

i) If $d_n \leq C_0 n^{-\alpha}$ for any $n \geq 1$, then $\tau_n \leq C_0\tilde{\beta}_n n^{-\alpha}$, with

$$\tilde{\beta}_n := \begin{cases} 2\left(1+\eta^{-1}\right), & \text{if } n = 1 \\ 2^{3\alpha+1}\ell_2\gamma_n^{-2}, & \text{if } n \geq 2. \end{cases}$$

If we write n as $n = 4\ell + k$ (with $\ell \in \{0, 1, \dots\}$ and $k \in \{0, 1, 2, 3\}$), then $\ell_2 = 2\left(\ell + \lceil \frac{k}{4} \rceil\right)$.

ii) If $d_n \leq C_0 e^{-c_1 n^\alpha}$ for $n \geq 1$ and $C_0 \geq 1$, then $\tau_n \leq C_0\tilde{\beta}_n e^{-c_2 n^{-\alpha}}$, with $c_2 = c_1 2^{-2\alpha-1}$ and

$$\tilde{\beta}_n := \begin{cases} 2\left(1+\eta^{-1}\right), & \text{if } n = 1 \\ \sqrt{2n}\gamma_n^{-1}, & \text{if } n \geq 2. \end{cases}$$

Proof.

i) We show by induction that $\tilde{\beta}_n$ is larger than the coefficient β_n defined in lemma 4.8. If $n = 1$, $\tilde{\beta}_1 = \beta_1$. Then, for $n > 1$, given that $(\gamma_n)_{n=1}^\infty$ is monotonically decreasing,

$$\beta_n \leq \gamma_n^{-1}\sqrt{2\ell_2\beta_{\ell_1}}\left(2\sqrt{2}\right)^\alpha \leq \gamma_n^{-1}\sqrt{2\ell_2\tilde{\beta}_{\ell_1}}\left(2\sqrt{2}\right)^\alpha,$$

where we used the recurrence hypothesis $\beta_{\ell_1} \leq \tilde{\beta}_{\ell_1}$ in the second inequality. Furthermore, since $\tilde{\beta}_{\ell_1} \leq 2^{3\alpha+1}\ell_2\gamma_n^{-2}$, it follows that:

$$\beta_n \leq \gamma_n^{-1}\sqrt{2\ell_2 2^{3\alpha+1}\ell_2\gamma_n^{-2}}\left(2\sqrt{2}\right)^\alpha = 2^{3\alpha+1}\ell_2\gamma_n^{-2} = \tilde{\beta}_n.$$

ii) The result is straightforward and follows from the definition of β_n given in lemma 4.9.

■

If $(\Lambda_n)_{n=1}^\infty$ is constant, corollary 4.10 shows that we obtain exactly the same result as the one derived in [8] for the exponential case (recalled in lemma 4.3). In the case of polynomial decay, the result of corollary 4.10 provides a slightly degraded result with respect to the one in [8] (recalled in lemma 4.2). The most important difference relies on the fact that in [8] a convergence rate of order $\mathcal{O}(n^{-\alpha+1/2+\varepsilon})$ is obtained whereas the present results yields a convergence in $\mathcal{O}(n^{-\alpha+1})$. It has so far not been possible to derive better convergence rates in the polynomial case for a general behavior of the sequence $(\Lambda_n)_{n=1}^\infty$. However, under the refined assumption

$$\Lambda_n = \mathcal{O}(n^\zeta),$$

which is a typical behavior observed in numerical applications, Lemma 4.11 below shows that, in this case, the convergence is of order $\mathcal{O}(n^{-\alpha+\zeta+1/2+\varepsilon})$, which is consistent with the result of [8], in case of a constant Λ_n ($\zeta = 0$).

Lemma 4.11 *If, for $n > 0$, $d_n \leq C_0 n^{-\alpha}$ and $\gamma_n^{-1} \leq C_\zeta n^\zeta$, with constants $C_0,\ C_\zeta,\ \alpha,\ \zeta > 0$, then for any $\beta > 1/2$, we have $\tau_n \leq C_1 n^{-\alpha+\zeta+\beta}$, $n \in \mathbb{N}^*$, where*

$$C_1 := \max\left\{C_0 2^{\frac{2\alpha^2}{\zeta}} \left(\frac{\zeta+\beta}{\beta-\frac{1}{2}}\right)^\alpha \max\left(1; C_\zeta^{\frac{\zeta+\beta}{\zeta}}\right);\ \max_{n=1,\ldots,2\lfloor 2(\zeta+\beta)\rfloor+1} n^{\alpha-\zeta-\beta}\right\}.$$

Note that in the above lemma, the constant β has no connection with β_n defined in section 2.

