SINGULAR VALUE DECOMPOSITION IN SOBOLEV SPACES: PART I

MAZEN ALI AND ANTHONY NOUY

M. Ali: Ulm University, Institute for Numerical Mathematics

A. Nouy: Centrale Nantes, LMJL (UMR CNRS 6629)

Abstract. A well known result from functional analysis states that any compact operator between Hilbert spaces admits a singular value decomposition (SVD). This decomposition is a powerful tool that is the workhorse of many methods both in mathematics and applied fields. A prominent application in recent years is the approximation of high-dimensional functions in a low-rank format. This is based on the fact that, under certain conditions, a tensor can be identified with a compact operator and SVD applies to the latter. One key assumption for this application is that the tensor product norm is not weaker than the injective norm. This assumption is not fulfilled in Sobolev spaces, which are widely used in the theory and numerics of partial differential equations. Our goal is the analysis of the SVD in Sobolev spaces.

This work consists of two parts. In this manuscript (part I), we address low-rank approximations and minimal subspaces in H^1. We analyze the H^1-error of the SVD performed in the ambient L^2-space. In part II, we will address variants of the SVD in norms stronger than the L^2-norm. We will provide a few numerical examples that support our theoretical findings.

1. Introduction

1.1. **Low-Rank Approximation.** We begin by providing a few motivating examples where the need for low-rank approximation arises. We do not aim to provide a comprehensive overview and refer instead to, e.g., [9, 10, 8, 5, 11] for more details.

Consider a prototypical situation in numerical approximation where a function $f \in C([0,1]^d)$ is approximated by a discrete function on a finite grid

$$f_{k_1,\ldots,k_d} := f(x_{k_1,\ldots,x_d}), \quad x_{k_1,\ldots,x_d} := (k_1 h, \ldots, k_d h) \in [0,1]^d, \quad 0 \le k_j \le n, \tag{1.1}$$

where $h = 1/n$ and $n+1$ is the number of grid points in each dimension. Simply storing such an approximation requires storing at least $(n+1)^d$ values – and this number grows exponentially in d. This quickly becomes unfeasible for problems with a large dimension d.

Low-rank approximation is a widely used tool to address high-dimensional problems. Suppose we can approximate f in the form

$$f \approx f_r := \sum_{k=1}^{r} f_k^1 \otimes \ldots \otimes f_k^d.$$

Storing or evaluating the right-hand-side f_r requires only the knowledge of the rd one-dimensional entries f_k^j and thus reduces the cost to nrd. Thus, this is a significant reduction in cost, *provided r is small* in some sense (hence, the term *low-rank*).

Prominent applications where such high-dimensional problems appear arise in quantum mechanics and quantum chemistry. Wave functions describing physical states of non-relativistic quantum systems are functions in the space $L^2(\mathbb{R}^{3N})$, where N is the number of elementary particles. Thus, for multi-particle systems the dimension $d = 3N$ is large even for comparatively simple models. A wave function $\psi \in L^2(\mathbb{R}^{3N})$ is then given as a solution to a PDE, such as the Schrödinger equation, and ψ is approximated via, e.g., a finite difference method with point values as in (1.1), or through the coefficients of some basis expansion $\psi \approx \sum_{k_1,\ldots,k_d} c_{k_1,\ldots,k_d} \varphi_{k_1,\ldots,k_d}$, where $\varphi_{k_1,\ldots,k_d}$ is some multi-dimensional basis of functions such as finite elements or wavelets. See [9] for more details.

E-mail address: mazen.ali@uni-ulm.de, anthony.nouy@ec-nantes.fr.

2010 *Mathematics Subject Classification.* 46N40 (primary), 65J99 (secondary).

Key words and phrases. Singular Value Decomposition (SVD), Higher-Order Singular Value Decomposition (HOSVD), Low-Rank Approximation, Tensor Intersection Spaces, Sobolev Spaces, Minimal Subspaces .

Yet another application is the approximation of operators or inverse operators for either solving or pre-conditioning equations. Suppose a matrix or an operator are given in the form

$$A = \sum_{k=1}^{d} \mathrm{id}_1 \otimes \ldots \otimes \mathrm{id}_{k-1} \otimes A_k \otimes \mathrm{id}_{k+1} \otimes \ldots \otimes \mathrm{id}_d \,. \tag{1.2}$$

Assume that a function φ applied to A has the integral representation

$$\varphi(A) = \int_\Omega \exp(AF(t))G(t)dt,$$

for some $\Omega \subset \mathbb{R}$ and functions F and G. This is valid for, e.g., the inverse function $\varphi(x) = x^{-\sigma}$ and a self-adjoint positive A. Then, since A has the form as in (1.2), a quadrature rule for approximating the integral

$$\varphi(x) = \int_\Omega \exp(xF(t))G(t)dt, \quad x \in \mathbb{R},$$

provides a separable approximation to $\varphi(A)$. For more details we refer to [10, 8].

1.2. **The Singular Value Decomposition.** Let $T : H_1 \to H_2$ be a continuous compact linear operator between Hilbert spaces. Then, for any $x \in H_1$

$$Tx = \sum_{k=1}^{\infty} \sigma_k \langle x, \psi_k \rangle_{H_1} \phi_k, \tag{1.3}$$

for a non-negative non-increasing sequence $\{\sigma_k\}_{k\in\mathbb{N}}$ and orthonormal systems $\{\psi_k\}_{k\in\mathbb{N}} \subset H_1$ and $\{\phi_k\}_{k\in\mathbb{N}} \subset H_2$. The representation (1.3) is known as the *singular value decomposition* of T, or SVD for short. It is both a powerful analysis tool and an approximation tool. Perhaps the most important feature of this decomposition can be summarized as

$$\left\| T - \sum_{k=1}^{r} \sigma_k \langle \cdot, \psi_k \rangle_{H_1} \phi_k \right\| = \sigma_{r+1} = \inf_{\operatorname{rank}(A) \le r} \|T - A\|,$$

where $\|\cdot\|$ refers to the standard operator norm and where the infimum is taken over all operators A from H_1 to H_2 with rank bounded by r. I.e., (1.3) gives both the optimal approximation with rank $\le r$ (for any r), obtained by truncating the SVD, and the singular values $\{\sigma_k\}_{k\in\mathbb{N}}$ provide the best approximation errors.

SVD has many applications in both mathematics and applied sciences. To name a few: computation of pseudoinverse, determination of rank (and null space, range), least-squares minimization, principal component analysis, proper orthogonal decomposition, data compression, quantum entanglement. For recent applications in model reduction see [12, 1]. The subject of this work is the application of SVD to low-rank approximation of functions, see [5].

A function u in the tensor product $H_1 \otimes H_2$ of two Hilbert spaces H_1 and H_2 possesses a decomposition

$$u = \sum_{k=1}^{\infty} \sigma_k \psi_k \otimes \phi_k, \tag{1.4}$$

if the norm on the tensor product space is not weaker than the injective norm. This guarantees that u can be identified with a compact operator and thus (1.3) applies. These conditions are certainly satisfied for functions between finite dimensional spaces. There are also important examples of infinite dimensional spaces, where this is satisfied as well. The most prominent example is the space of square integrable functions $L^2(\Omega_1 \times \Omega_2)$.

Low-rank approximations are of essential importance when dealing with tensor product spaces $\bigotimes_{j=1}^d H_j$, $d \gg 2$. There is no known generalization of SVD to $d > 2$. However, if we consider the vector space isomorphism (given appropriate norms), $\bigotimes_{j=1}^d H_j \cong \left(\bigotimes_{j\in\alpha} H_j\right) \otimes \left(\bigotimes_{j\in\alpha^c} H_j\right)$, $\alpha \subset \{1, \ldots, d\}$, we can apply SVD in the latter tensor space since this is again a two dimensional tensor product. This is known as the *higher-order singular value decomposition* (HOSVD), see [4, 7]. Thus, the theory for $d = 2$ can be recycled for higher dimensions. This applies to high-dimensional kernel operators in $L^2(\bigtimes_{j=1}^d \Omega_j)$.

There are two other works[1] that considered the related questions of regularity and error estimation of the L^2-SVD. In [14] the author showed that the L^2-SVD inherits the regularity of the original function. In [6] the author investigated L^∞-error control of the L^2-SVD for functions with sufficient smoothness by using the Gagliardo-Nirenberg inequality.

An important example where (1.4) does not apply are multi-dimensional Sobolev spaces. The Sobolev norm on the tensor product space is not weaker than the injective norm and thus Sobolev functions can not be identified with compact operators. Another way of framing this from an approximation standpoint: we can not apply SVD to functions while controlling the Sobolev norm. However, not all hope is lost, since Sobolev spaces are "in between" spaces where SVD applies. E.g., the space of square integrable functions or the space of functions with mixed smoothness. Moreover, Sobolev spaces such as $H^1(\Omega)$ can be identified with an intersection of tensor product spaces, where SVD applies in each of the spaces in the intersection.

