Journal of Complexity 86 (2025) 101887

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

On the approximation of vector-valued functions by volume sampling ☆

Daniel Kressner [a], Tingting Ni [b], André Uschmajew [c,*]

[a] ANCHP, Institute of Mathematics, EPFL, CH-1015 Lausanne, Switzerland
[b] SYCAMORE Lab, EPFL, CH-1015 Lausanne, Switzerland
[c] Institute of Mathematics & Centre for Advanced Analytics and Predictive Sciences, University of Augsburg, 86159 Augsburg, Germany

A R T I C L E   I N F O

Article history:
Received 6 April 2023
Received in revised form 25 June 2024
Accepted 19 July 2024
Available online 31 July 2024

MSC:
41A46
65F55
68W20

Keywords:
Lebesgue-Bochner spaces
Low-rank approximation
Volume sampling
Average widths

A B S T R A C T

Given a Hilbert space $\mathcal{H}$ and a finite measure space Ω, the approximation of a vector-valued function $f : \Omega \to \mathcal{H}$ by a k-dimensional subspace $\mathcal{U} \subset \mathcal{H}$ plays an important role in dimension reduction techniques, such as reduced basis methods for solving parameter-dependent partial differential equations. For functions in the Lebesgue–Bochner space $L^2(\Omega; \mathcal{H})$, the best possible subspace approximation error $d_k^{(2)}$ is characterized by the singular values of f. However, for practical reasons, $\mathcal{U}$ is often restricted to be spanned by point samples of f. We show that this restriction only has a mild impact on the attainable error; there always exist k samples such that the resulting error is not larger than $\sqrt{k+1} \cdot d_k^{(2)}$. Our work extends existing results by Binev et al. (2011) [3] on approximation in supremum norm and by Deshpande et al. (2006) [8] on column subset selection for matrices.

☆ Communicated by T. Ullrich.

* Corresponding author.
E-mail address: andre.uschmajew@uni-a.de (A. Uschmajew).

https://doi.org/10.1016/j.jco.2024.101887

1. Introduction

Let $(\Omega, \mathcal{A}, \mu)$ be a finite measure space with $\mu(\Omega) > 0$. Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. In this work, we consider the approximation of a vector-valued function $f : \Omega \to \mathcal{H}$ by its projection to a finite-dimensional subspace $\mathcal{U} \subset \mathcal{H}$. For $1 \leq p \leq \infty$, possible error measures are then the approximation numbers

$$d_k^{(p)}(f) = \inf \left\{ \| f - g \|_{L^p(\Omega;\mathcal{H})} : \mathcal{U} \text{ is a } k\text{-dimensional subspace of } \mathcal{H} \text{ and } g \in L^p(\Omega; \mathcal{U}) \right\},$$

where $\| \cdot \|_{L^p(\Omega;\mathcal{H})}$ defines the norm on the Lebesgue-Bochner space $L^p(\Omega; \mathcal{H})$.

The number $d_k^{(p)}(f)$ measures the best possible approximation of f in $L^p(\Omega; \mathcal{H})$ by a function g that maps (almost all points in Ω) into some k-dimensional subspace $\mathcal{U} \subseteq \mathcal{H}$. Obviously, for given $\mathcal{U}$ the best choice for g is $g(y) = P_{\mathcal{U}} f(y)$ almost everywhere, where $P_{\mathcal{U}}$ is the $\mathcal{H}$-orthogonal projection onto $\mathcal{U}$. With this notation, we have

$$d_k^{(p)}(f) = \inf \left\{ \| f - P_{\mathcal{U}} f \|_{L^p(\Omega;\mathcal{H})} : \mathcal{U} \text{ is a } k\text{-dimensional subspace of } \mathcal{H} \right\}. \tag{1}$$

In the special case $p = \infty$, the quantity

$$d_k^{(\infty)}(f) = \inf \left\{ \operatorname*{ess\,sup}_{y \in \Omega} \| f(y) - P_{\mathcal{U}} f(y) \| : \mathcal{U} \text{ is a } k\text{-dimensional subspace of } \mathcal{H} \right\}$$

is the Kolmogorov width [14,16] of the "essential" image of f in $\mathcal{H}$. For general $p < \infty$, the quantities $d_k^{(p)}(f)$ are called *average* Kolmogorov widths and can even be generalized to non Hilbert-space settings, see, e.g., [15, Section 2.2] and references therein. Several results on the asymptotic order of $d_k^{(p)}(f)$ are available for the case that $\mathcal{H}$ is a Sobolev space and f is a Gaussian random variable; see [19] for an overview.