Proof. It follows from the monotonicity of $(\tau_n)_{n=1}^\infty$ and inequality (24) for $N = K = n$ and any $1 \leq m < n$ that:

$$\tau_{2n} \leq \sqrt{2n} \frac{1}{\prod_{i=1}^{n} \gamma_{n+i}^{1/n}} \tau_n^\delta d_m^{1-\delta}, \quad \delta := \frac{m}{n}. \tag{33}$$

Given $\beta > 1/2$, we define $m := \left\lfloor \frac{\beta-\frac{1}{2}}{\zeta+\beta} \right\rfloor + 1$ (so that $m < n$ for $n > 2(\zeta+\beta) > 2\zeta+1$). It follows that

$$\delta = \frac{m}{n} \in \left(\frac{\beta-\frac{1}{2}}{\zeta+\beta}, \frac{\beta-\frac{1}{2}}{\zeta+\beta} + \frac{1}{n}\right). \tag{34}$$

We prove our claim by contradiction. Suppose it is not true and M is the first value where $\tau_M > C_1 M^{-\alpha+\zeta+\beta}$. Clearly, because of the definition of C_1 and the fact that $\tau_n \leq 1$, we must have $M > 2\lfloor 2(\zeta+\beta)\rfloor + 1$ (since $M \geq 2\lfloor 2(\zeta+\beta)\rfloor + 2$). We first consider the case $M = 2n$, and therefore $n \geq \lfloor 2(\zeta+\beta)\rfloor + 1$. From (33), we have:

$$\begin{aligned} C_1(2n)^{-\alpha+\zeta+\beta} < \tau_{2n} &\leq \sqrt{2n} \frac{1}{\prod_{i=1}^{n} \gamma_{n+i}^{1/n}} \tau_n^\delta d_m^{1-\delta} \\ &\leq \sqrt{2n} C_\zeta (2n)^\zeta C_1^\delta n^{\delta(-\alpha+\zeta+\beta)} C_0^{1-\delta} (\delta n)^{-\alpha(1-\delta)}, \end{aligned} \tag{35}$$

where we have used the fact that $\tau_n \leq C_1 n^{-\alpha+\zeta+\beta}$ and $d_m \leq C_0 m^{-\alpha}$. It follows that

$$C_1^{1-\delta} < 2^{\alpha-\beta+\frac{1}{2}} C_\zeta C_0^{1-\delta} \delta^{-\alpha(1-\delta)} n^{\frac{1}{2}+\delta(\zeta+\beta)-\beta}$$

and therefore

$$C_1 < 2^{\frac{\alpha-\beta+\frac{1}{2}}{1-\delta}} C_\zeta^{\frac{1}{1-\delta}} C_0 \delta^{-\alpha} n^{\frac{\zeta+\beta}{1-\delta}\left(\delta - \frac{\beta-\frac{1}{2}}{\zeta+\beta}\right)}.$$

Since, for $n \geq \lfloor 2(\zeta+\beta)\rfloor + 1 > 2(\zeta+\beta)$, we have

$$\delta < \frac{\beta-\frac{1}{2}}{\zeta+\beta} + \frac{1}{n} < \frac{\beta}{\zeta+\beta}, \tag{36}$$

then,

$$\frac{1}{1-\delta} < \frac{\zeta+\beta}{\zeta}. \tag{37}$$

Hence,

$$\frac{\zeta+\beta}{1-\delta}\left(\delta - \frac{\beta-\frac{1}{2}}{\zeta+\beta}\right) < \left(\frac{\zeta+\beta}{1-\delta}\right)\frac{1}{n} < \frac{(\zeta+\beta)^2}{\zeta}\frac{1}{n}, \tag{38}$$

where we have used inequalities (36) and (37). By using (38), it follows that

$$n^{\frac{\zeta+\beta}{1-\delta}\left(\delta - \frac{\beta-\frac{1}{2}}{\zeta+\beta}\right)} < n^{\frac{(\zeta+\beta)^2}{\zeta}\frac{1}{n}} < 2^{\frac{(\zeta+\beta)^2}{\zeta}}. \tag{39}$$

This yields:

$$C_1 < 2^{\frac{\alpha-\beta+\frac{1}{2}}{1-\delta}} C_\zeta^{\frac{1}{1-\delta}} C_0 \delta^{-\alpha} 2^{\frac{(\zeta+\beta)^2}{\zeta}} < 2^{\left(\frac{\zeta+\beta}{\zeta}\right)(\alpha+\zeta+\frac{1}{2})} C_\zeta^{\frac{1}{1-\delta}} C_0 \delta^{-\alpha}. \tag{40}$$

Furthermore, for $-\alpha+\zeta+\beta < 0$ (which is the meaningful case), and using the fact that $\beta > \frac{1}{2}$, we have:

$$2^{\frac{\zeta+\beta}{\zeta}\left(\alpha+\zeta+\frac{1}{2}\right)} < 2^{\frac{\alpha}{\zeta}(\alpha+\zeta+\beta)} < 2^{\frac{2\alpha^2}{\zeta}} \quad \text{and} \quad C_\zeta^{\frac{1}{1-\delta}} < \max\left(1; C_\zeta^{\frac{\zeta+\beta}{\zeta}}\right). \tag{41}$$

Also, from (34), we have

$$\delta^{-\alpha} < \left(\frac{\zeta+\beta}{\beta-\frac{1}{2}}\right)^{\alpha}. \tag{42}$$

By inserting inequalities (41) and (42) in (40), the desired contradiction follows:

$$C_1 < C_0 2^{\frac{2\alpha^2}{\zeta}} \left(\frac{\zeta+\beta}{\beta-\frac{1}{2}}\right)^{\alpha} \max\left(1; C_\zeta^{\frac{\zeta+\beta}{\zeta}}\right).$$