The purpose of this work is to analyze if and how SVD can be applied to approximate functions in a Sobolev space. We work with the prototype $H^1(\Omega)$, which frequently arises as the solution space of partial differential equations. The results can be naturally extended to the spaces $H^k(\Omega)$, $k > 1$. The paper is organized as follows. In Section 2 we briefly review some of the basics of tensor spaces. In Section 3 we discuss low-rank approximation and minimal subspaces in Sobolev spaces. In Section 4 we analyze the H^1-error of the SVD performed in the ambient L^2 space (L^2-SVD).

2. Preliminaries

We briefly review some of the theory on tensor spaces and minimal subspaces. Most of the following material can be found in [5], some of it in [3]. We use the notation $A \lesssim B \Leftrightarrow A \leq CB$, for some constant $C > 0$ independent of A or B. Similarly for $\gtrsim$; and $\sim$ if both $\lesssim$ and $\gtrsim$ hold. We use $\cong$ to denote vector space isomorphisms, with equivalent norms where relevant.

2.1. Algebraic Tensor Spaces.

Let $V = X \otimes_a Y$ be an algebraic tensor product space, where X and Y are vector spaces. Briefly, it is the space of all sums of the form $v = \sum_{k=1}^r x \otimes y$, $x \in X,\ y \in Y,\ r \in \mathbb{N}$, where the tensor product $\otimes$ is bilinear on $X \times Y$. See [5, Chapter 3.2] for a precise definition of the tensor product.

This construction can be extended for more than two vector spaces to obtain the tensor space $V = {}_a\bigotimes_{j=1}^d X_j$, with elements $v = \sum_{k=1}^r \bigotimes_{j=1}^d x_j,\ x_j \in X_j,\ r \in \mathbb{N}$. We will sometimes require the isomorphic representations

$$ {}_a\bigotimes_{j=1}^{d} X_j \cong X_i \otimes_a \left({}_a\bigotimes_{j \neq i} X_j \right) \cong \left({}_a\bigotimes_{j \in \alpha} X_j \right) \otimes_a \left({}_a\bigotimes_{i \in \alpha^c} X_i \right), $$

where $\alpha \subset \{1, \dots, d\},\ \alpha^c = \{1, \dots, d\} \setminus \alpha$.

2.2. Tensor Norms and Banach Tensor Spaces.

If we are given a norm $\|\cdot\|$ on the vector space $V = {}_a\bigotimes_{j=1}^d X_j$, we can consider the completion w.r.t. that norm.

Definition 2.1 (Topological Tensor Product)**.** *The space*

$$ {}_{\|\cdot\|}\bigotimes_{j=1}^{d} X_j := \overline{{}_a\bigotimes_{j=1}^{d} X_j}^{\|\cdot\|}, $$

is called a topological tensor product.

Let each of the X_j be a normed vector space. Since $\|\cdot\|$ induces a topology on V and with the product topology on $\bigtimes_{j=1}^d X_j$, we can ask if $\otimes : \bigtimes_{j=1}^d X_i \to V$ is continuous. In fact, many useful properties in the analysis of tensor product spaces require even stronger conditions. For ease of presentation, we list the definitions for $d = 2$.

Definition 2.2 (Crossnorms)**.** *A norm on $V = X \otimes_a Y$ is called a crossnorm if $\|x \otimes y\| = \|x\|_X \|y\|_Y$. It is called a reasonable crossnorm if it is a crossnorm and $\|x^* \otimes y^*\|^* = \|x^*\|_{X^*} \|y^*\|_{Y^*}$, $x^* \in X^*,\ y^* \in Y^*$ where $\|\cdot\|^*$ denotes the standard dual norm on the topological dual Z^* of a space Z.*

[1] That we are aware of.

It is called a uniform crossnorm if it is a reasonable crossnorm and $\|A \otimes B\| = \|A\|\|B\|$, $A \in \mathcal{L}(X,X)$, $B \in \mathcal{L}(Y,Y)$, *with the standard operator norms and where* $\mathcal{L}(X,Y)$ *denotes the space of continuous linear operators from* X *to* Y.

There are two important examples of reasonable crossnorms which are the strongest and the weakest crossnorms (see [3, Chapter 1.1.2] for a justification of the terminology).

Definition 2.3 (Projective and Injective Norms)**.** *The projective norm on* $V = X \otimes_a Y$ *is defined as*

$$\|v\|_\wedge := \inf\left\{\sum_{i=1}^m \|x_i\|_X \|y_i\|_Y : v = \sum_{i=1}^m x_i \otimes y_i\right\},$$

where the infimum is taken over all possible representations of v. *The injective norm on* $V = X \otimes_a Y$ *is defined as*

$$\|v\|_\vee := \sup_{\varphi \in X^*\setminus\{0\},\, \psi \in Y^*\setminus\{0\}} \frac{|(\varphi \otimes \psi)v|}{\|\varphi\|_{X^*}\|\psi\|_{Y^*}}.$$

By [5, Proposition 4.68], we have that for any reasonable crossnorm $\|\cdot\|$, $\|\cdot\|_\vee \lesssim \|\cdot\| \lesssim \|\cdot\|_\wedge$. In this work we will frequently require the following definition.

Definition 2.4 (Hilbert Tensor Space with Canonical Norm)**.** *Let* $H = H_1 \otimes_a H_2$ *be an algebraic tensor product of two Hilbert spaces* H_1 *and* H_2. *The canonical inner product (and associated canonical norm) on* H *is defined such that* $\langle x_1 \otimes y_1, x_2 \otimes y_2\rangle = \langle x_1, x_2\rangle_{H_1} \cdot \langle y_1, y_2\rangle_{H_2}$. *By linearity this definition extends to any* $v \in H$. *The canonical norm is a uniform crossnorm.*

2.3. **Sobolev Spaces.** For the remainder of this work we will require the spaces

$$L^2(\Omega) = L^2(\bigtimes_{j=1}^d \Omega_j), \quad H^1(\Omega) = H^1(\bigtimes_{j=1}^d \Omega_j).$$

We use the shorthand notation $\|\cdot\|_0$ to denote the L^2 norm and $\|\cdot\|_1$ to denote the H^1 norm. This notation will be used both for the tensor product space and the one dimensional components, where the difference should be clear from context. We use $H^1_{\mathrm{mix}}(\Omega)$ to denote spaces of functions with mixed smoothness with the corresponding norm $\|\cdot\|_{\mathrm{mix}}$.

We have

$$L^2(\Omega) \cong {}_{\|\cdot\|_0}\bigotimes_{j=1}^d L^2(\Omega_j), \quad H^1_{\mathrm{mix}}(\Omega) \cong {}_{\|\cdot\|_{\mathrm{mix}}}\bigotimes_{j=1}^d H^1(\Omega_j),$$

where $\|\cdot\|_0$ (resp. $\|\cdot\|_{\mathrm{mix}}$) are uniform crossnorms defined from the norms $\|\cdot\|_0$ (resp. $\|\cdot\|_1$) on the individual spaces $L^2(\Omega_j)$ (resp. $H^1(\Omega_j)$).

We frequently require spaces of functions differentiable in only one direction

$$H^{e_k} := H^1(\Omega_k) \otimes_{\|\cdot\|_{e_k}} \left({}_a\bigotimes_{j\neq k} L^2(\Omega_j)\right),$$

where for $e_k = (\delta_{1k}, \ldots, \delta_{dk})$ being the k-th canonical vector, the norm is defined via

$$\|v\|_{e_k}^2 := \|v\|_0^2 + \left\|\frac{\partial}{\partial x_k} v\right\|_0^2.$$

As in Definition 2.3, we can define the projective and injective norms on $H^1(\Omega_k)_a \otimes \left({}_a\bigotimes_{j\neq k} L^2(\Omega_j)\right)$. We denote these norms by $\|\cdot\|_{\wedge(e_k)}$ and $\|\cdot\|_{\vee(e_k)}$, respectively. The space $H^1(\Omega)$ can be identified with the intersection space

$$H^1(\Omega) \cong \bigcap_{k=1}^d H^{e_k}, \tag{2.1}$$

where the latter is equipped with the intersection norm $\|\cdot\| := \max_{1\le k\le d} \|\cdot\|_{e_k}$, or any equivalent norm. The utility in this representation lies in the fact that $\|\cdot\|_{e_k}$ is the canonical norm on the Hilbert tensor space H^{e_k} and thus SVD applies (see Section 2.5). For each $1 \le k \le d$, we get a different decomposition.

2.4. **Minimal Subspaces and Tensor Formats.** For a tensor in the algebraic tensor space $X \otimes_a Y$, with X and Y Hilbert spaces, the SVD gives the representation $v = \sum_{k=1}^r \sigma_k \psi_k \otimes \phi_k$. Letting $U_1 := \operatorname{span}\{\psi_k : 1 \le k \le r\}$, $U_2 := \operatorname{span}\{\phi_k : 1 \le k \le r\}$, we have the obvious statement $u \in U_1 \otimes_a U_2$. More importantly, these spaces are minimal in the sense that if $u \in V_1 \otimes_a V_2$, then $U_1 \subset V_1$ and $U_2 \subset V_2$. Spaces U_1 and U_2 are called the *minimal subspaces* of u and they can be defined in a more general setting.