In many situations, one hopes or even expects that the above widths decay rapidly as the dimension k of the subspace increases. One then aims at constructing an actual k-dimensional subspace $\mathcal{U}$ that results in an error close to $d_k^{(p)}(f)$. A popular approach is to sample f at k well chosen points $y_1, \ldots, y_k \in \Omega$ and define $\mathcal{U}$ as the span of $f(y_1), \ldots, f(y_k)$. For example, reduced basis methods [11,17] for solving parameter-dependent partial differential equations commonly use a greedy strategy for selecting parameter points $y_1, \ldots, y_k$ successively. Similarly, orthogonal matching pursuit [18] and the empirical interpolation method [2] can be viewed as greedy strategies. Recently, deep neural networks have been demonstrated to be effective at approximating f from (noisy) samples; see, e.g., [1].

Existing convergence analyses [2,4,9,18] of such sample-based methods typically establish error bounds that remain *qualitatively* close to (1), but the involved prefactors can be huge, often growing exponentially with k. For example, the result in [4, Section 2] for $p = \infty$ shows that a greedy selection of k points leads to an error (nearly) bounded by $2^{k+1}(k+1)d_k^{(\infty)}(f)$; this bound has been improved to $2^{k+1}/\sqrt{3} \cdot d_k^{(\infty)}(f)$ in [3, Theorem 4.4]. The results from [3] also show that greedy algorithms recover algebraic or exponential decays of the Kolmogorov widths. On the other hand, the factor 2^k is, in general, unavoidable when using a greedy selection and this raises the question whether the use of samples necessarily leads to large prefactors or whether there exist sample selections leading to more favorable bounds (1). Again for the case $p = \infty$ it was shown in [3, Theorem 4.1] that there always exist sample points $y_1, \ldots, y_k$ for which the prefactor can indeed be reduced to $k+1$. The proof of this result is nonconstructive and uses a maximum volume argument, a classical tool in approximation theory. A maximum volume principle also underlies a result from [10] on the rank-k approximation of matrices in the elementwise maximum norm by cross interpolation. However, there is evidence [8] that such maximum volume arguments do not extend to the case $p = 2$, corresponding to approximation in Frobenius norm for matrices, but randomized sampling arguments can be used instead.

Main result In this note, we treat the case $p = 2$ and establish a prefactor $\sqrt{k+1}$.

Theorem 1. *Let $f \in L^2(\Omega; \mathcal{H})$. Then there exists a measurable set $\hat{\mathbf{\Omega}} \subseteq \Omega \times \cdots \times \Omega = \Omega^k$ of positive product measure $\mu^{\otimes k}$ such that for $\mu^{\otimes k}$-almost every $\mathbf{y} = (y_1, \ldots, y_k) \in \hat{\mathbf{\Omega}}$ it holds that*

$$d_k^{(2)}(f) \le \|f - \Pi_{\mathbf{y}} f\|_{L^2(\Omega;\mathcal{U})} \le \sqrt{k+1} \cdot d_k^{(2)}(f), \tag{2}$$

where $\Pi_{\mathbf{y}} f$ denotes the $\mathcal{H}$-orthogonal projection of f onto the span of $f(y_1), \ldots, f(y_k)$.

Here we do not assume f to be continuous. In general, the elements f of $L^2(\Omega; \mathcal{H})$ are equivalence classes of functions which are only μ-almost everywhere equal. The theorem is formulated in such a way that it applies to the whole equivalence class, in the sense that the set $\hat{\mathbf{\Omega}}$ is the same for every representative. Note that the set of $\mathbf{y} \in \hat{\mathbf{\Omega}}$ that actually satisfy (2) can be different for two representatives, but only by a set of product measure zero. The rest of this note is concerned with the proof of Theorem 1.

Finite-dimensional setting The finite-dimensional analogue of Theorem 1 is due to Deshpande et al. [8] and reads as follows: Given a matrix $A = [a_1, \ldots, a_n] \in \mathbb{R}^{m\times n}$ and $k < m \le n$, there exist k column indices $j_1, \ldots, j_k \in \{1, \ldots, n\}$ such that

$$(\sigma_{k+1}^2 + \cdots + \sigma_m^2)^{1/2} \le \|A - P_{j_1,\ldots,j_k} A\|_F \le \sqrt{k+1} \cdot (\sigma_{k+1}^2 + \cdots + \sigma_m^2)^{1/2}, \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\sigma_1 \ge \cdots \ge \sigma_m \ge 0$ are the singular values of A, and $P_{j_1,\ldots,j_k}$ denotes the orthogonal projection onto $\operatorname{span}\{a_{j_1}, \ldots, a_{j_k}\}$. This result is included in Theorem 1 by considering the uniform measure on $\Omega = \{1, \ldots, n\}$, the Euclidean space $\mathcal{H} = \mathbb{R}^m$ and letting $f(i) = a_i$ for $i = 1, \ldots, n$. The proof in [8] uses a probabilistic method, showing that (3) holds in expectation when columns are sampled with a probability proportional to their induced volume.

