Numerische
Mathematik

CrossMark

Higher-order principal component analysis for the approximation of tensors in tree-based low-rank formats

Anthony Nouy[1]

Abstract

This paper is concerned with the approximation of tensors using tree-based tensor formats, which are tensor networks whose graphs are dimension partition trees. We consider Hilbert tensor spaces of multivariate functions defined on a product set equipped with a probability measure. This includes the case of multidimensional arrays corresponding to finite product sets. We propose and analyse an algorithm for the construction of an approximation using only point evaluations of a multivariate function, or evaluations of some entries of a multidimensional array. The algorithm is a variant of higher-order singular value decomposition which constructs a hierarchy of subspaces associated with the different nodes of the tree and a corresponding hierarchy of interpolation operators. Optimal subspaces are estimated using empirical principal component analysis of interpolations of partial random evaluations of the function. The algorithm is able to provide an approximation in any tree-based format with either a prescribed rank or a prescribed relative error, with a number of evaluations of the order of the storage complexity of the approximation format. Under some assumptions on the estimation of principal components, we prove that the algorithm provides either a quasi-optimal approximation with a given rank, or an approximation satisfying the prescribed relative error, up to constants depending on the tree and the properties of interpolation operators. The analysis takes into account the discretization errors for the approximation of infinite-dimensional tensors. For a tensor with finite and known rank in a tree-based format, the algorithm is able to recover the tensor in a stable way using a number of evaluations equal to the storage complexity of the representation of the tensor in this format. Several numerical examples illustrate the main results and the behavior of the algorithm for the approximation of high-dimensional functions using hierarchical Tucker or tensor train tensor formats, and the approximation of univariate functions using tensorization.

Mathematics Subject Classification 15A69 · 41A05 · 41A63 · 65D15 · 65J99

✉ Anthony Nouy
anthony.nouy@ec-nantes.fr

1   Centrale Nantes, LMJL, UMR CNRS 6629, 1 rue de la Noë, 44321 Nantes, France

1 Introduction

The approximation of high-dimensional functions is one of the most challenging tasks in computational science. Such high-dimensional problems arise in many domains of physics, chemistry, biology or finance, where the functions are the solutions of high-dimensional partial differential equations (PDEs). Such problems also typically arise in statistics or machine learning, for the estimation of high-dimensional probability density functions, or the approximation of the relation between a certain random variable and some predictive variables, the typical task of supervised learning. The approximation of high-dimensional functions is also required in optimization or uncertainty quantification problems, where the functions represent the response of a system (or model) in terms of some parameters. These problems require many evaluations of the functions and are usually intractable when one evaluation requires a specific experimental set-up or one run of a complex numerical code.

The approximation of high-dimensional functions from a limited number of information on the functions requires exploiting low-dimensional structures of functions. This usually call for nonlinear approximation tools [10,43]. A prominent approach consists of exploiting the sparsity of functions relatively to a basis, a frame, or a more general dictionary of functions [4,6,44]. Another approach consists of exploiting low-rank structures of multivariate functions, interpreted as elements of tensor spaces, which is related to notions of sparsity in (uncountably infinite) dictionaries of separable functions. For a multivariate function $v(x_1, \ldots, x_d)$ defined on a product set $\mathcal{X}_1 \times \cdots \times \mathcal{X}_d$, which is here identified with a tensor of order d, a natural notion of rank is the *canonical rank*, which is the minimal integer r such that

$$v(x_1, \ldots, x_d) = \sum_{k=1}^{r} v_k^1(x_1) \ldots v_k^d(x_d)$$

for some univariate functions v_k^ν defined on $\mathcal{X}_\nu$. For $d = 2$, this corresponds to the unique notion of rank, which coincides with the matrix rank when the variables take values in finite index sets and v is identified with a matrix. A function with low canonical rank r has a number of parameters which scales only linearly with r and d. However, it turns out that this format has several drawbacks when $d > 2$ (see, e.g., [9,22]), which makes it unsuitable for approximation. Then, other notions of rank have been introduced. For a subset of dimensions α in $\{1, \ldots, d\}$, the α-rank of a function v is the minimal integer $\mathrm{rank}_\alpha(v)$ such that

$$v(x_1, \ldots, x_d) = \sum_{k=1}^{\mathrm{rank}_\alpha(v)} v_k^\alpha(x_\alpha) v_k^{\alpha^c}(x_{\alpha^c})$$

for some functions v_k^α and $v_k^{\alpha^c}$ of complementary groups of variables $x_\alpha = (x_\nu)_{\nu \in \alpha} \in \mathcal{X}_\alpha$ and $x_{\alpha^c} = (x_\nu)_{\nu \in \alpha^c} \in \mathcal{X}_{\alpha^c}$, with α^c the complementary subset of α in $\{1, \ldots, d\}$. Approximation formats can then be defined by imposing α-ranks for a collection of subsets α. More precisely, if A is a collection of subsets in $\{1, \ldots, d\}$, we define an approximation format