Likewise, if $M = 2n+1$, then $M \geq 2\lfloor 2(\zeta+\beta)\rfloor + 3$, which implies that $n \geq \lfloor 2(\zeta+\beta)\rfloor + 1$:

$$C_1 2^{-\alpha+\zeta+\beta}(2n)^{-\alpha+\zeta+\beta} < C_1(2n+1)^{-\alpha+\zeta+\beta} < \tau_{2n+1} \leq \tau_{2n}. \tag{43}$$

But, since we have from equation (35)

$$\tau_{2n} \leq \sqrt{2n} C_\zeta (2n)^\zeta C_1^\delta n^{\delta(-\alpha+\zeta+\beta)} C_0^{1-\delta} (\delta n)^{-\alpha(1-\delta)}, \tag{44}$$

then, following the same argument as above, we get:

$$C_1 < C_0 2^{(\frac{\zeta+\beta}{\zeta})(\frac{1}{2}+2\alpha-\beta)} \left(\frac{\zeta+\beta}{\beta-\frac{1}{2}}\right)^{\alpha} \max\left(1; C_\zeta^{\frac{\zeta+\beta}{\zeta}}\right) < C_0 2^{\frac{2\alpha^2}{\zeta}} \left(\frac{\zeta+\beta}{\beta-\frac{1}{2}}\right)^{\alpha} \max\left(1; C_\zeta^{\frac{\zeta+\beta}{\zeta}}\right),$$

where we have used the fact that $\beta > 1/2$ in the last inequality. ■

4.4 Convergence rates of the interpolation error

Thanks to the convergence rates obtained for $(\tau_n)_{n=1}^\infty$ in section 4.3, the following rates are readily obtained for the interpolation error of GEIM.

Theorem 4.12

i) If $d_n \leq C_0 n^{-\alpha}$ for any $n \geq 1$, then $\|\varphi - \mathcal{J}_n[\varphi]\|_\mathcal{X} \leq (1+\Lambda_n) C_0 \beta_n n^{-\alpha}$ for any $\varphi \in F$, where β_n is given in lemma 4.8.

ii) If $d_n \leq C_0 e^{-c_1 n^\alpha}$ for $n \geq 1$ and $C_0 \geq 1$, $\|\varphi - \mathcal{J}_n[\varphi]\|_\mathcal{X} \leq (1+\Lambda_n) C_0 \beta_n e^{-c_2 n^\alpha}$ for any $\varphi \in F$, where β_n and c_2 are defined as in lemma 4.9.

Proof. It follows from equation (7) and the definition of τ_n that, $\forall \varphi \in F$, $\|\varphi - \mathcal{J}_n[\varphi]\|_\mathcal{X} \leq (1+\Lambda_n)\|\varphi - P_n(\varphi)\|_\mathcal{X} \leq (1+\Lambda_n)\tau_n$. We conclude the proof by bounding τ_n by using lemmas 4.8 and 4.9. ■
If $(\Lambda_n)_{n=1}^\infty$ is a monotonically increasing sequence, we have a more precise behavior:

Corollary 4.13 *Let $(\Lambda_n)_{n=1}^\infty$ be a monotonically increasing sequence. Then,*

i) if $d_n \leq C_0 n^{-\alpha}$ for any $n \geq 1$, then

$$\forall \varphi \in F, \quad \|\varphi - \mathcal{J}_n[\varphi]\|_\mathcal{X} \leq \begin{cases} 2C_0\left(1+\eta^{-1}\right)(1+\Lambda_1), & \text{if } n = 1. \\ C_0 2^{3\alpha+1} \ell_2 (1+\Lambda_n)^3 \eta^{-2} n^{-\alpha}, & \text{if } n \geq 2. \end{cases}$$

If we write n as $n = 4\ell + k$ (with $\ell \in \{0, 1, \dots\}$ and $k \in \{0,1,2,3\}$), then $\ell_2 = 2\left(\ell + \lceil \frac{k}{4} \rceil\right)$.

ii) if $d_n \leq C_0 e^{-c_1 n^\alpha}$ for $n \geq 1$ and $C_0 \geq 1$, then (remember $c_2 = c_1 2^{-2\alpha-1}$)

$$\forall \varphi \in F, \quad \|\varphi - \mathcal{J}_n[\varphi]\|_\mathcal{X} \leq \begin{cases} 2C_0\left(1+\eta^{-1}\right)(1+\Lambda_1), & \text{if } n = 1, \\ C_0\sqrt{2}(1+\Lambda_n)^2 \eta^{-1}\sqrt{n} e^{-c_2 n^\alpha}, & \text{if } n \geq 2, \end{cases}$$

iii) if $d_n \leq C_0 n^{-\alpha}$ and $\gamma_n^{-1} \leq C_\zeta n^\zeta$ for any $n \geq 1$, then for any $\beta > 1/2$,

$$\forall \varphi \in F, \quad \|\varphi - \mathcal{J}_n[\varphi]\|_\mathcal{X} \leq \eta C_\zeta C_1 n^{-\alpha+2\zeta+\beta},$$

where the parameter C_1 is defined as in lemma 4.11.