An example from [8] shows that the prefactor $\sqrt{k+1}$ in (3) can, in general, not be improved. In follow-up work, Deshpande and Rademacher [7] derived a more efficient algorithm for computing a column subset selection satisfying (3); see [5] for further improvements concerning its numerical realization. For fixed $j_1, \ldots, j_k$, even tiny changes of A may result in a significantly larger error $\|A - P_{j_1,\ldots,j_k} A\|_F$ [6]. This lack of continuity makes it difficult to combine the algorithms from [5,7] with a discretization of the infinite-dimensional setting. For the time being, Theorem 1 is an existence result based on a probabilistic argument and it remains an open problem to design an efficient sampling procedure that leads to an error (approximately) bounded by $\sqrt{k+1} \cdot d_k^{(2)}(f)$.

While our main strategy for the proof of Theorem 1 follows [8], several nontrivial modifications are needed in order to address the infinite-dimensional case.

Bochner integral and measurable functions Before proceeding, let us recall the basic definitions regarding the Bochner integral. A function $f : \Omega \to \mathcal{H}$ is called *strongly μ-measurable* if it is the μ-almost everywhere pointwise strong limit of simple functions $f_\ell : \Omega \to \mathcal{H}$. It is called Bochner integrable if, in addition, $\int_\Omega \|f - f_\ell\| \, d\mu \to 0$ for $\ell \to \infty$. In this case the Bochner integral $\int_\Omega f \, d\mu$ of f is defined as the limit of integrals of the simple functions f_ℓ which then does not depend on the choice of the sequence f_ℓ. Since the real-valued functions $y \mapsto \|f(y) - f_\ell(y)\|$ are themselves strongly μ-measurable, the integrals $\int_\Omega \|f - f_\ell\|$ in this definition are well-defined. In particular, the integrands are $\mathcal{A}$-measurable functions if μ is a complete measure, or otherwise μ-almost everywhere equal to an $\mathcal{A}$-measurable functions [13, Proposition 1.1.16]. Note that a similar reasoning will be implicitly assumed at other occurrences in the paper when composing real-valued functions from strongly μ-measurable $\mathcal{H}$-valued functions without further mentioning.

The Hilbert space $L^2(\Omega; \mathcal{H})$ consists of equivalence classes of Bochner integrable functions f for which $\int_\Omega \|f\|^2 \, d\mu < \infty$. While it is common practice to not distinguish between a function and its equivalence class, we will often work with pointwise arguments of particular representatives. For more details on Bochner integrals and the spaces $L^p(\Omega; \mathcal{H})$ we refer to [13].

2. Schmidt decomposition

The Hilbert space $L^2(\Omega;\mathcal{H})$ is isometrically isomorphic to the space $HS(L^2(\Omega);\mathcal{H})$ of Hilbert–Schmidt operators from $L^2(\Omega)$ to $\mathcal{H}$. This isometry is realized by associating $f \in L^2(\Omega;\mathcal{H})$ with the bounded integral operator

$$T_f : L^2(\Omega) \to \mathcal{H}, \qquad v \mapsto T_f v = \int_\Omega f v \, \mathrm{d}\mu.$$

To see this, let $\{u_i : i \in I\}$ and $\{v_j : j \in J\}$ be orthonormal bases of $\mathcal{H}$ and $L^2(\Omega)$, respectively, with J being countable. Then it can be routinely verified by properties of the Bochner integral that $\{u_i v_j : i \in I,\ j \in J\}$ is an orthonormal basis of $L^2(\Omega;\mathcal{H})$. The integral operator associated with $u_i v_j$ is the rank-one operator $u_i \langle v_j, \cdot \rangle_{L^2(\Omega)}$. These rank-one operators form an orthonormal basis of $HS(L^2(\Omega);\mathcal{H})$; see, e.g., [12, Theorem 4.4.5]. This shows that $f \mapsto T_f$ is an isometric isomorphism between $L^2(\Omega;\mathcal{H})$ and $HS(L^2(\Omega);\mathcal{H})$ and leads to the following well-known decomposition of f.

Theorem 2 *(Schmidt decomposition). Let $f \in L^2(\Omega;\mathcal{H})$. There exist at most countable orthonormal systems $\{u_i \in \mathcal{H} : i = 1,2,\dots,r\}$, $\{v_i \in L^2(\Omega,\mu) : i = 1,2,\dots,r\}$ (with $r \in \mathbb{N} \cup \{+\infty\}$) and singular values $\sigma_1 \geq \sigma_2 \geq \dots$ with $\sigma_i > 0$ for $i = 1,2,\dots,r$ such that*

$$f = \sum_{i=1}^{r} \sigma_i u_i v_i,$$

with the series converging in $L^2(\Omega;\mathcal{H})$. It holds that

$$f(y) \in \overline{\operatorname{span}}\{u_i : i = 1,2,\dots,r\} \quad \mu\text{-a.e.}$$

and

$$f(y) = \sum_{i=1}^{r} \sigma_i u_i v_i(y) \quad \mu\text{-a.e.}, \tag{4}$$

with the series converging in $\mathcal{H}$.