$$\mathcal{T}_r^A = \{v : \mathrm{rank}_\alpha(v) \le r_\alpha, \alpha \in A\} = \bigcap_{\alpha \in A} \mathcal{T}_{r_\alpha}^{\{\alpha\}},$$

where $r = (r_\alpha)_{\alpha \in A}$ is a tuple of integers. When A is a tree-structured collection of subsets (a subset of a dimension partition tree), $\mathcal{T}_r^A$ is a tree-based tensor format whose elements admit a hierarchical and data-sparse representation. Tree-based tensor formats are tree tensor networks, i.e. tensor networks with tree-structured graphs [36]. They include the hierarchical Tucker (HT) format [21] and the tensor-train (TT) format [38]. Tree-based formats have many favorable properties that make them favorable for numerical use. As an intersection of subsets of tensors with bounded α-rank, $\alpha \in A$, these formats inherit most of the nice properties of the low-rank approximation format for order-two tensors. In particular, under suitable assumptions on tensor norms, best approximation problems in the set $\mathcal{T}_r^A$ are well-posed [13,16]. Also, the α-rank of a tensor can be computed through singular value decomposition, and the notion of singular value decomposition can be extended (in different ways) to these formats [8,17,37]. Another interesting property, which is not exploited in the present paper, is the fact that the set $\mathcal{T}_r^A$ is a differentiable manifold [14,15,23,45], which has interesting consequences in optimization or model order reduction of dynamical systems in tensor spaces [30]. There are only a few results available on the approximation properties of tree-based formats [42]. However, it has been observed in many domains of applications that tree-based formats have a high approximation power (or expressive power). Hierarchical tensor formats have been recently identified with deep neural networks with a particular architecture [7].

The reader is referred to the monograph [20] and surveys [1,18,27,28,34,35] for an introduction to tensor numerical methods and an overview of recent developments in the field.

This paper is concerned with the problem of computing an approximation of a function $u(x_1, \ldots, x_d)$ using point evaluations of this function, where evaluations can be selected adaptively. This includes problems where the function represents the output of a black-box numerical code, a system or a physical experiment for a given value of the input variables $(x_1, \ldots, x_d)$. This also includes the solution of high-dimensional PDEs with a probabilistic interpretation, where Monte-Carlo methods can be used to obtain point evaluations of their solutions. This excludes problems where evaluations of the functions come as an unstructured data set. A multivariate function $u(x_1, \ldots, x_d)$ is here considered as an element of a Hilbert tensor space $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_d$ of real-valued functions defined on a product set $\mathcal{X}_1 \times \cdots \times \mathcal{X}_d$ equipped with a probability measure. This includes the case of multidimensional arrays when the variables x_ν take values in finite sets $\mathcal{X}_\nu$. In this case, a point evaluation corresponds to the evaluation of an entry of the tensor.

Several algorithms have been proposed for the construction of approximations in tree-based formats using point evaluations of functions or entries of tensors. Let us mention algorithms that use adaptive and structured evaluations of tensors [2,39] and statistical learning approaches that use unstructured (random) evaluations of functions [5,11,12,19]. Let us also mention the recent work [31] for the approximation in Tucker format, with an approach similar to the one proposed in the present paper.

In the present paper, we propose and analyse a new algorithm which is based on a particular extension of the singular value decomposition for the tree-based format $\mathcal{T}_r^A$ which allows us to construct an approximation using only evaluations of a function (or entries of a tensor). The proposed algorithm constructs a hierarchy of subspaces U_α of functions of groups of variables x_α, for all $\alpha \in A$, and associated interpolation operators I_{U_α} which are oblique projections onto U_α. For the construction of U_α for a particular node $\alpha \in A$, we interpret the function u as a random variable $u(\cdot, x_{\alpha^c})$ depending on a set of random variables x_{α^c} with values in the space of functions of the variables x_α. Then U_α is obtained by estimating the principal components of this function-valued random variable using random samples $u(\cdot, x_{\alpha^c}^k)$. In practice, we estimate the principal components from interpolations $I_{V_\alpha} u(\cdot, x_{\alpha^c}^k)$ of these samples on a subspace V_α which is a certain approximation space when α is a leaf of the tree, or the tensor product of subspaces $\{U_\beta\}_{\beta \in S(\alpha)}$ associated with the sons $S(\alpha)$ of the node α when α is not a leaf of the tree. This construction only requires evaluations of u on a product set of points which is the product of an interpolation grid in $\mathcal{X}_\alpha$ (unisolvent for the space V_α), and a random set of points in $\mathcal{X}_{\alpha^c}$. It is a sequential construction going from the leaves to the root of the tree.