Definition 2.5 (Minimal Subspaces). *Let $\|\cdot\| \gtrsim \|\cdot\|_\vee$ be a norm on $V = {}_a\bigotimes_{j=1}^d X_j$. For any $v \in \overline{V}^{\|\cdot\|}$ the j-th minimal subspace is defined as*

$$U_j^{\min}(v) := \overline{\operatorname{span}\left\{\varphi(v) : \varphi = \bigotimes_{k=1}^d \varphi_k,\ \varphi_j = \mathrm{id}_j,\ \varphi_k \in (X_k)^*,\ k \neq j)\right\}}^{\|\cdot\|_{X_j}},$$

where id_j denotes the identity operator on X_j. This definition can be naturally extended to $U_\alpha^{\min}(v)$ for any $\alpha \subset \{1, \dots, d\}$.

The question whether

$$v \in \overline{{}_a\bigotimes_{j=1}^d U_j^{\min}(v)}^{\|\cdot\|} \tag{2.2}$$

is not trivial for topological tensors $v \in \overline{V}^{\|\cdot\|}$. A positive answer requires further structure of the component spaces and the tensor norm.

Definition 2.6 (Grassmanian). *Let X be a Banach space. A closed subspace $U \subset X$ is called direct or complemented if there exists a closed subspace W such that $X = U \oplus W$ is a direct sum. The set $\mathbb{G}(X)$ of all complemented subspaces in X is called the Grassmanian.*

Any closed subspace U of a Hilbert space X belongs to $\mathbb{G}(X)$. An important example where (2.2) is satisfied is when all X_j are Hilbert spaces and $\|\cdot\|$ is the canonical norm. The Sobolev space $H^1(\Omega)$ does not have this property. In particular, $\|\cdot\| \gtrsim \|\cdot\|_\vee$ does not hold. However, $H^1(\Omega)$ is isomorphic to an intersection of tensor spaces, where each individual space in the intersection satisfies (2.2). This property is frequently exploited in our work.

Ultimately we are interested in low-rank approximations. For $d = 2$, there is only one choice of a low-rank format. However, for $d > 2$ there are many possible low-rank tensor formats. The two most basic tensor formats are the following.

Definition 2.7 (Canonical Format). *Let $r \in \mathbb{N}$. The r-term (or canonical) format in $V = {}_a\bigotimes_{j=1}^d X_j$ is defined as*

$$\mathcal{R}_r(V) := \left\{ v = \sum_{k=1}^r \bigotimes_{j=1}^d x_j^k : x_j^k \in X_j \right\}.$$

Definition 2.8 (Tucker Format). *For $\boldsymbol{r} = (r_1, \dots, r_d) \in \mathbb{N}^d$, the Tucker format in $V = {}_a\bigotimes_{j=1}^d X_j$ is defined as*

$$\mathcal{T}_{\boldsymbol{r}}(V) := \left\{ v \in V : \dim U_j^{\min}(v) \le r_j \right\}.$$

2.5. **Tensors as Operators and Singular Value Decomposition.** Let $\mathcal{F}(Y, X)$ denote the space of finite rank operators from Y to X, $\mathcal{K}(Y, X)$ denote the space of compact operators from Y to X and $\mathcal{N}(Y, X)$ denote the space of nuclear operators from Y to X. Then, for any reasonable crossnorm $\|\cdot\|$ we get the inclusions (see [5, Corollary 4.84])

$$\mathcal{N}(Y, X) \cong X \otimes_{\|\cdot\|_\wedge} Y^* \subset X \otimes_{\|\cdot\|} Y^* \subset X \otimes_{\|\cdot\|_\vee} Y^* \cong \overline{\mathcal{F}(Y, X)}^{\|\cdot\|_{X \leftarrow Y}} \subset \mathcal{K}(Y, X).$$

An important example and the subject of this work is the case when X and Y are Hilbert spaces. Then, $X^* \cong X$ and $Y^* \cong Y$. This implies that if $\|\cdot\|$ is a reasonable crossnorm, then $X \otimes_{\|\cdot\|} Y \subset \mathcal{K}(Y, X)$. Since we can apply the singular value decomposition in $\mathcal{K}(Y, X)$, this gives a representation for any

$v \in X \otimes_{\|\cdot\|} Y$, $v = \sum_{k=1}^{\infty} \sigma_k \psi_k \otimes \phi_k$, for a decreasing non-negative sequence $\{\sigma_k\}_{k\in\mathbb{N}}$ and orthonormal systems $\{\psi_k\}_{k\in\mathbb{N}} \subset X$, $\{\phi_k\}_{k\in\mathbb{N}} \subset Y$. Moreover, this provides us with the best low-rank approximations

$$\|v - \sum_{k=1}^{r} \sigma_k x_k \otimes y_k\|_{\vee} = \sigma_{r+1} = \inf_{v_r = \sum_{k=1}^{r} x_k \otimes y_k} \|v - v_r\|_{\vee},$$

for any $r \in \mathbb{N}$. The rank of v is the smallest r such that $\sigma_{r+1} = 0$ ($r = \infty$ if no such r exists). For the canonical norm on Hilbert tensor spaces we get

$$\|v - \sum_{k=1}^{r} \sigma_k x_k \otimes y_k\|^2 = \sum_{k=r+1}^{\infty} (\sigma_k)^2 = \inf_{v_r = \sum_{k=1}^{r} x_k \otimes y_k} \|v - v_r\|^2.$$

The corresponding space is the space of Hilbert Schmidt operators from Y to X, $X \otimes_{\|\cdot\|} Y \cong \mathrm{HS}(Y, X)$. A typical example is the space of square integrable functions $L^2(\Omega) \cong L^2(\Omega_1) \otimes_{\|\cdot\|_0} L^2(\Omega_2)$, where we consider product domains $\Omega = \Omega_1 \times \Omega_2$. The Sobolev space $H^1(\Omega)$, on the other hand, is not equipped with the canonical norm. The Hilbert tensor space that is the tensor product of one dimensional Sobolev spaces with the canonical norm corresponds to the space $H^1_{\mathrm{mix}}(\Omega)$ of functions with mixed smoothness. The singular value decomposition does not apply in H^1 directly, which is the motivation for this work.

The above does not extend to $d > 2$ directly. However, we have the following vector space isomorphism.

Definition 2.9 (Matricisation). *The matricisation $\mathcal{M}_\alpha$ with $\emptyset \neq \alpha \subsetneq \{1, \dots, d\}$ is the linear map defined by*

$$\begin{aligned} \mathcal{M}_\alpha : {}_a\bigotimes_{j=1}^{d} X_j &\to \left({}_a\bigotimes_{j\in\alpha} X_j\right) \otimes_a \left({}_a\bigotimes_{j\in\alpha^c} X_j\right), \\ \bigotimes_{j=1}^{d} x_j &\mapsto \left(\bigotimes_{j\in\alpha} x_j\right) \otimes \left(\bigotimes_{j\in\alpha^c} x_j\right), \end{aligned}$$

where the definition can be extended to any $x \in {}_a\bigotimes_{j=1}^{d} X_j$ by linearity. Moreover, this definition can be extended to topological tensors, if the norms in the domain and image of $\mathcal{M}_\alpha$ are compatible, i.e., $\mathcal{M}_\alpha$ and $\mathcal{M}_\alpha^{-1}$ are continuous.

Definition 2.10 (HOSVD). *Let α be as in* (2.9) *and let $V = {}_{\|\cdot\|}\bigotimes_{j=1}^{d} X_j$, where all X_j are Hilbert spaces and $\|\cdot\|$ is the canonical norm. Then, $\mathcal{M}_\alpha$ is an linear isometric isomorphism from V to the Hilbert tensor space $\left({}_a\bigotimes_{j\in\alpha} X_j\right) \otimes_{\|\cdot\|_\alpha} \left({}_a\bigotimes_{i\in\alpha^c} X_i\right)$, endowed with the canonical norm $\|\cdot\|_\alpha$. Thus, we can apply SVD for any α. Set $\alpha = \{j\}$ and let $\{\psi_k^j\}_{k\in\mathbb{N}} \subset X_j$ denote the X_j-orthonormal singular functions obtained from the SVD of $\mathcal{M}_{\{j\}}(x)$. Then, there exists a unique sequence $\boldsymbol{x} \in \ell_2(\mathbb{N}^d)$ such that $x = \sum_{k_1,\dots,k_d=1}^{\infty} \boldsymbol{x}_{k_1,\dots,k_d} \psi_{k_1}^1 \otimes \cdots \otimes \psi_{k_d}^d$. This representation is called the higher-order singular value decomposition (HOSVD) of x.*

Other types of decompositions can be obtained by considering SVDs of α-matricisations for all α in a dimension partition tree over $\{1, \cdots, d\}$. These decompositions are called hierarchical HOSVDs. For details and precise definitions see [5, Sections 8.3, 11.3].

The approximation obtained by truncating the HOSVD is not optimal anymore but rather quasi-optimal, as recalled in the following theorem. The proof can be found in [5, Theorem 10.3].