Proof. By the singular value decomposition of $T = T_f$ there exist σ_i, u_i, and v_i with the stated properties such that $T = \sum_{i=1}^{r} \sigma_i u_i \langle v_i, \cdot \rangle_{L^2(\Omega)}$, where the sum converges in Hilbert–Schmidt norm; see, e.g., [12, Theorem 4.3.2]. The isometric isomorphism discussed above then implies the claimed series representation of f in $L^2(\Omega;\mathcal{H})$.

To show the second part of the theorem, we first note that the range of T is the separable Hilbert space $\mathcal{H}_0 = \overline{\operatorname{span}}\{u_i : i = 1,2,\dots,r\}$. As a consequence,

$$\int_A f \, \mathrm{d}\mu = \int_\Omega f \chi_A \, \mathrm{d}\mu = T\chi_A \in \mathcal{H}_0 \quad \forall A \in \mathcal{A}.$$

Since the measure is finite, this implies – by Proposition 1.2.13 in [13] – that $f(y) \in \mathcal{H}_0$ for μ-almost every $y \in \Omega$. Since $\{u_i\}$ is an orthonormal basis of $\mathcal{H}_0$, this in turn implies

$$f(y) = \sum_{i=1}^{r} \langle f(y), u_i \rangle u_i \quad \mu\text{-a.e.} \tag{5}$$

For every i, the function strongly $y \mapsto \langle f(y), u_i \rangle$ is μ-measurable and hence μ-a.e. equal to an $\mathcal{A}$-measurable function [13, Proposition 1.1.16]. Then, for every $A \in \mathcal{A}$,

$$\int_A \langle f(y), u_i \rangle \, d\mu = \left\langle \int_A f(y)\, d\mu, u_i \right\rangle$$
$$= \langle T\chi_A, u_i \rangle = \sum_{j=1}^{r} \sigma_j \langle u_j, u_i \rangle \langle v_j, \chi_A \rangle_{L^2(\Omega)} = \int_A \sigma_i v_i(y)\, d\mu$$

where we have used that the Bochner integral can be interchanged with bounded linear functionals; see, e.g., [13, Eq. (1.2)]. This implies

$$\langle f(y), u_i \rangle = \sigma_i v_i(y) \quad \mu\text{-a.e.},$$

which together with (5) shows (4). □

The singular values are uniquely determined by f and independent of its representative. The number $r \in \mathbb{N} \cup \{+\infty\}$ of positive singular values is called the rank of f, denoted by $\operatorname{rank}(f)$. By the Schmidt–Mirsky theorem [12, Theorem 4.4.7] and the isomorphism explained above, it follows that the truncated function

$$f_k = \sum_{i=1}^{k} \sigma_i u_i v_i = P_{\mathcal{U}_k} f, \quad \mathcal{U}_k = \operatorname{span}\{u_1, \ldots, u_k\},$$

is the best approximation of f by a function of rank at most k in the $L^2(\Omega; \mathcal{H})$-norm. As any other projection of f onto a k-dimensional subspace has rank at most k, it follows

$$d_k^{(2)}(f) = \|f - f_k\|_{L^2(\Omega;\mathcal{H})} = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots)^{1/2}. \tag{6}$$

3. Expected volume

For $k \geq 1$, we consider the product measure space $(\Omega^k, \mathcal{A}^{\otimes k}, \boldsymbol{\mu})$, where $\mathcal{A}^{\otimes k}$ denotes the product σ-algebra (the smallest σ-algebra containing Cartesian products $A_1 \times \cdots \times A_k$ of sets $A_1, \ldots, A_k \in \mathcal{A}$) and $\boldsymbol{\mu} := \mu^{\otimes k}$ denotes the product measure (the unique measure on $\mathcal{A}^{\otimes k}$ satisfying $\boldsymbol{\mu}(A_1 \times \cdots \times A_k) = \mu(A_1) \cdots \mu(A_k)$). Let $f \in L^2(\Omega; \mathcal{H})$. Given k sample points $\mathbf{y} = (y_1, \ldots, y_k)$, the Gramian of $f(y_1), \ldots, f(y_k) \in \mathcal{H}$ is

$$G^{(k)}(\mathbf{y}) = \begin{bmatrix} \langle f(y_1), f(y_1) \rangle & \cdots & \langle f(y_1), f(y_k) \rangle \\ \vdots & & \vdots \\ \langle f(y_k), f(y_1) \rangle & \cdots & \langle f(y_k), f(y_k) \rangle \end{bmatrix} \in \mathbb{R}^{k \times k}.$$

Its determinant

$$\det G^{(k)}(\mathbf{y}) = \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \prod_{i=1}^{k} \langle f(y_i), f(y_{\pi(i)}) \rangle,$$

where $\mathcal{S}_k$ denotes the set of all permutations of $(1, \ldots, k)$, is often called the volume of the $f(y_1), \ldots, f(y_k)$. Note that in regard of the equivalence class of f this quantity is only $\boldsymbol{\mu}$-almost everywhere uniquely defined. We therefore regard $\det G^{(k)}(\mathbf{y})$ as an equivalence class of $\boldsymbol{\mu}$-almost everywhere equal $\boldsymbol{\mu}$-measurable real-valued functions. In turn, the Lebesgue integral of $\det G^{(k)}$, called the expected volume, can be defined. The following lemma shows that this quantity can be computed from the singular values of f, in analogy to [8, Lemma 3.1].