Proof. *i*) and *ii*) easily follow from corollary 4.10 and *iii*) is derived by using lemma 4.11. ■

5 Convergence rates of GEIM in a Hilbert space

In this section, $\mathcal{X}$ is a Hilbert space equipped with its induced norm $\|f\|_{\mathcal{X}} = (f,f)_{\mathcal{X}}$, where $(.,.)_{\mathcal{X}}$ is the scalar product in $\mathcal{X}$. In the same spirit as in the case of a Banach space, we define the sequence $(\tau_n)_{n=1}^{\infty}$ as in formula (15) but now, for any $f \in F$, $P_n(f)$ corresponds to the unique element of X_n that is the orthogonal projection of f onto X_n. Note that lemma 4.1 still holds in the Hilbert setting. We derive convergence rates for the interpolation error by applying the same strategy as in the Banach space case. In section 5.1, we derive convergence rates for $(\tau_n)_{n=1}^{\infty}$ as an intermediate step. We compare to [8] in corollary 3.3 by taking $\gamma_n = \gamma$ in our results. The results of [8] read:

Lemma 5.1 (Corollary $3.3-(ii)$ **of [8])** *If* $d_n \leq C_0 n^{-\alpha}$ *for* $n = 1,2,\dots,$ *then* $\tau_n \leq C_1 n^{-\alpha}$, $n = 1,2,\dots,$ *with* $C_1 = 2^{5\alpha+1}\gamma^{-2}C_0$.

Lemma 5.2 (Corollary $3.3-(iii)$ **of [8])** *If* $d_n \leq C_0 e^{-c_1 n^{\alpha}}$ *for* $n = 1,2,\dots,$ *then* $\tau_n < \sqrt{2C_0}\gamma^{-1}e^{-c_2 n^{\alpha}}$, $n = 1,2,\dots,$ *where* $c_2 = 2^{-1-2\alpha}c_1$.

5.1 Convergence rates for $(\tau_n)_{n=1}^{\infty}$

Like in the Banach space case, we start by bounding the sequence $(\tau_n)_{n=1}^{\infty}$ with respect to $(d_n)_{n=1}^{\infty}$. This is done in theorem 5.3 (the analogue of theorem 4.5). It yields corollaries 5.4 and 5.5, that are the analogue of corollaries 4.6 and 4.7. The major difference with respect to the Banach space case is the absence of a factor $\sqrt{n}$ in corollaries 5.4 and 5.5. It will be the key to obtain improved results in Hilbert spaces.

Theorem 5.3 *For any* $N \geq 0$, *consider a weak greedy algorithm for which* (17) *holds. Then, for any* $K \geq 2,\ 1 \leq m < K$,

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

Proof. See appendix B. ■

Corollary 5.4 *For* $n \geq 2$,

$$\tau_n \leq \sqrt{2}\frac{1}{\prod_{i=1}^{n}\gamma_i^{1/n}}\min_{1\leq m<n} d_m^{\frac{n-m}{n}}. \tag{45}$$

In particular, for any $\ell \geq 1$

$$\tau_{2\ell} \leq \sqrt{2}\frac{1}{\prod_{i=1}^{2\ell}\gamma_i^{\frac{1}{2\ell}}}\sqrt{d_\ell}. \tag{46}$$

Corollary 5.5 *For* $N \geq 0$, $K \geq 2$ *and* $1 \leq m < K$:

$$\tau_{N+K} \leq \frac{1}{\prod_{i=1}^{K}\gamma_{N+i}^{1/K}}\sqrt{2}\tau_{N+1}^{m/K}d_m^{1-m/K}. \tag{47}$$

Proof. The proofs of corollaries 5.4 and 5.5 follow very similar guidelines as the ones for corollaries 4.6 and 4.7. The only difference is that here the staring point is theorem 5.3 instead of 4.5. ■

Using theorem 5.3, we derive decay rates of the sequence $(\tau_n)_{n=1}^{\infty}$ when $(d_n)_{n=1}^{\infty}$ has a polynomial or an exponential decay. In lemmas 5.6 and 5.7, no assumption on the behavior of $(\Lambda_n)_{n=1}^{\infty}$ is made.

Lemma 5.6 *For any* $n \geq 1$, *let* $n = 4\ell + k$ *(where* $\ell \in \{0,1,\dots\}$ *and* $k \in \{0,1,2,3\}$*). If, for* $n \geq 1$, $d_n \leq C_0 n^{-\alpha}$ *with* $C_0 > 0$, *then* $\tau_n \leq C_0\beta_n n^{-\alpha}$, *where*

$$\beta_n = \beta_{4\ell+k} := \begin{cases} 2\left(1+\eta^{-1}\right) & \text{if } n = 1 \\ \sqrt{2\beta_{\ell_1}}\dfrac{1}{\prod_{i=1}^{\ell_2}\gamma_{\ell_1-\lceil\frac{k}{4}\rceil+i}^{\frac{1}{\ell_2}}}(2\sqrt{2})^{\alpha} & \text{if } n \geq 2 \end{cases}$$

and $\ell_1 = 2\ell + \lfloor\frac{2k}{3}\rfloor$, $\ell_2 = 2\left(\ell + \lceil\frac{k}{4}\rceil\right)$.