Theorem 2.11 (HOSVD truncation). *In the setting of Definition 2.10, let $\boldsymbol{r} := (r_1, \dots, r_d) \in \mathbb{N}^d$ and let $P_{r_j}^j$ be the orthogonal projection from X_j onto $U_{r_j}^j(x) = \operatorname{span}\left\{\psi_k^j : 1 \le k \le r_j\right\}$, where $\psi_k^j \in X_j$ are the singular functions obtained via the HOSVD. Then $x_{\boldsymbol{r}} := \mathcal{P}_{\boldsymbol{r}} x$, with $\mathcal{P}_{\boldsymbol{r}} = \bigotimes_{j=1}^{d} P_{r_j}^j x$, is called the truncated HOSVD with multilinear (Tucker) rank $\boldsymbol{r}$, and the truncation error satisfies*

$$\|x - x_{\boldsymbol{r}}\|^2 \le \sum_{j=1}^{d} \sum_{i=r_j+1}^{\infty} (\sigma_i^j)^2 \le d \inf_{v \in \mathcal{T}_{\boldsymbol{r}}(V)} \|x - v\|^2,$$

where $\{\sigma_i^j\}_{i\in\mathbb{N}}$ are the singular values of $\mathcal{M}_{\{j\}}(x)$.

Similar statements can be obtained for the hierarchical HOSVD, with different constants [4, 7].

3. Low-Rank Approximations in H^1

Before we continue with our analysis of low-rank approximations, we clarify what is meant by an algebraic tensor in $H^1(\Omega)$. So far we defined algebraic tensors only on tensor product spaces. In the case of intersection spaces there are several candidates, since there are multiple tensor product spaces involved. As the following lemma shows, all possible choices lead to the same algebraic tensor space, as long as we require H^1 regularity.

Lemma 3.1 ([5, Proposition 4.104])**.**

$$_a\bigotimes_{j=1}^d L^2(\Omega_j) \bigcap H^1(\Omega) = {}_a\bigotimes_{j=1}^d H^1(\Omega_j).$$

Proof. To show how the algebraic tensors in $L^2(\Omega)$ inherit H^1 regularity, we detail the proof in a more rigorous way than in [5, Proposition 4.104]. The inclusion "$\supset$" is obvious. For the inclusion "$\subset$", let $u \in {}_a\bigotimes_{j=1}^d L^2(\Omega_j) \bigcap H^1(\Omega)$. For a fixed $1 \le k \le d$, we have

$$u \in L^2(\Omega_k) \otimes_a \left({}_a\bigotimes_{j \neq k} L^2(\Omega_j)\right) \cap \overline{H^1(\Omega_k)_a \bigotimes_{j \neq k} L^2(\Omega_j)}^{\|\cdot\|_{e_k}}. \tag{3.1}$$

Then, there is a number $r \in \mathbb{N}$ and functions $\{v_l\}_{l=1}^r \subset L^2(\Omega_k)$, $\{w_l\}_{l=1}^r \subset {}_a\bigotimes_{j \neq k} L^2(\Omega_j)$ such that $u = \sum_{l=1}^r v_l \otimes w_l$. By [5, Lemma 3.13], w.l.o.g., we can assume $\{v_l\}_{l=1}^r$ and $\{w_l\}_{l=1}^r$ to be linearly independent. Thus, we can choose a dual basis $\{\varphi_l\}_{l=1}^r \subset \left({}_a\bigotimes_{j \neq k} L^2(\Omega_j)\right)^*$ such that $\varphi_k(w_l) = \delta_{kl}$. Note that the mapping

$$\mathrm{id}_k \otimes \varphi_l : H^1(\Omega_k) \otimes_{\|\cdot\|_{e_k}} \left({}_a\bigotimes_{j \neq k} L^2(\Omega_j)\right) \to H^1(\Omega_k)$$

is continuous for all $1 \le l \le r$, since $\|\cdot\|_{e_k}$ is a reasonable crossnorm.

Moreover, by (3.1), there exist $\{v_l^m \otimes w_l^m\}_{\substack{1 \le l \le m, \\ 1 \le m < \infty}} \subset H^1(\Omega_k) \otimes_a \left({}_a\bigotimes_{j \neq k} L^2(\Omega_j)\right)$ such that

$$\lim_{m \to \infty} \left\| u - \sum_{l=1}^m v_l^m \otimes w_l^m \right\|_{e_k} = 0.$$

Thus, since $\mathrm{id}_k \otimes \varphi_i$ is continuous

$$\begin{aligned} 0 &= \lim_{m \to \infty} \left\| \mathrm{id}_k \otimes \varphi_i \left(u - \sum_{l=1}^m v_l^m \otimes w_l^m \right) \right\|_1 \\ &= \lim_{m \to \infty} \left\| \sum_{l=1}^r v_l \varphi_i(w_l) - \sum_{l=1}^m v_l^m \varphi_i(w_l^m) \right\|_1 \\ &= \lim_{m \to \infty} \left\| v_i - \sum_{l=1}^m v_l^m \varphi_i(w_l^m) \right\|_1 . \end{aligned}$$

And thus $v_i \in \overline{\operatorname{span}\left\{v_l^m : 1 \le l \le m,\ 1 \le m < \infty\right\}}^{\|\cdot\|_1} \subset H^1(\Omega_k)$. Since i and k were chosen arbitrarily, this shows $u \in \bigcap_{k=1}^d H^1(\Omega_k) \otimes_a \left({}_a\bigotimes_{j \neq k} L^2(\Omega_j)\right)$. Finally, by [5, Lemma 6.11] $\bigcap_{k=1}^d H^1(\Omega_k) \otimes_a \left({}_a\bigotimes_{j \neq k} L^2(\Omega_j)\right) = {}_a\bigotimes_{j=1}^d H^1(\Omega_j)$. This completes the proof. □

3.1. Existence of Low-Rank Approximations.

First, we address the question of existence of low-rank approximations of a function $u \in H^1(\Omega)$. Since for $d > 2$ and $r > 1$ the set $\mathcal{R}_r(V)$ is not closed even for the case $V = L^2(\Omega)$ (see [5, Section 9.4.1]), we only consider Tucker formats.

In analogy to Definition 2.5, for $u \in H^1(\Omega)$ we define the subspace

$$U^j(u) := \operatorname{span}\left\{\varphi(u) : \varphi = \bigotimes_{k=1}^{d} \varphi_k,\ \varphi_j = \mathrm{id}_j,\ \varphi_k \in \left(L^2(\Omega_k)\right)^*,\ k \neq j\right\}. \tag{3.2}$$

Note that for $u \in {}_a\bigotimes_{j=1}^d H^1(\Omega_j)$, $U^j(u)$ is a closed subspace of $H^1(\Omega_j)$ and the definition coincides with the case $u \in {}_a\bigotimes_{j=1}^d L^2(\Omega_j)$. Since in this case u can be written as $u = \sum_{k=1}^r \bigotimes_{i=1}^d v_k^i$, for some $r \in \mathbb{N}$, any φ from (3.2) applied to u yields $\varphi(u) = \sum_{k=1}^r v_k^j \left(\prod_{i\neq j} \varphi_i(v_k^i)\right)$. And thus $U^j(u) \subset \operatorname{span}\left\{v_k^j : 1 \leq k \leq r\right\}$. The subspace $U^j(u) \subset H^1(\Omega_j)$ is finite dimensional and is closed in any norm.

Thus, together with Lemma 3.1, we can define the Tucker manifold for $\boldsymbol{r} = (r_1, \ldots, r_d)$ and $V = {}_a\bigotimes_{j=1}^d H^1(\Omega_j)$ in the same way as in Definition 2.8. This set remains weakly closed in $H^1(\Omega)$. To show this, we first require the following lemma.

Lemma 3.2. $\mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d H^1(\Omega_j)\right) = {}_a\bigotimes_{j=1}^d H^1(\Omega_j) \cap \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d L^2(\Omega_j)\right)$.

Proof. The inclusion "$\subset$" is trivial.

For the other inclusion, assume $v \in {}_a\bigotimes_{j=1}^d H^1(\Omega_j) \cap \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d L^2(\Omega_j)\right)$. Since $v \in {}_a\bigotimes_{j=1}^d H^1(\Omega_j)$, $U^j(v) \subset H^1(\Omega_j)$ and by [5, Lemma 6.11] $v \in {}_a\bigotimes_{j=1}^d U^j(v) \subset {}_a\bigotimes_{j=1}^d L^2(\Omega_j)$. In particular, since $v \in \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d L^2(\Omega_j)\right)$, $\dim U^j(v) \leq r_j$ for all $1 \leq j \leq d$. Hence, $v \in \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d H^1(\Omega_j)\right)$. This completes the proof. $\square$

Theorem 3.3. $\mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d H^1(\Omega_j)\right)$ *is weakly closed and therefore proximinal in* $H^1(\Omega)$.