Lemma 3. *With the notation introduced above, it holds that*

$$\int_{\Omega^k} \det G^{(k)} \, \mathrm{d}\boldsymbol{\mu} = \sum_{\substack{j_1, \ldots, j_k = 1 \\ j_1, \ldots, j_k \text{ mutually distinct}}}^{r} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2.$$

Proof. Consider the Schmidt decomposition from Theorem 2. For every finite $n \le r$, let

$$f_n(y) = \sum_{i=1}^{n} \sigma_i u_i v_i(y).$$

Since the v_i are $\mathcal{A}$-measurable, the corresponding determinant

$$\det G_n^{(k)}(\mathbf{y}) = \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \prod_{i=1}^{k} \langle f_n(y_i), f_n(y_{\pi(i)}) \rangle = \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \prod_{i=1}^{k} \sum_{j=1}^{n} \sigma_j^2 v_j(y_i) v_j(y_{\pi(i)})$$

is a nonnegative $\mathcal{A}^{\otimes k}$-measurable function, which satisfies, using Tonelli's theorem,

$$\begin{aligned}
\int_{\Omega^k} \det G_n^{(k)} \, d\boldsymbol{\mu} &= \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \int_{\Omega^k} \prod_{i=1}^{k} \sum_{j=1}^{n} \sigma_j^2 v_j(y_i) v_j(y_{\pi(i)}) \, \mathrm{d}\boldsymbol{\mu} \\
&= \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \int_{\boldsymbol{\Omega}} \prod_{i=1}^{k} v_{j_i}(y_i) v_{j_{\pi(i)}}(y_i) \, \mathrm{d}\boldsymbol{\mu} \\
&= \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \prod_{i=1}^{k} \langle v_{j_i}, v_{j_{\pi(i)}} \rangle_{L^2(\Omega)} \\
&= \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \sum_{\pi \in \mathcal{S}_k} \operatorname{sign}(\pi) \prod_{i=1}^{k} \delta_{j_i, j_{\pi(i)}} \\
&= \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \det(P_{j_1,\dots,j_k}),
\end{aligned}$$

with the matrix $[P_{j_1,\dots,j_k}]_{\alpha,\beta} = \delta_{j_\alpha, j_\beta}$ for $\alpha, \beta = 1, \dots, k$. If all $j_1, \dots, j_d$ are distinct, then $\det(P_{j_1,\dots,j_k}) = 1$, otherwise $\det(P_{j_1,\dots,j_k}) = 0$, since then $P_{j_1,\dots,j_k}$ contains identical rows. Therefore,

$$\int_{\Omega^k} \det G_n^{(k)} \, d\boldsymbol{\mu} = \sum_{\substack{j_1,\dots,j_k=1 \\ j_1,\dots,j_k \text{ mutually distinct}}}^{n} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2.$$

If r is finite, this proves the claim by taking $n = r$. Otherwise we take the limit $n \to \infty$ and argue that

$$\int_{\Omega^k} \det G^{(k)} \, d\boldsymbol{\mu} = \lim_{n \to \infty} \int_{\Omega^k} \det G_n^{(k)} \, d\boldsymbol{\mu}$$

by dominated convergence. This is possible because $\det G_n^{(k)}(\mathbf{y}) \to \det G^{(k)}(\mathbf{y})$ $\boldsymbol{\mu}$-almost everywhere (a consequence of (4)) and

$$\left| \det G_n^{(k)}(\mathbf{y}) \right| \le \sum_{\pi \in \mathcal{S}_k} \left| \prod_{i=1}^{k} \langle f_n(y_i), f_n(y_{\pi(i)}) \rangle \right| \le \sum_{\pi \in \mathcal{S}_k} \prod_{i=1}^{k} \| f_n(y_i) \|^2 \le \sum_{\pi \in \mathcal{S}_k} \prod_{i=1}^{k} \| f(y_i) \|^2.$$

Since $f \in L^2(\Omega; \mathcal{H})$, the right hand side is a dominating integrable function on Ω^k. $\square$

4. Proof of Theorem 1

Let us assume that $\operatorname{rank}(f) \geq k$. Then we can define

$$\varrho = \frac{\det G^{(k)}}{\int_{\Omega^k} \det G^{(k)} \, \mathrm{d}\boldsymbol{\mu}}.$$