Proof. The proof is similar with the one of lemma 4.8: the case $n = 1$ directly follows from lemma 4.4 and if $n \geq 2$, we write $n = N + K$ with $N \geq 0$ and $K \geq 2$. Corollary 5.5 yields

$$\tau_{N+K} \leq \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{1/K}} \sqrt{2} \tau_{N+1}^{m/K} d_m^{1-m/K}.$$

By using that $d_m \leq C_0 m^{-\alpha}$ and the recurrence hypothesis $\tau_{N+1} \leq \beta_{N+1}(N+1)^{-\alpha}$, it follows that

$$\tau_{N+K} \leq C_0 \sqrt{2} \frac{1}{\prod_{i=1}^{K} \gamma_{N+i}^{\frac{1}{K}}} \beta_{N+1}^{\frac{m}{K}} \xi(N,K,m)^{\alpha} (N+K)^{-\alpha},$$

where $\xi(N,K,m) = \dfrac{N+K}{m} \left(\dfrac{m}{N+1} \right)^{\frac{m}{K}}$ for any $1 \leq m < K$ and any given index $n = N + K \geq 2$, where $N \geq 0$ and $K \geq 2$. It suffices now to decompose any $n \geq 2$ as $n = 4\ell + k$ with $\ell \in \{0, 1, \dots\}$ and $k \in \{0, 1, 2, 3\}$ and use the same choices of N, K and m described in the proof of lemma 4.8 to derive the result. ■

Lemma 5.7 *If, for $n \geq 1$, $d_n \leq C_0 e^{-c_1 n^{\alpha}}$ with $C_0 \geq 1$, then $\tau_n \leq C_0 \beta_n e^{-c_2 n^{\alpha}}$, where $c_2 := c_1 2^{-2\alpha-1}$ and $\beta_1 = 2(1+\eta^{-1})$, $\beta_n := \sqrt{2} \dfrac{1}{\prod_{i=1}^{2\lfloor \frac{n}{2} \rfloor} \gamma_i^{\frac{1}{2\lfloor \frac{n}{2} \rfloor}}}$ for $n \geq 2$.*

Proof. The proof is the same as lemma 4.9 but uses corollary 5.4 instead of corollary 4.6. ■

As in the case of Banach spaces, it is important to study convergence rates in the case where $(\Lambda_n)_{n=1}^{\infty}$ is monotonically increasing. The following corollary accounts for it.

Corollary 5.8 *If $(\Lambda_n)_{n=1}^{\infty}$ is a monotonically increasing sequence then*

i) if $d_n \leq C_0 n^{-\alpha}$ for any $n \geq 1$, then $\tau_n \leq C_0 \tilde{\beta}_n n^{-\alpha}$, with

$$\tilde{\beta}_n := \begin{cases} 2\left(1+\eta^{-1}\right), & \text{if } n = 1 \\ 2^{3\alpha+1} \gamma_n^{-2}, & \text{if } n \geq 2. \end{cases}$$

ii) if $d_n \leq C_0 e^{-c_1 n^{\alpha}}$ for $n \geq 1$ and $C_0 \geq 1$, then $\tau_n \leq C_0 \tilde{\beta}_n e^{-c_2 n^{-\alpha}}$, with

$$\tilde{\beta}_n := \begin{cases} 2\left(1+\eta^{-1}\right), & \text{if } n = 1 \\ \sqrt{2} \dfrac{1}{\gamma_n}, & \text{if } n \geq 2. \end{cases}$$

Proof. The proof is derived by following the same guidelines as the proof of corollary 4.10. ■ As a direct consequence of corollary 5.8, if γ_n is constant, we recover slightly better results than the ones in [8] for $n \geq 2$ (see lemmas 5.1 and 5.2 above).

5.2 Convergence rates of the interpolation error

Following similar guidelines as in the case of Banach spaces, the following rates can easily be derived for the interpolation error of GEIM.

Theorem 5.9

1. *If, for $n \geq 1$, $d_n \leq C_0 n^{-\alpha}$, with $C_0 > 0$, then $\|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}} \leq (1+\Lambda_n) C_0 \beta_n n^{-\alpha}$ for any $\varphi \in F$, where the parameter β_n is defined as in lemma 5.6.*
2. *If, for $n \geq 1$, $d_n \leq C_0 e^{-c_1 n^{\alpha}}$ with $C_0 \geq 1$, then $\|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}} \leq (1+\Lambda_n) C_0 \beta_n e^{-c_2 n^{\alpha}}$ for any $\varphi \in F$, where β_n and c_2 are defined as in lemma 5.7.*

Corollary 5.10 *If $(\Lambda_n)_{n=1}^{\infty}$ is a monotonically increasing sequence, then:*

- *if $d_n \leq C_0 n^{-\alpha}$ for any $n \geq 1$, then for any $\varphi \in F$,*

$$\|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}} \leq \begin{cases} 2C_0\left(1+\eta^{-1}\right)(1+\Lambda_1), & \text{if } n=1. \\ C_0 2^{3\alpha+1}(1+\Lambda_n)^3\eta^{-2}n^{-\alpha}, & \text{if } n \geq 2. \end{cases}$$

- *if $d_n \leq C_0 e^{-c_1 n^{\alpha}}$ for $n \geq 1$ and $C_0 \geq 1$, then for any $\varphi \in F$,*

$$\|\varphi - \mathcal{J}_n[\varphi]\|_{\mathcal{X}} \leq \begin{cases} 2C_0\left(1+\eta^{-1}\right)(1+\Lambda_1), & \text{if } n=1, \\ C_0\sqrt{2}(1+\Lambda_n)^2\eta^{-1}e^{-c_2 n^{\alpha}}, & \text{if } n \geq 2, \end{cases}$$

where $c_2 = 2^{-2\alpha-1}$.