Proof. Let $\{v_n\}_{n\in\mathbb{N}} \subset \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d H^1(\Omega_j)\right)$ satisfy $v_n \rightharpoonup v$ in $H^1(\Omega)$. Since $\left(L^2(\Omega)\right)^* \subset \left(H^1(\Omega)\right)^*$, $v_n \rightharpoonup v$ in $L^2(\Omega)$. By Lemma 3.2, $\{v_n\}_{n\in\mathbb{N}} \subset \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d L^2(\Omega_j)\right)$, i.e., $\dim U^j(v_n) \leq r_j$ for all $1 \leq j \leq d$. By [5, Theorem 6.24], $\dim U^j(v) \leq \liminf_{n\to\infty} \dim U^j(v) \leq r_j$, and thus $v \in \mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d H^1(\Omega_j)\right)$. Since $H^1(\Omega)$ is a reflexive Banach space, the set $\mathcal{T}_{\boldsymbol{r}}\left({}_a\bigotimes_{j=1}^d H^1(\Omega_j)\right)$ is proximinal. $\square$

3.2. Minimal Subspaces.

The subspaces from (3.2) inherit H^1 regularity.

Lemma 3.4. *For* $u \in H^1(\Omega)$, $U^j(u) \subset H^1(\Omega_j)$.

Proof. Let $v \in {}_a\bigotimes_{j=1}^d H^1(\Omega_j)$ and $\varphi_{[j]} := \bigotimes_{k=1}^d \varphi_k$, $\varphi_j = \mathrm{id}_j$, $\varphi_k \in L^2(\Omega_k)^*$, $k \neq j$. Clearly, $\varphi_{[j]}(v) \in H^1(\Omega_j)$. By [5, Lemma 4.97] and [5, Proposition 4.68]

$$\|\varphi_{[j]}(v)\|_1 \lesssim \|v\|_{\vee(e_j)} \lesssim \|v\|_{e_j}.$$

Thus, $\varphi_{[j]} : \left({}_a\bigotimes_{j=1}^d H^1(\Omega_j), \|\cdot\|_1\right) \to H^1(\Omega_j)$ is a continuous linear mapping. Since ${}_a\bigotimes_{j=1}^d H^1(\Omega_j)$ is dense in $H^1(\Omega)$, $\varphi_{[j]}$ can be uniquely extended to a bounded linear mapping on $H^1(\Omega)$ with the same operator norm, i.e., $\varphi_{[j]}(v) \in H^1(\Omega_j)$ is well defined for $v \in H^1(\Omega)$ and the statement follows. $\square$

Before we proceed, we would like to clarify that there are several possible definitions for minimal subspaces when considering $u \in H^1(\Omega)$. First, there are two possible choices for the dual space leading to

$$U_a^j(u) := \operatorname{span}\left\{\varphi(u) : \varphi = \bigotimes_{k=1}^{d} \varphi_k,\ \varphi_j = \mathrm{id},\ \varphi_k \in \left(L^2(\Omega_k)\right)^*,\ k \neq j\right\},$$

$$U_b^j(u) := \operatorname{span}\left\{\varphi(u) : \varphi = \bigotimes_{k=1}^{d} \varphi_k,\ \varphi_j = \mathrm{id},\ \varphi_k \in \left(H^1(\Omega_k)\right)^*,\ k \neq j\right\},$$

where clearly $U_a^j(u) = U^j(u)$. Second, there are two possible choices for the completion norm, which overall leads to four possible definitions

$$U_I^j(u) := \overline{U_a^j(u)}^{\|\cdot\|_0},\ U_{II}^j(u) := \overline{U_a^j(u)}^{\|\cdot\|_1},$$
$$U_{III}^j(u) := \overline{U_b^j(u)}^{\|\cdot\|_0},\ U_{IV}^j(u) := \overline{U_b^j(u)}^{\|\cdot\|_1}.$$

The space $U_I^j(u)$ is the minimal subspace of u as a function in $L^2(\Omega)$. For $d=2$, $U_{II}^1(u)$ (resp. $U_{II}^2(u)$) is the minimal subspace of u as a function in $H^{(1,0)}$ (resp. $H^{(0,1)}$), $U_{III}^1(u)$ (resp. $U_{III}^2(u)$) is the minimal subspace of u as a function in $H^{(0,1)}$ (resp. $H^{(1,0)}$) and $U_{IV}^j(u)$ is the minimal subspace of u as a function in $H^1_{\text{mix}}(\Omega)$. Since we want to consider precisely $u \in H^1(\Omega)$, we consider $u \in H^{(1,0)}$ and choose the variant $U_{II}^1(u)$ for the left minimal subspace, and $u \in H^{(0,1)}$ with the variant $U_{II}^2(u)$ for the right minimal subspace. Analogously for $d > 2$.

With the preceding lemma we may now define $U_j^{\min}(u) = \overline{U_a^j(u)}^{\|\cdot\|_1} \subset H^1(\Omega_j)$. This space differs from $U^j(u)$ in case $u \notin {}_a\bigotimes_{j=1}^d H^1(\Omega_j)$. We want to check if property (2.2) still holds for H^1 functions. To this end, we require the following assumption.

Assumption 3.5. *Let $\mathcal{P}^j : H^1(\Omega_j) \to U_j^{\min}(u)$ be an orthogonal projection. We assume $\mathcal{P}^j$ is continuous in L^2*

$$\sup_{\substack{v_j \in H^1(\Omega_j),\\ v \neq 0}} \frac{\|\mathcal{P}^j v_j\|_0}{\|v\|_0} < \infty.$$

Remark 3.6. *We will frequently encounter Assumption 3.5 in the following sections. We will discuss sufficient conditions for this assumption to be satisfied in part II of this series. We will see that this assumption is not necessarily satisfied. In fact, we conjecture that there are functions $u \in H^1(\Omega)$ which do not satisfy the statement of Theorem 3.7. The proof of this, however, seems to be not trivial.*

Proposition 3.7. *Let $u \in H^1(\Omega)$ and assume 3.5 is satisfied. Then, it holds $u \in {}_{\|\cdot\|_1}\bigotimes_{j=1}^d U_j^{\min}(u)$.*

Proof. Since $u \in H^1(\Omega)$, by (2.1), $u \in H^1(\Omega_j) \otimes_{\|\cdot\|_{e_j}} \left({}_a\bigotimes_{k\neq j} L^2(\Omega_k)\right)$. The space $U_j^{\min}(u) \subset H^1(\Omega_j)$ is a closed subspace of the Hilbert space $H^1(\Omega_j)$. Thus, $U_j^{\min}(u) \in \mathbb{G}(H^1(\Omega_j))$. Moreover, $\|\cdot\|_{e_j}$ is a uniform crossnorm. This holds for any $1 \leq j \leq d$ and thus by [5, Theorem 6.29] we obtain

$$u \in \bigcap_{k=1}^d U_k^{\min}(u) \otimes_{\|\cdot\|_{e_k}} \left({}_a\bigotimes_{j\neq k} L^2(\Omega_j)\right).$$

Next, following the arguments of [5, Theorem 6.28], consider the orthogonal projection $P^j : H^1(\Omega_j) \to U_j^{\min}(u)$. Let $\mathcal{P}^j := P^j \otimes \left(\bigotimes_{k\neq j}^d \mathrm{id}_k\right)$. This is a linear continuous mapping from $H^1(\Omega_j) \otimes_{\|\cdot\|_{e_j}} \left({}_a\bigotimes_{k\neq j} L_2(\Omega_k)\right)$ to $U_j^{\min}(u) \otimes_{\|\cdot\|_{e_j}} \left({}_a\bigotimes_{k\neq j} L^2(\Omega_k)\right)$, with $\|\mathcal{P}^j\| = \|P^j\| = 1$ (since $\|\cdot\|_{e_j}$ is a uniform crossnorm).

Take a sequence $\{u_n\}_{n\in\mathbb{N}} \subset U_j^{\min}(u) \otimes_a \left({}_a\bigotimes_{k\neq j} L^2(\Omega_k)\right)$ such that $\lim_{n\to\infty} \|u-u_n\|_{e_j} = 0$. Clearly, $\mathcal{P}^j(u_n) = u_n$ and

$$\begin{aligned}\|u - \mathcal{P}^j u\|_{e_j} &\leq \|u-u_n\|_{e_j} + \|u_n - \mathcal{P}^j u\|_{e_j} = \|u-u_n\|_{e_j} + \|\mathcal{P}^j(u_n-u)\|_{e_j}\\ &\leq 2\|u-u_n\|_{e_j}.\end{aligned}$$

Taking the limit with n, we obtain $u = \mathcal{P}^j u$.