By Lemma 3, the denominator is positive. We have $\varrho \geq 0$ ($\boldsymbol{\mu}$-almost everywhere) and $\int_{\Omega^k} \varrho \, \mathrm{d}\boldsymbol{\mu} = 1$. Thus, we can interpret ϱ as a probability density function on the product measure space. We now aim to bound

$$\mathbb{E}_\varrho\big(\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})}\big) = \int_{\Omega^k} \|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \varrho(\mathbf{y}) \, \mathrm{d}\boldsymbol{\mu}(\mathbf{y}),$$

where $\Pi_{\mathbf{y}} f : \Omega \to \mathcal{H}$ is the $\mathcal{H}$-orthogonal projection of f onto $\operatorname{span}\{f(y_1), \dots, f(y_k)\}$. Note that the projector $\Pi_{\mathbf{y}}$ depends on the chosen representative f in the equivalence class; indeed, the subspaces spanned by $f(y_1), \dots, f(y_k)$ can be completely different for two different representatives in the same equivalence class, but only on a set of zero product measure. The statement and proof of the following theorem, which is the analog to [8, Theorem 1.3], pay attention to these subtleties.

Theorem 4. *Given $f \in L^2(\Omega; \mathcal{H})$, consider the Schmidt decomposition (4) and assume $r = \operatorname{rank}(f) \geq k$. Then the following holds:*

(i) *For every fixed $\mathbf{y} \in \Omega^k$ the function $y' \mapsto f(y') - \Pi_{\mathbf{y}} f(y')$, with $\Pi_{\mathbf{y}} f$ defined as above, defines an equivalence class of Bochner integrable functions which belongs to $L^2(\Omega; \mathcal{H})$. The nonnegative function $\mathbf{y} \mapsto \|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})}$ is therefore well-defined.*
(ii) *The class of $\boldsymbol{\mu}$-almost everywhere equal functions $\mathbf{y} \mapsto \|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \cdot \varrho(\mathbf{y})$ contains a representative which is $\mathcal{A}^{\otimes k}$-measurable. Hence $\mathbb{E}_\varrho(\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})})$ is well-defined.*
(iii) *It holds that*

$$\sum_{i=k+1}^{r} \sigma_i^2 \leq \mathbb{E}_\varrho\big(\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})}\big) \leq (k+1) \sum_{i=k+1}^{r} \sigma_i^2.$$

Proof. Ad (i). Consider a fixed representative of f. Then the orthogonal projector $T = \Pi_{\mathbf{y}}$ onto $\operatorname{span}\{f(y_1), \dots, f(y_k)\}$ is a bounded linear operator on $\mathcal{H}$. In turn, Tf is Bochner-integrable. Moreover, $\|f(y') - Tf(y')\| = \|(I - T)f(y')\| \leq \|f(y')\|$ for every $y' \in \Omega$. Therefore, $\int_\Omega \|f(y') - Tf(y')\|^2 \, \mathrm{d}\mu \leq \int_\Omega \|f(y')\|^2 \, \mathrm{d}\mu < \infty$. Obviously, for different choices of representatives f, the functions $f - \Pi_{\mathbf{y}} f$ are $\boldsymbol{\mu}$-almost everywhere equal.

Ad (ii). Take a representative of f that is everywhere a pointwise strong limit of simple functions [13, Proposition 1.1.16]. The corresponding volume function $\mathbf{y} \mapsto \det G^{(k)}(\mathbf{y})$ is then $\mathcal{A}^{\otimes k}$-measurable. Likewise, the function $(\mathbf{y}, y') \mapsto \det G^{(k+1)}(\mathbf{y}, y')$ is $\mathcal{A}^{\otimes (k+1)}$-measurable. By Tonelli's theorem, the nonnegative function

$$g(\mathbf{y}) = \int_\Omega \det G^{(k+1)}(\mathbf{y}, y') \, \mathrm{d}\mu(y')$$

is then $\mathcal{A}^{\otimes k}$-measurable. Since the representative f is fixed, the projector $\Pi_{\mathbf{y}}$ is defined in a meaningful way for every $\mathbf{y} \in \Omega^k$. We claim that for every $\mathbf{y}$ it holds that

$$g(\mathbf{y}) = \det G^{(k)}(\mathbf{y}) \int_\Omega \|f(y') - \Pi_{\mathbf{y}} f(y')\|^2 \, \mathrm{d}\mu(y'). \tag{7}$$