6 Final remarks

We have analyzed the convergence rates of the interpolation error in GEIM in the case of polynomially or exponentially decaying Kolmogorov n-widths of F. The impact on this convergence rate of the Lebesgue constant appears as multiplicative factors of order $\mathcal{O}(\Lambda_n^2)$ or $\mathcal{O}(\Lambda_n^3)$. Given that, for reasonable enough dictionaries Σ, it has been observed in practical applications that $(\Lambda_n)_{n=1}^{\infty}$ is linear in the worst case scenario (see [1], [9], [13], [11]), our results prove that a decay of order $\mathcal{O}(n^{-3})$ in $d_n(F, \mathcal{X})$ should be enough to ensure the convergence of the interpolation errors of GEIM.

A Proof of Theorem 4.5

We begin by recalling a preliminary lemma for matrices that is proven in [8].

Lemma A.1 *Let $G = (g_{i,j})$ be a $K \times K$ lower triangular matrix with rows $\boldsymbol{g_1}, \ldots, \boldsymbol{g_K}$, W be any m dimensional subspace of $\mathbb{R}^K$, and P be the orthogonal projection of $\mathbb{R}^K$ onto W. Then,*

$$\prod_{i=1}^{K} g_{i,i}^2 \leq \left\{\frac{1}{m}\sum_{i=1}^{K}\|P\boldsymbol{g_i}\|_{\ell_2}^2\right\}^m \left\{\frac{1}{K-m}\sum_{i=1}^{K}\|\boldsymbol{g_i} - P\boldsymbol{g_i}\|_{\ell_2}^2\right\}^{K-m} \tag{48}$$

where $\|.\|_{\ell_2}$ is the euclidean norm of a vector in $\mathbb{R}^K$.

For the proof of theorem 4.5, we consider a lower triangular matrix $A = (a_{i,j})_{i,j=1}^{\infty}$ defined in the following way. For each $j = 1, \ldots,$ we let $\lambda_j \in \mathcal{X}'$ be the bounded linear functional of norm one that satisfies:

$$(i)\ \lambda_j(X_j) = 0, \qquad (ii)\ \lambda_j(\varphi_j) = \operatorname{dist}(\varphi_j, X_j), \tag{49}$$

where $X_j = \operatorname{span}\{\varphi_0, \ldots, \varphi_{j-1}\}$, $j = 1, 2, \ldots,$ is the interpolating space given by the greedy algorithm of GEIM. The existence of such a functional is a consequence of the Hahn-Banach theorem. We let A be the matrix with entries

$$a_{i,j} = \lambda_j(\varphi_i).$$

The matrix A has the following properties:

Q1: The diagonal elements of A satisfy $\gamma_n\tau_n \leq a_{n,n} \leq \tau_n$.

Q2: For every $j < i$, $|a_{i,j}| \leq \operatorname{dist}(\varphi_i, X_j) \leq \tau_j$.

Q3: For every $j > i$, $a_{i,j} = 0$.

Proof.

Q1: We have

$$a_{j,j} = \lambda_j(\varphi_j) = \operatorname{dist}(\varphi_j, X_j) = \|\varphi_j - P_j(\varphi_j)\|_{\mathcal{X}} \leq \max_{\varphi \in F}\|\varphi - P_j(\varphi)\|_{\mathcal{X}} = \tau_j.$$

Lemma 4.1 directly yields the second part of the inequality: $a_{j,j} \geq \gamma_j\tau_j$.

Q2: For any $j < i$ and any $g \in X_j$,

$$|a_{i,j}| = |\lambda_j(\varphi_i)| = |\lambda_j(\varphi_i - g)| \leq \|\lambda_j\|_{\mathcal{X}'} \|\varphi_j - g\|_{\mathcal{X}},$$

where we have used the fact that $\lambda_j(g) = 0$ because $g \in X_j$. Therefore, since $\|\lambda_j\|_{\mathcal{X}'} = 1$, we have $|a_{i,j}| \leq \|\varphi_j - g\|_{\mathcal{X}}, \quad \forall\, g \in X_j$. Thus, $|a_{i,j}| \leq \|\varphi_i - P_j(\varphi_i)\|_{\mathcal{X}} \leq \tau_j$.

Q3: Clearly, for $j > i$, $a_{i,j} = \lambda_j(\varphi_i) = 0$ because $\varphi_i \in X_j$ in this case.