Since this holds for any $1 \leq j \leq d$, we get

$$u = \left(\prod_{j=1}^d \mathcal{P}^j\right) u = \left(\bigotimes_{j=1}^d P^j\right) u. \tag{3.3}$$

Next, we require a separable representation for u that converges in $H^1(\Omega)$. This is possible for $H^1(\Omega)$ by choosing a complete $H^1(\Omega)$-orthonormal system of elementary tensor products (e.g., a Fourier basis) or an $H^1(\Omega)$ Riesz basis of wavelets. Let $\{\bigotimes_{j=1}^d \psi_{k_j}^j : (k_j)_{j=1}^d \in \mathbb{N}^d\}$ be such a system. Then, there exists a sequence $\boldsymbol{u} = (\boldsymbol{u}_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}^d} \in \ell_2(\mathbb{N}^d)$ such that

$$u = \lim_{n\to\infty} \sum_{k_1=1}^{n} \cdots \sum_{k_d=1}^{n} \boldsymbol{u}_{\boldsymbol{k}} \bigotimes_{j=1}^{d} \psi_{k_j}^j, \tag{3.4}$$

with convergence in $\|\cdot\|_1$. Since $\|\cdot\|_{e_j} \le \|\cdot\|_1, \quad 1 \le j \le d$, (3.4) converges in $\|\cdot\|_{e_j}$ for all $1 \le j \le d$ as well. Thus, by (3.3) and Assumption 3.5

$$u = \left(\bigotimes_{j=1}^{d} P^j\right) u = \lim_{n\to\infty} \sum_{k_1=1}^{n} \cdots \sum_{k_d=1}^{n} \boldsymbol{u}_{\boldsymbol{k}} \bigotimes_{j=1}^{d} P^j(\psi_{k_j}^j),$$

for any $1 \le i \le d$ and with convergence in $H^1(\Omega)$. We take $u_n := \sum_{k_1=1}^{n} \cdots \sum_{k_d=1}^{n} \boldsymbol{u}_{\boldsymbol{k}} \bigotimes_{j=1}^{d} P^j(\psi_{k_j}^j)$. Clearly, $u_n \in {}_a\bigotimes_{j=1}^d U_j^{\min}(u)$ and by above $\lim_{n\to\infty} \|u - u_n\|_1 = 0$. This completes the proof. □

4. H^1-Error Analysis of L^2-SVD

The singular value decomposition can be utilized to obtain spaces $U_j^{\min}(u)$ and low-rank approximations therein. Interestingly, the resulting spaces are not necessarily the same depending on the interpretation of $u \in H^1(\Omega)$. In the following we restrict the exposition to $d = 2$.

If we consider $u \in L^2(\Omega)$, then u can be identified with a compact operator from $L^2(\Omega_2)$ to $L^2(\Omega_1)$ such that

$$u[w] = \int_{\Omega_2} u(\cdot, y) w(y) dy. \tag{4.1}$$

for $w \in L^2(\Omega_2)$. The adjoint $u^* : L^2(\Omega_1) \to L^2(\Omega_2)$ is given by $u^*[v] = \int_{\Omega_1} u(x,\cdot) v(x) dx$, for $v \in L^2(\Omega_1)$. Thus, a left singular vector ψ of u satisfies $uu^*[\psi] = \int_{\Omega_2} u(\cdot, y) \int_{\Omega_1} u(x,y)\psi(x) dx dy = \lambda\psi$, for some $\lambda \in \mathbb{R}^+$, and the accompanying right singular vector satisfies $u^*u[\phi] = \int_{\Omega_1} u(x,\cdot) \int_{\Omega_2} u(x,y)\phi(y) dy dx = \lambda\phi$.

Since u is a compact operator, we can find an L^2-orthonormal system of left and right singular vectors, which we denote by $\{\psi_k\}_{k\in\mathbb{N}}$ and $\{\phi_k\}_{k\in\mathbb{N}}$, respectively, and the corresponding singular values by $\{\sigma_k^{00} = \sqrt{\lambda_k^{00}}\}_{k\in\mathbb{N}}$, sorted by decreasing values such that $u = \lim_{r\to\infty} \sum_{k=1}^r \sigma_k^{00} \psi_k \otimes \phi_k$, in $\|\cdot\|_0$. We have the identities $U_1(u) = \operatorname{span}\{\psi_k : k \in \mathbb{N},\ \sigma_k^{00} > 0\}$, $U_2(u) = \operatorname{span}\{\phi_k : k \in \mathbb{N},\ \sigma_k^{00} > 0\}$. The SVD provides both optimal low-rank approximations of given rank and an error estimator in the sense that

$$\|u - \sum_{k=1}^{r} \sigma_k^{00} \psi_k \otimes \phi_k\|_0 = \inf_{g \in \mathcal{R}_r(L^2(\Omega))} \|u - g\|_0,$$
$$\|u - \sum_{k=1}^{r} \sigma_k^{00} \psi_k \otimes \phi_k\|_0^2 = \sum_{k=r+1}^{\infty} (\sigma_k^{00})^2.$$

For the case $u \notin L^2(\Omega_1) \otimes_a L^2(\Omega_2)$, $\lambda_k^{00} > 0$ for all $k \in \mathbb{N}$. Otherwise, we only require finitely many ψ_k's and ϕ_k's. Letting $\gamma_k = \frac{1}{\lambda_k^{00}} \int_{\Omega_2} \frac{\partial}{\partial x} u(\cdot, y) \int_{\Omega_1} u(x,y)\psi_k(x) dx dy$, we have that

$$\begin{aligned}
\|\gamma_k\|_0^2 &= \frac{1}{(\lambda_k^{00})^2} \int_{\Omega_1} \left(\int_{\Omega_2} \frac{\partial}{\partial x} u(s,y) \int_{\Omega_1} u(x,y)\psi_k(x) dx dy \right)^2 ds \\
&\le \frac{1}{(\lambda_k^{00})^2} \int_{\Omega_1} \left(\int_{\Omega_2} \frac{\partial}{\partial x} u(s,y) \left(\int_{\Omega_1} u^2(x,y) dx \right)^{1/2} \left(\int_{\Omega_1} \psi_k^2(x) dx \right)^{1/2} dy \right)^2 ds \\
&\le \frac{1}{(\lambda_k^{00})^2} \|\psi_k\|_0^2 \int_{\Omega_1} \left(\left(\int_{\Omega_2} (\frac{\partial}{\partial x} u(s,y))^2 dy \right)^{1/2} \left(\int_{\Omega_2} \int_{\Omega_1} u^2(x,y) dx dy \right)^{1/2} \right)^2 ds \\
&= \frac{1}{(\lambda_k^{00})^2} \|u\|_0^2 \left\| \frac{\partial}{\partial x} u \right\|_0^2,
\end{aligned} \tag{4.2}$$

and for all $\varphi \in C_c^\infty(\Omega_1)$

$$
\begin{aligned}
\int_{\Omega_1} \varphi(s)\gamma_k(s) &= \int_{\Omega_1} \varphi(s)\frac{1}{\lambda_k^{00}} \int_{\Omega_2} \frac{\partial}{\partial x} u(s,y) \int_{\Omega_1} u(x,y)\varphi_k(x) dx dy ds \\
&= \frac{1}{\lambda_k^{00}} \int_{\Omega_2}\int_{\Omega_1} \varphi(s)\frac{\partial}{\partial s} u(s,y) ds \int_{\Omega_1} u(x,y)\varphi_k(x) dx dy \\
&= -\frac{1}{\lambda_k^{00}} \int_{\Omega_2}\int_{\Omega_1} \frac{d}{ds}\varphi(s) u(s,y) ds \int_{\Omega_1} u(x,y)\varphi_k(x) dx dy \\
&= -\int_{\Omega_1} \frac{d}{ds}\varphi(s)\frac{1}{\lambda_k^{00}} \int_{\Omega_2} u(s,y) \int_{\Omega_1} u(x,y)\varphi_k(x) dx dy ds \\
&= -\int_{\Omega_1} \frac{d}{ds}\varphi(s)\psi_k(s) ds, \qquad (4.3)
\end{aligned}
$$

so that $\gamma_k = \frac{d}{dx}\psi_k$. Analogously for ϕ_k,

$$
\frac{d}{dx}\phi_k = \frac{1}{\lambda_k^{00}} \int_{\Omega_1} \frac{\partial}{\partial y} u(x,\cdot) \int_{\Omega_2} u(x,y)\phi_k(y) dy dx.
$$

Thus, $\psi_k \in H^1(\Omega_1)$, $\phi_k \in H^1(\Omega_2)$ for all $k \in \mathbb{N}$ and, consistently with Lemma 3.4, $U_1(u) \subset H^1(\Omega_1)$ and $U_2(u) \subset H^1(\Omega_2)$. The best rank r approximation in L^2,

$$
u_r := \sum_{k=1}^r \sigma_k^{00} \psi_k \otimes \phi_k, \qquad (4.4)
$$

makes sense in H^1 and we can consider the error $\|u - u_r\|_1$.