After division by $\int_{\Omega^k} \det G^{(k)}\, \mathrm{d}\boldsymbol{\mu} > 0$, this implies part (ii). To establish (7), we first note that if $f(y_1), \ldots, f(y_k)$ are linearly dependent then (7) trivially holds because it implies $\det G^{(k)}(\mathbf{y}) = 0$ as well as $g(\mathbf{y}) = 0$ (since $\det G^{(k+1)}(\mathbf{y}, y') = 0$ for all $y' \in \Omega$). We may therefore assume that $f(y_1), \ldots, f(y_k)$ are linearly independent. Then $G^{(k)}(\mathbf{y})$ is invertible and for all $y' \in \Omega$ it holds that

$$G^{(k+1)}(\mathbf{y}, y') = \begin{bmatrix} G^{(k)}(\mathbf{y}) & w_{\mathbf{y},y'} \\ w_{\mathbf{y},y'}^T & \|f(y')\|^2 \end{bmatrix}$$
$$= \begin{bmatrix} I & 0 \\ w_{\mathbf{y},y'}^T G^{(k)}(\mathbf{y})^{-1} & 1 \end{bmatrix} \begin{bmatrix} G^{(k)}(\mathbf{y}) & w_{\mathbf{y},y'} \\ 0 & \|f(y')\|^2 - w_{\mathbf{y},y'}^T G^{(k)}(\mathbf{y})^{-1} w_{\mathbf{y},y'} \end{bmatrix},$$

with the vector

$$w_{\mathbf{y},y'} = [\langle f(y_1), f(y')\rangle, \ldots, \langle f(y_k), f(y')\rangle]^T \in \mathbb{R}^k.$$

It is straightforward to show that the orthogonal projection $\Pi_{\mathbf{y}} f(y')$ of $f(y')$ onto the span of $f(y_1), \ldots, f(y_k)$ satisfies $\|\Pi_{\mathbf{y}} f(y')\|^2 = w_{\mathbf{y},y'}^T G^{(k)}(\mathbf{y})^{-1} w_{\mathbf{y},y'}$. Hence we have

$$\det G^{(k+1)}(\mathbf{y}, y') = \det G^{(k)}(\mathbf{y}) \cdot \|f(y') - \Pi_{\mathbf{y}} f(y')\|^2,$$

which yields (7).

Ad (iii). The lower bound follows immediately from (6). It remains to prove the upper bound. Using the same representative for f as in (ii), by (7) we have

$$\mathbb{E}_{\varrho}\big(\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})}\big) \cdot \int_{\Omega^k} \det G^{(k)}\, \mathrm{d}\boldsymbol{\mu} = \int_{\Omega^{k+1}} \det G^{(k+1)}\, \mathrm{d}\mu^{\otimes(k+1)}.$$

Applying Lemma 3 to both sides gives

$$\mathbb{E}_{\varrho}\big(\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})}\big) = \frac{\sum_{j_1,\ldots,j_{k+1}} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \sigma_{j_{k+1}}^2}{\sum_{j_1,\ldots,j_k} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2}, \tag{8}$$

where the summations are performed over mutually distinct indices ranging from 1 to r. As there are $(k+1)!$ different ways of arranging $k+1$ mutually distinct numbers, one deduces that

$$\begin{aligned}\sum_{j_1,\ldots,j_{k+1}} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \sigma_{j_{k+1}}^2 &= (k+1)! \sum_{j_1<\cdots<j_{k+1}} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \sigma_{j_{k+1}}^2 \\ &= (k+1)! \sum_{j=k+1}^{r} \sigma_j^2 \sum_{j_1<\cdots<j_k<j} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \\ &\le (k+1)! \sum_{j=k+1}^{r} \sigma_j^2 \sum_{j_1<\cdots<j_k} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2 \\ &= (k+1) \sum_{j=k+1}^{r} \sigma_j^2 \sum_{j_1,\ldots,j_k} \sigma_{j_1}^2 \cdots \sigma_{j_k}^2.\end{aligned}$$

Inserting this inequality into (8) completes the proof. □

We are ready to state the main result of this section, which in light of (6) also proves Theorem 1.

Theorem 5. *Let $f \in L^2(\Omega; \mathcal{H})$ with rank $r \in \mathbb{N} \cup \{\infty\}$. There exists a measurable set $\hat{\mathbf{\Omega}} \subseteq \Omega^k$ of positive product measure such that for every representative f in the equivalence class and $\boldsymbol{\mu}$-almost every sample tuple $\mathbf{y} \in \hat{\mathbf{\Omega}}$ it holds that*

$$\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \le (k+1) \sum_{i=k+1}^{r} \sigma_i^2.$$

Proof. For the moment, let us assume that $r \ge k$. In this case, the proof boils down to the fact that the probability for a random variable to not being larger than its expected value is positive. For completeness, we provide the full argument. As argued in the proof of Theorem 4(ii) there exists a suitable representative f such that the two functions $\mathbf{y} \mapsto \varrho(\mathbf{y})$ and $\mathbf{y} \mapsto \|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \cdot \varrho(\mathbf{y})$ are both $\mathcal{A}^{\otimes k}$-measurable. Hence, the sets $B = \{\mathbf{y} \in \Omega^k : \varrho(\mathbf{y}) > 0\}$ and

$$\begin{aligned}\hat{\mathbf{\Omega}} &= \left\{ \mathbf{y} \in B : \|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \le (k+1) \sum_{i=k+1}^{r} \sigma_i^2 \right\} \\ &= \left\{ \mathbf{y} \in B : \|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \cdot \varrho(\mathbf{y}) - (k+1) \sum_{i=k+1}^{r} \sigma_i^2 \cdot \varrho(\mathbf{y}) \le 0 \right\}\end{aligned}$$

both belong to $\mathcal{A}^{\otimes k}$. It is then a standard argument to show that we must have $\boldsymbol{\mu}(\hat{\mathbf{\Omega}}) > 0$, since otherwise Theorem 4(iii) would be violated. For any other representative of f, the values of $\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})}$ remain the same for $\boldsymbol{\mu}$-almost all $\mathbf{y}$. This completes the proof in the case $r \ge k$.