■

We can now prove theorem 4.5: **Proof.** For a given $K \geq 2$, consider the $K \times K$ matrix G formed by the rows and columns of A with indices from $\{N+1, \ldots, N+K\}$. Let Y_m be a subspace of $\mathcal{X}$ of dimension $\leq m$ (we recall that $1 \leq m < K$). For each i, there exists an element $h_i \in Y_m$ such that

$$\|\varphi_i - h_i\|_{\mathcal{X}} = \operatorname{dist}(\varphi_i, Y_m) \leq d_{Y_m},$$

where $d_{Y_m} \coloneqq \max_{\varphi \in F} \operatorname{dist}(\varphi, Y_m)$. Therefore

$$|\lambda_j(\varphi_i) - \lambda_j(h_i)| = |\lambda_j(\varphi_i - h_i)| \leq \|\lambda_j\|_{\mathcal{X}'} \|\varphi_i - h_i\|_{\mathcal{X}} \leq d_{Y_m}. \tag{50}$$

We now consider the vectors $(\lambda_{N+1}(h), \ldots, \lambda_{N+K}(h))\,,\ h \in X_m$. They span a space $W \subset \mathbb{R}^K$ of dimension $\leq m$. We assume that $\dim(W) = m$ (a slight notational adjustment has to be made if $\dim(W) < m$). It follows from (50) that each row $\boldsymbol{g_i}$ of G can be approximated by a vector from W in the ℓ_∞ norm to accuracy d_{Y_m}, and therefore in the ℓ_2 norm to accuracy $\sqrt{K} d_{Y_m}$. Let P be the orthogonal projection of $\mathbb{R}^K$ onto W. Hence, we have

$$\|\boldsymbol{g_i} - P\boldsymbol{g_i}\|_{\ell_2} \leq \sqrt{K} d_{Y_m}, \quad i = 1, \ldots, K. \tag{51}$$

Also, from the property Q2, $\|P\boldsymbol{g_i}\|_{\ell_2} \leq \|\boldsymbol{g_i}\|_{\ell_2} \leq \left(\sum_{j=1}^{i} \tau_{N+j}^2\right)^{1/2}$, and therefore

$$\sum_{i=1}^{K} \|P\boldsymbol{g_i}\|_{\ell_2}^2 \leq \sum_{i=1}^{K} \sum_{j=1}^{i} \tau_{N+j}^2 \leq K \sum_{i=1}^{K} \tau_{N+i}^2. \tag{52}$$

Next, we apply lemma A.1 for this G and W and use property Q1 and estimates (51) and (52) to derive

$$\begin{aligned}
\prod_{i=1}^{K} \gamma_{N+i}^2 \tau_{N+i}^2 &\leq \left\{ \frac{K}{m} \sum_{i=1}^{K} \tau_{N+i}^2 \right\}^m \left\{ \frac{K^2}{K-m} d_{Y_m}^2 \right\}^{K-m} \\
&= K^{K-m} \left(\frac{K}{m} \right)^m \left(\frac{K}{K-m} \right)^{K-m} \left\{ \sum_{i=1}^{K} \tau_{N+i}^2 \right\}^m d_{Y_m}^{2(K-m)} \\
&\leq 2^K K^{K-m} \left\{ \sum_{i=1}^{K} \tau_{N+i}^2 \right\}^m d_{Y_m}^{2(K-m)},
\end{aligned}$$

where we have used the fact that $x^{-x}(1 - x^{x-1}) \leq 2$ for $0 < x < 1$. The proof follows by taking the infimum over all subspaces Y_m of $\mathcal{X}$ of dimension $\leq m$. ■

B Proof of Theorem 5.3

In this section, $\mathcal{X}$ is a Hilbert space. We denote by $(\varphi_n^*)_{n \geq 0}$ the orthonormal system obtained from $(\varphi_n)_{n \geq 0}$ by Gram-Schmidt orthonormalisation. It follows that the orthogonal projector P_n from $\mathcal{X}$ onto X_n can be written as $P_n \varphi = \sum_{i=0}^{n-1} (\varphi, \varphi_i^*)_{\mathcal{X}} \varphi_i^*$, for $n \geq 1$. In particular, $\varphi_n = P_{n+1}\varphi_n = \sum_{j=0}^{n} a_{n,j} \varphi_j^*$, with $a_{n,j} = (\varphi_n, \varphi_j^*)_{\mathcal{X}},\ j \leq n$. There is no loss of generality in assuming that the infinite dimensional Hilbert space $\mathcal{X}$ is $\ell_2(\mathbb{N})$ and that $\varphi_j^* = e_j$, where e_j is the sequence with all entries zero except the j-th entry which is 1. In a similar manner as in the Banach space case, we associate with the greedy procedure of GEIM the lower triangular matrix:

$$A := (a_{i,j})_{i,j=0}^{\infty}, \quad a_{i,j} := 1,\ j > i.$$

The following two properties characterize any lower triangular matrix A generated by such a greedy algorithm.

S1: The diagonal elements of A satisfy $\gamma_n\tau_n \le |a_{n,n}| \le \tau_n$.

S2: For every $m \ge n$, one has $\sum_{j=n}^{m} a_{m,j}^2 \le \tau_n^2$.

Proof.

S1: For any $n \ge 1$, since $\varphi_n - P_n\varphi_n = a_{n,n}\varphi_n^*$, it follows that for any $n \ge 1$, $|a_{n,n}| = \|\varphi_n - P_n\varphi_n\| \le \tau_n$. The fact that $|a_{n,n}| \ge \gamma_n\tau_n$ directly follows from lemma 4.1.

S2: For $m \ge n$, $\sum_{j=n}^{m} a_{m,j}^2 = \|\varphi_m - P_n\varphi_m\|^2 \le \max_{\varphi\in F}\|\varphi - P_n\varphi\|^2 = \tau_n^2$.