Theorem 4.1. *Let $u \in H^1(\Omega)$ and u_r be its best rank r approximation in L^2 defined by (4.4). We have*

$$
\|u_r\|_1^2 = \sum_{k=1}^r (\sigma_k^{00})^2 \left(1 + \left\|\frac{d}{dx}\psi_k\right\|_0^2 + \left\|\frac{d}{dy}\phi_k\right\|_0^2\right) \qquad (4.5)
$$

If

$$
\lim_{r\to\infty} \|u_r\|_1 < \infty \qquad (4.6)
$$

then $\|u - u_r\|_1 \to 0$ and

$$
\|u - u_r\|_1^2 = \sum_{k=r+1}^\infty (\sigma_k^{00})^2 \left(1 + \left\|\frac{d}{dx}\psi_k\right\|_0^2 + \left\|\frac{d}{dy}\phi_k\right\|_0^2\right). \qquad (4.7)
$$

Proof. Clearly, if u_r converges to some $\tilde{u} \in H^1(\Omega)$, $u = \tilde{u}$ a.e. by the simple inequality

$$
\|u - \tilde{u}\|_0 \le \inf_{r\in\mathbb{N}} \{\|u - u_r\|_0 + \|\tilde{u} - u_r\|_1\} = 0. \qquad (4.8)
$$

We have

$$
\begin{aligned}
\|u_r\|_1^2 &= \left\|\sum_{k=1}^r \sigma_k^{00}\psi_k \otimes \phi_k\right\|_1^2 = \sum_{k,l=1}^r \sigma_k^{00}\sigma_l^{00} \langle \psi_k \otimes \phi_k, \psi_l \otimes \phi_l\rangle_1 \\
&= \sum_{k,l=1}^r \sigma_k^{00}\sigma_l^{00} \Bigg(\langle \psi_k \otimes \phi_k, \psi_l \otimes \phi_l\rangle_0 + \left\langle \frac{d}{dx}\psi_k \otimes \phi_k, \frac{d}{dx}\psi_l \otimes \phi_l \right\rangle_0 \\
&\quad + \left\langle \psi_k \otimes \frac{d}{dy}\phi_k, \psi_l \otimes \frac{d}{dy}\phi_l \right\rangle_0 \Bigg) \\
&= \sum_{k,l=1}^r \sigma_k^{00}\sigma_l^{00} \left(\delta_{kl}\delta_{kl} + \left\|\frac{d}{dx}\psi_k\right\|_0^2 \delta_{kl} + \delta_{kl}\left\|\frac{d}{dy}\phi_k\right\|_0^2\right) \\
&= \sum_{k=1}^r (\sigma_k^{00})^2 \left(1 + \left\|\frac{d}{dx}\psi_k\right\|_0^2 + \left\|\frac{d}{dy}\phi_k\right\|_0^2\right).
\end{aligned}
$$

Thus, $(\|u_r\|_1^2)_{r\in\mathbb{N}}$ is a positive increasing sequence. If (4.6) holds, then $(\|u_r\|_1^2)_{r\in\mathbb{N}}$ converges. Then, for $m \geq r$ $\|u_m - u_r\|_1^2 = \|u_m\|_1^2 - \|u_r\|_1^2$, which proves that u_r is Cauchy and therefore converges. Taking the limit and by (4.8), we obtain $\|u - u_r\|_1 \to 0$. The proof of (4.7) follows similarly as above. □

Remark 4.2. *Equation* (4.7) *is thus a recipe for constructing low-rank approximations via the L^2-SVD but with error control in H^1. Assumption* (4.6) *particularly holds when u is a numerical approximation to the solution of a PDE.*

We can not expect (4.6) *to hold in general. Specifically, in* (4.2) *we applied twice the Cauchy-Schwarz inequality, which is known to be sharp. Since $\lambda_k^{00} = (\sigma_k^{00})^2$, this would imply that* (4.6) *is not satisfied and u_r diverges in $H^1(\Omega)$. On the other hand, we can think of cases where* (4.6) *is satisfied, such as in the case of a Fourier basis.*

We ask what are the possible conditions on ψ_k and ϕ_k for (4.6) *to be satisfied? Note that this condition is similar to well-known estimates from approximation theory, specifically approximation via wavelets or, more generally, multi-scale approximation. There sufficient conditions include the existence of a uniformly bounded family of projectors that satisfy direct and inverse inequalities.*

Translated into our setting, sufficient conditions for (4.6) *look as follows. Define the subspaces $S_l := \operatorname{span}\left\{\psi_k : 1 \leq k \leq 2^l\right\} \subset U_1^{\min}(u)$, $l \in \mathbb{N}_0$. We require the* Jackson (direct) *inequality to be satisfied*

$$\inf_{v_l \in S_l} \|f - v_l\|_0 \lesssim 2^{-sl}\|f\|_1, \quad \forall f \in U_1^{\min}(u),$$

for some $s > 1$ and the Bernstein (indirect) *inequality $\|v_l\|_1 \lesssim 2^{\bar{s}l}\|v_l\|_0$, for all $v_l \in S_l$, for some $\bar{s} > 1$. Analogously for the space generated by the ϕ_k's. For more details we refer to* [13, Theorem 5.12] *and* [2]*.*

We conclude this part by extending the result to $d \geq 2$. Let $u \in H^1(\Omega)$ with $\Omega = \times_{j=1}^d \Omega_j$, $x = (x_1, \ldots, x_d) \in \Omega$. Then, for any $1 \leq i \leq d$, we can consider the integral operator

$$u_i : L^2(\Omega_i) \to L^2(\bigtimes_{j\neq i} \Omega_j), \quad u_i[w] = \int_{\Omega_i} u(\cdot, \ldots, x_i, \ldots, \cdot) w(x_i) dx_i, w \in L^2(\Omega_i).$$

As before, we can consider the singular vectors $\{\psi_k^i : k \in \mathbb{N}\}$, and the corresponding eigenvalues $\{\lambda_k^i \in \mathbb{R}^+ : k \in \mathbb{N}\}$. The derivatives are given by

$$\frac{d}{dx_i}\psi_k^i = \int_{\times_{j\neq i}\Omega_j} \frac{\partial}{\partial x_i} u(\ldots, x_{i-1}, \cdot, x_{i+1}, \ldots) \int_{\Omega_i} u(x)\psi_k^i(x_i) dx.$$

with the familiar estimate $\left\|\frac{d}{dx_i}\psi_k^i\right\|_0 \leq \frac{1}{\lambda_k^i}\|u\|_0 \left\|\frac{\partial}{\partial x_i} u\right\|_0$. The identity

$$U_i(u) = \operatorname{span}\left\{\psi_k^i : k \in \mathbb{N},\ \sigma_k^i = \sqrt{\lambda_k^i} > 0\right\}, \quad 1 \leq i \leq d,$$

holds. Define the subspace $B_{r_i}^i := \operatorname{span}\left\{\psi_k^i : 1 \leq k \leq r_i\right\}$, and the corresponding L^2-orthogonal projector $P_{r_i}^i : L^2(\Omega_i) \to B_{r_i}^i$, for $1 \leq i \leq d$. Then, for $\boldsymbol{r} = (r_1, \ldots, r_d) \in \mathbb{N}^d$, define

$$\mathcal{P}_{\boldsymbol{r}} := \bigotimes_{j=1}^{d} P_{r_j}^j, \quad \text{and} \quad u_{\boldsymbol{r}} := \mathcal{P}_{\boldsymbol{r}} u. \tag{4.9}$$

The projection $\mathcal{P}_{\boldsymbol{r}}$ is the HOSVD projection from Theorem 2.11. Before we proceed, we require the following lemma, which is an extension of Theorem 4.1.

Lemma 4.3. *Let $u \in H^1(\Omega)$ and $\mathcal{P}_{r_j}^j = \mathrm{id}_1 \otimes \cdots \otimes P_{r_j}^j \otimes \cdots \otimes \mathrm{id}_d$. We have*

$$\|\mathcal{P}_{r_j}^j u\|_{e_j}^2 = \sum_{k=1}^{r_j} (\sigma_k^j)^2 \left(1 + \left\|\frac{d}{dx}\psi_k^j\right\|_0^2\right).$$

If $\lim_{r_j\to\infty} \|\mathcal{P}_{r_j}^j u\|_{e_j} < \infty$, then $\|u - \mathcal{P}_{r_j}^j u\|_{e_j} \to 0$ and

$$\|u - \mathcal{P}_{r_j}^j u\|_{e_j}^2 = \sum_{k=r_j+1}^{\infty} (\sigma_k^j)^2 \left(1 + \left\|\frac{d}{dx}\psi_k^j\right\|_0^2\right)$$

Proof. We consider the matricisation

$$\mathcal{M}_{\{j\}}(u) : H^{e_j} \to H^1(\Omega_j) \otimes_{\|\cdot\|_{(1,0)}} L^2(\bigtimes_{i\neq j} \Omega_i).$$

This is a linear isometric isomorphism since $\|\cdot\|_{e_j}$ and $\|\cdot\|_{(1,0)}$ are canonical norms (induced by the same norms). The space $H^1(\Omega_j) \otimes_{\|\cdot\|_{(1,0)}} L^2(\bigtimes_{i\neq j} \Omega_i)$ is a Hilbert tensor space of order 2 equipped with the canonical norm, with the H^1 norm on the left and L^2 norm on the right. Thus, we can apply Theorem 4.1 to $\mathcal{M}_{\{j\}}(u)$ and the statement follows. ☐

For the H^1 error we get the following result. Remark 4.2 applies here as well.