In the case $\operatorname{rank}(f) = r < k$ we can apply the theorem to r instead of k. Then

$$\|f - \Pi_{(y_1,\dots,y_r)} f\|^2_{L^2(\Omega;\mathcal{H})} \le (r+1) \sum_{i=r+1}^{r} \sigma_i^2 = 0$$

for all tuples $(y_1, \dots, y_r)$ in a subset $\tilde{\mathbf{\Omega}}$ of Ω^r with positive $\mu^{\otimes r}$-measure. Then for all $\mathbf{y} \in \tilde{\mathbf{\Omega}} \times \Omega^{k-r}$, which has positive $\boldsymbol{\mu}$-measure, we also have

$$\|f - \Pi_{\mathbf{y}} f\|^2_{L^2(\Omega;\mathcal{H})} \le \|f - \Pi_{(y_1,\dots,y_r)} f\|^2_{L^2(\Omega;\mathcal{H})} = 0$$

since the subspaces on which one projects only get larger. □

Acknowledgements

The work of A.U. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 506561557. The authors thank the reviewers for their detailed reading and constructive feedback.

References

[1] B. Adcock, S. Brugiapaglia, N. Dexter, S. Morage, Deep neural networks are effective at learning high-dimensional Hilbert-valued functions from limited data, in: J. Bruna, J. Hesthaven, L. Zdeborova (Eds.), Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference, in: Proceedings of Machine Learning Research, vol. 145, PMLR, 2022, pp. 1–36.
[2] M. Barrault, Y. Maday, N.C. Nguyen, A.T. Patera, An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, C. R. Math. Acad. Sci. Paris 339 (9) (2004) 667–672.
[3] 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.
[4] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme, G. Turinici, A priori convergence of the greedy algorithm for the parametrized reduced basis method, ESAIM: Math. Model. Numer. Anal. 46 (3) (2012) 595–603.
[5] A. Cortinovis, D. Kressner, Low-rank approximation in the Frobenius norm by column and row subset selection, SIAM J. Matrix Anal. Appl. 41 (4) (2020) 1651–1673.

[6] Alice Cortinovis, 2023, Personal communication.
[7] A. Deshpande, L. Rademacher, Efficient volume sampling for row/column subset selection, in: 2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010, 2010, pp. 329–338.
[8] A. Deshpande, L. Rademacher, S. Vempala, G. Wang, Matrix approximation and projective clustering via volume sampling, Theory Comput. 2 (2006) 225–247.
[9] Z. Drmač, S. Gugercin, A new selection operator for the discrete empirical interpolation method—improved a priori error bound and extensions, SIAM J. Sci. Comput. 38 (2) (2016) A631–A648.
[10] S.A. Goreinov, E.E. Tyrtyshnikov, The maximal-volume concept in approximation by low-rank matrices, in: Structured Matrices in Mathematics, Computer Science, and Engineering, I, Boulder, CO, 1999, in: Contemp. Math., vol. 280, Amer. Math. Soc., Providence, RI, 2001, pp. 47–51.
[11] J.S. Hesthaven, G. Rozza, B. Stamm, Certified Reduced Basis Methods for Parametrized Partial Differential Equations, Springer, Cham, 2016.
[12] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, John Wiley & Sons, Ltd., Chichester, 2015.
[13] T. Hytönen, J. van Neerven, M. Veraar, L. Weis, Analysis in Banach Spaces. Vol. I. Martingales and Littlewood-Paley Theory, Springer, Cham, 2016.
[14] A. Kolmogoroff, Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse, Ann. Math. (2) 37 (1) (1936) 107–110.
[15] Y. Malykhin, Widths and rigidity, arXiv:2205.03453, 2022.
[16] A. Pinkus, n-Widths in Approximation Theory, Springer-Verlag, Berlin, 1985.
[17] A. Quarteroni, A. Manzoni, F. Negri, Reduced Basis Methods for Partial Differential Equations, Springer, Cham, 2016.
[18] V. Temlyakov, Greedy Approximation, Cambridge University Press, Cambridge, 2011.
[19] H. Wang, Probabilistic and average linear widths of weighted Sobolev spaces on the ball equipped with a Gaussian measure, J. Approx. Theory 241 (2019) 11–32.