■ We can now prove theorem 5.3: **Proof.** For a given $K \ge 2$, consider the $K \times K$ matrix $G = (g_{i,j})$ formed by the rows and columns of A with indices from $\{N+1, \dots, N+K\}$. Each row $\boldsymbol{g_i}$ is the restriction of φ_{N+i} to the coordinates $N+1, \dots, N+K$. Let Y_m be a subspace of $\mathcal{X}$ of dimension $\le m$. Then, $\operatorname{dist}(\varphi_{N+i}, Y_m) \le d_{Y_m}$, $i = 1, \dots, K$. Let $\tilde{W}$ be the linear subspace which is the restriction of Y_m to the coordinates $N+1, \dots, N+K$. In general, $\dim(\tilde{W}) \le m$. Let W be a m dimensional space, $W \subset \operatorname{span}\{e_{N+1}, \dots, e_{N+K}\}$, such that $\tilde{W} \subset W$ and P and $\tilde{P}$ are the projections in $\mathbb{R}^K$ onto W and $\tilde{W}$, respectively. Clearly,

$$\|P\boldsymbol{g_i}\|_{\ell_2} \le \|\boldsymbol{g_i}\|_{\ell_2} \le \tau_{N+1}, \quad i = 1, \dots, K, \tag{53}$$

where we have used property S2 in the last inequality. Note that

$$\|\boldsymbol{g_i} - P\boldsymbol{g_i}\|_{\ell_2} \le \|\boldsymbol{g_i} - \tilde{P}\boldsymbol{g_i}\|_{\ell_2} = \operatorname{dist}\left(\boldsymbol{g_i}, \tilde{W}\right) \le \operatorname{dist}\left(\varphi_{N+i}, Y_m\right) \le d_{Y_m}, \quad i = 1, \dots, K. \tag{54}$$

It follows from property S1 that

$$\prod_{i=1}^{K} |a_{N+i,N+i}|^2 \ge \prod_{i=1}^{K} \gamma_{N+i}^2\tau_{N+i}^2. \tag{55}$$

To derive the result, we apply lemma A.1 for this G and W, use estimates (53), (54) and (55) and take the infimum over all subspaces of $\mathcal{X}$ of dimension $\le m$. ■

References

[1] M. Barrault, Y. Maday, N. Nguyen, and A. Patera, *An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations.*, C. R. Acad. Sci. Paris, Série I., 339 (2004), pp. 667–672.

[2] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, and P. Krysl, *Meshless methods: an overview and recent developments*, Computer Methods in Applied Mechanics and Engineering, 139 (1996), pp. 3–47.

[3] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods.*, SIAM J. Math. Anal., 43 (2011), pp. 1457–1472.

[4] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data Assimilation in Reduced Modeling*, ArXiv e-prints, (2015).

[5] M. D. Buhmann, *Radial basis functions*, Acta Numerica 2000, 9 (2000), pp. 1–38.

[6] A. Cohen and R. DeVore, *Kolmogorov widths under holomorphic mappings*, IMA Journal of Numerical Analysis, (2015), p. dru066.

[7] R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in Banach spaces*, arXiv preprint arXiv:1602.06342, (2016).

[8] R. A. DeVore, G. Petrova, and P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constructive Approximation, (2012), pp. 1–12.

[9] M. Grepl, Y. Maday, N. Nguyen, and A. Patera, *Efficient reduced-basis treatment of non-affine and nonlinear partial differential equations.*, ESAIM, Math. Model. Numer. Anal., 41(3) (2007), pp. 575–605.

[10] Y. Maday and O. Mula, *A Generalized Empirical Interpolation Method: Application of reduced basis techniques to data assimilation*, in Analysis and Numerics of Partial Differential Equations, F. Brezzi, P. Colli Franzone, U. Gianazza, and G. Gilardi, eds., vol. 4 of Springer INdAM Series, Springer Milan, 2013, pp. 221–235.

[11] Y. Maday, O. Mula, A. T. Patera, and M. Yano, *The Generalized Empirical Interpolation Method: Stability theory on Hilbert spaces with an application to the Stokes equation*, Computer Methods in Applied Mechanics and Engineering, 287 (2015), pp. 310–334.

[12] Y. Maday, O. Mula, and G. Turinici, *A priori convergence of the Generalized Empirical Interpolation Method*, in 10th international conference on Sampling Theory and Applications (SampTA 2013), 2013, pp. 168–171.

[13] Y. Maday, N. Nguyen, A. Patera, and G. Pau, *A general multipurpose interpolation procedure: the magic points*, Comm. Pure Appl. Anal., 8(1) (2009), pp. 383–404.

[14] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano, *A Parameterized-Background Data-Weak approach to variational data assimilation: formulation, analysis, and application to acoustics*, International Journal for Numerical Methods in Engineering, 102 (2015), pp. 933–965.

[15] Y. Maday and B. Stamm, *Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces.*, SIAM J. Sci. Comput., 35 (2013).

[16] C. Micchelli and T. Rivlin, *Lectures on optimal recovery*, Springer, 1985.

[17] C. Micchelli, T. Rivlin, and S. Winograd, *The optimal recovery of smooth functions*, Numerische Mathematik, 26 (1976), pp. 191–200.

[18] A. Pinkus, *n-widths in approximation theory*, vol. 7 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], Springer-Verlag, Berlin, 1985.

[19] M. L. Stein, *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media, 2012.