Theorem 4.4. *Let $u \in H^1(\Omega)$ and $u_{\boldsymbol{r}}$ be defined by (4.9). We have*

$$\frac{1}{d}\sum_{j=1}^{d}\sum_{k=1}^{r_j}(\sigma_k^j)^2 \le \|u_{\boldsymbol{r}}\|_1^2 \le \sum_{j=1}^{d}\sum_{k=1}^{r_j}(\sigma_k^j)^2\left(1+\left\|\frac{d}{dx}\psi_k^j\right\|_0^2\right). \tag{4.10}$$

Define the constants $\Gamma_j(r_j) := \sup_{v\in B^j_{r_j}} \frac{\|v\|_1}{\|v\|_0}$. If

$$\lim_{\min_j r_j\to\infty}\sum_{j=1}^{d}\sum_{k=1}^{r_j}(\sigma_k^j)^2\left(1+\left\|\frac{d}{dx}\psi_k^j\right\|_0^2+\sum_{i\neq j}\Gamma_i^2(r_i)\right) < \infty \tag{4.11}$$

then $\|u-u_{\boldsymbol{r}}\|_1 \to 0$ and

$$\|u-u_{\boldsymbol{r}}\|_1^2 \sim \sum_{j=1}^{d}\sum_{k=r_j+1}^{\infty}(\sigma_k^j)^2\left(1+\left\|\frac{d}{dx}\psi_k^j\right\|_0^2\right). \tag{4.12}$$

Proof. Let $\mathcal{P}^j_{r_j}$ be defined as in Lemma 4.3. The projection $\prod_{i\neq j}^d \mathcal{P}^i_{r_i}$ is orthogonal in the $\|\cdot\|_{e_j}$ norm. The lower bound in (4.10) is an immediate consequence of [5, Theorem 10.3]. For the upper bound we get by applying Lemma 4.3

$$\begin{aligned}\|\mathcal{P}_{\boldsymbol{r}}u\|_1^2 &\le \sum_{j=1}^d \|\mathcal{P}_{\boldsymbol{r}}u\|_{e_j}^2 = \sum_{j=1}^d\left\|\left(\prod_{i\neq j}^d \mathcal{P}^i_{r_i}\right)\mathcal{P}^j_{r_j}u\right\|_{e_j}^2\\ &\le \sum_{j=1}^d\left\|\mathcal{P}^j_{r_j}u\right\|_{e_j}^2 = \sum_{j=1}^{d}\sum_{k=1}^{r_j}(\sigma_k^j)^2\left(1+\left\|\frac{d}{dx}\psi_k^j\right\|_0^2\right).\end{aligned}$$

Next, observe that we can bound the $\|\cdot\|_1$ norm of $P^j_{r_j}$ as follows

$$\|P^j_{r_j}v\|_1 \le \Gamma_j(r_j)\|\|P^j_{r_j}v\|_0 \le \Gamma_j(r_j)\|v\|_0 \le \Gamma_j(r_j)\|v\|_1,$$

for any $v\in H^1(\Omega_j)$. Thus, since $\|\cdot\|_{e_j}$ is a uniform crossnorm on H^{e_j}, $\|\mathcal{P}^j_{r_j}\|_{e_j} = \|P^j_{r_j}\|_1$, so that we can bound $\|\mathcal{P}^j_{r_j}\|_1 \le \Gamma_j(r_j)$.

Let $u_{\boldsymbol{m}} := \mathcal{P}_{\boldsymbol{m}}u$ be an HOSVD approximation as in (4.9) with $m_j > r_j$ for all j. Then, since $\prod_{i\neq j}^d \mathcal{P}^i_{r_i}$ is orthogonal w.r.t. $\|\cdot\|_{e_j}$ and applying again Lemma 4.3, we have

$$\begin{aligned}\|u_{\boldsymbol{m}}-u_{\boldsymbol{r}}\|_1^2 &\le \sum_{j=1}^d \|(\mathcal{P}_{\boldsymbol{m}}-\mathcal{P}_{\boldsymbol{r}})u\|_{e_j}^2\\ &\le \sum_{j=1}^d\left[\left\|(\mathcal{P}^j_{m_j}-\mathcal{P}^j_{r_j})\prod_{i\neq j}^d\mathcal{P}^i_{r_i}u\right\|_{e_j}+\left\|\mathcal{P}^j_{m_j}\left(\prod_{i\neq j}^d\mathcal{P}^i_{m_i}-\prod_{i\neq j}^d\mathcal{P}^i_{r_i}\right)u\right\|_{e_j}\right]^2\\ &\le 2\sum_{j=1}^{d}\sum_{k=r_j+1}^{m_j}(\sigma_k^j)^2\left(1+\left\|\frac{d}{dx}\psi_k^j\right\|_0^2\right)+\Gamma_j^2(m_j)\sum_{i\neq j}\sum_{k=r_i+1}^{m_i}(\sigma_k^i)^2.\end{aligned}$$

If (4.11) holds, then $u_{\boldsymbol{r}}$ is a Cauchy sequence in H^1 and by uniqueness of the limit we must have $\|u-u_{\boldsymbol{r}}\|_1 \to 0$.

Finally, we show the bounds in (4.12). The mapping $\mathbf{id} = \bigotimes_{j=1}^{d} \mathrm{id}_j$ is an orthogonal projection in the $\|\cdot\|_{e_j}$ norm and $\mathrm{Im}\left(\prod_{i\neq j}^{d} \mathcal{P}_{r_i}^i\right) \subset \mathrm{Im}\,(\mathbf{id})$. Thus, for any $1 \leq j \leq d$

$$\|(\mathbf{id} - \mathcal{P}_{\boldsymbol{r}})u\|_1^2 \geq \|(\mathbf{id} - \mathcal{P}_{\boldsymbol{r}})u\|_{e_j}^2 = \left\| \left[\mathbf{id} - \left(\prod_{i\neq j}^{d} \mathcal{P}_{r_i}^i \right) \mathcal{P}_{r_j}^j \right] u \right\|_{e_j}^2$$
$$\geq \left\| (\mathbf{id} - \mathcal{P}_{r_j}^j)u \right\|_{e_j}^2 = \sum_{k=r_j+1}^{\infty} (\sigma_k^j)^2 \left(1 + \left\| \frac{d}{dx}\psi_k^j \right\|_0^2 \right),$$

where the last equality is due to Lemma 4.3. This shows the lower bound in (4.12).

For the upper bound

$$\|(\mathbf{id} - \mathcal{P}_{\boldsymbol{r}})u\|_1^2 \leq \sum_{j=1}^{d} \|(\mathbf{id} - \mathcal{P}_{\boldsymbol{r}})u\|_{e_j}^2$$
$$= \sum_{j=1}^{d} \left\| (\mathbf{id} - \prod_{i\neq j}^{d} \mathcal{P}_{r_i}^i)u \right\|_{e_j}^2 + \left\| \prod_{i\neq j}^{d} \mathcal{P}_{r_i}^i (\mathbf{id} - \mathcal{P}_{r_j}^j)u \right\|_{e_j}^2$$
$$\leq \sum_{j=1}^{d} \left[\sum_{k=r_j+1}^{\infty} (\sigma_k^j)^2 + \sum_{k=r_j+1}^{\infty} (\sigma_k^j)^2 \left(1 + \left\| \frac{d}{dx}\psi_k^j \right\|_0^2 \right) \right]$$
$$= \sum_{j=1}^{d} \sum_{k=r_j+1}^{\infty} (\sigma_k^j)^2 \left(2 + \left\| \frac{d}{dx}\psi_k^j \right\|_0^2 \right).$$

This completes the proof. □

References

[1] Benner, P., Cohen, A., Ohlberger, M., and Willcox, K., Eds. *Model reduction and approximation*, vol. 15 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2017. Theory and algorithms.

[2] Dahmen, W. Some remarks on multiscale transformations, stability, and biorthogonality. In *Wavelets, images, and surface fitting (Chamonix-Mont-Blanc, 1993)*. A K Peters, Wellesley, MA, 1994, pp. 157–188.

[3] Diestel, J., Fourie, J. H., and Swart, J. *The metric theory of tensor products*. American Mathematical Society, Providence, RI, 2008. Grothendieck's résumé revisited.

[4] Grasedyck, L. Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl. 31*, 4 (2009/10), 2029–2054.

[5] Hackbusch, W. *Tensor spaces and numerical tensor calculus*, vol. 42 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2012.

[6] Hackbusch, W. L^∞ estimation of tensor truncations. *Numer. Math. 125*, 3 (2013), 419–440.

[7] Hackbusch, W. Truncation of tensors in the hierarchical format. *SEMA Springer series / Sociedad Espanola de MatemÃ¡tica Aplicada* (2018). Submitted / accepted.

[8] Hackbusch, W., and Khoromskij, B. N. Low-rank kronecker-product approximation to multi-dimensional nonlocal operators. part II. HKT representation of certain operators. *Computing 76*, 3-4 (dec 2005), 203–225.

[9] Khoromskaia, V. *Tensor Numerical Methods in Quantum Chemistry*. de Gruyter Oldenbourg, 2018.

[10] Khoromskij, B. N. *Tensor Numerical Methods in Scientific Computing*. de Gruyter Oldenbourg, 2018.

[11] Nouy, A. Low-rank tensor methods for model order reduction. In *Handbook of Uncertainty Quantification*. Springer International Publishing, 2017, pp. 857–882.

[12] Quarteroni, A., Manzoni, A., and Negri, F. *Reduced basis methods for partial differential equations*, vol. 92 of *Unitext*. Springer, Cham, 2016. An introduction, La Matematica per il 3+2.

[13] Urban, K. *Wavelet methods for elliptic partial differential equations*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2009.

[14] Uschmajew, A. Regularity of tensor product approximations to square integrable functions. *Constr. Approx. 34*, 3 (2011), 371–391.