The proposed algorithm can be interpreted as an extension of principal component analysis for tree-based tensors which provides a statistical estimation of low-dimensional subspaces of functions of groups of variables for the representation of a multivariate function. It is able to provide an approximation $u^\star$ in any tree-based format $\mathcal{T}_r^A$ with either a prescribed rank r or a prescribed relative error (by adapting the rank r). For a given r, it has the remarkable property that it is able to provide an approximation in $\mathcal{T}_r^A$ with a number of evaluations equal to the storage complexity of the resulting approximation. Under some assumptions on the estimation of principal components, we prove that the algorithm, up to some discretization error ρ, provides with high probability a quasi-optimal approximation with a prescribed rank, i.e.

$$\|u - u^\star\| \le c \min_{v \in \mathcal{T}_r^A} \|u - v\| + \rho,$$

where the constant c depends on the set A and the properties of orthogonal projections and interpolation operators associated with principal subspaces. Also, under some assumptions on the estimation of principal components and discretization error, we prove that the algorithm with prescribed tolerance ϵ is able to provide an approximation $u^\star$ such that

$$\|u - u^\star\| \le \tilde{c}\epsilon \|u\|$$

holds with high probability, where the constant $\tilde{c}$ depends on the set A and the properties of projections and interpolation operators. Sharp inequalities are obtained by considering the properties of projection and interpolation operators when restricted to minimal subspaces of tensors. The analysis takes into account the discretization errors for the approximation of infinite-dimensional tensors. For a tensor with finite and known rank in a tree-based format, and when there is no discretization error, the algorithm is able to recover the tensor in a stable way using a number of evaluations

equal to the storage complexity of the representation of the tensor in this format. This algorithm may have important applications in the manipulation of big data, by providing a way to reconstruct a multidimensional array from a limited number of entries (tensor completion).

The outline of the paper is as follows. In Sect. 2, we introduce some definitions and properties of projections in Hilbert spaces, with a particular attention on Hilbert spaces of functions and projections based on point evaluations. In Sect. 3, we recall basic definitions on tensors and Hilbert tensor spaces of functions defined on measured product sets. Then we introduce some definitions and properties of operators on tensor spaces, with partial point evaluation functionals as a particular case. Finally, we introduce definitions and properties of projections on tensor spaces, with a particular attention on orthogonal projection and interpolation. In Sect. 4, we introduce tree-based low-rank formats in a general setting including classical HT and TT formats. In Sect. 5, we first introduce the notion of principal component analysis for multivariate functions and then propose an extension of principal component analysis to tree-based tensor format. This is based on a new variant of higher-order singular value decomposition of tensors in tree-based format. In Sect. 6, we present and analyse a modified version of the algorithm presented in Sect. 5 which only requires point evaluations of functions, and which is based on empirical principal component analyses and interpolations. In Sect. 7, the behavior of the proposed algorithm is illustrated and analysed in several numerical experiments.

2 Projections

For two vector spaces V and W equipped with norms $\|\cdot\|_V$ and $\|\cdot\|_W$ respectively, we denote by $L(V, W)$ the space of linear operators from V to W. We denote by $\mathcal{L}(V, W)$ the space of linear and continuous operators from V to W, with bounded operator norm $\|A\|_{V \to W} = \max_{\|v\|_V = 1} \|Av\|_W$. We denote by $V^* = L(V, \mathbb{R})$ the algebraic dual of V and by $V' = \mathcal{L}(V, \mathbb{R})$ the topological dual of V, and we let $\|\cdot\|_{V \to \mathbb{R}} = \|\cdot\|_{V'}$. We denote by $\langle \cdot, \cdot \rangle$ the duality pairing between a space and its dual. We let $L(V) := L(V, V)$ and $\mathcal{L}(V) := \mathcal{L}(V, V)$, and we replace the notation $\|\cdot\|_{V \to V}$ by $\|\cdot\|_V$, where the latter notation also stands for the norm on V.

2.1 Projections

Let V be a Hilbert space and U be a finite-dimensional subspace of V. An operator P is a *projection* onto a subspace U if $\text{Im}(P) = U$ and $Pu = u$ for all $u \in U$.

The *orthogonal projection* P_U onto U is a linear and continuous operator which associates to $v \in V$ the unique solution $P_U v \in U$ of

$$\|v - P_U v\|_V = \min_{u \in U} \|v - u\|_V,$$

or equivalently $(u, P_U v - v) = 0, \ \forall u \in U$. The orthogonal projection P_U has operator norm $\|P_U\|_V = 1$.

Let W be a finite-dimensional subspace of V^* such that

$$\dim(W) = \dim(U), \text{ and} \tag{1a}$$

$$\{u \in U : \langle w, u \rangle = 0 \quad \text{for all } w \in W\} = \{0\}, \tag{1b}$$

where the latter condition is equivalent to $U \cap {}^\perp W = \{0\}$, with ${}^\perp W$ the *annihilator* of W in V (see [33, Definition 1.10.4]). Under the above assumptions, we have that for any $v \in V$, there exists a unique $u \in U$ such that $\langle w, u - v \rangle = 0$ for all $w \in W$.[1] This allows to define the projection P_U^W onto U along W which is the linear operator on V which associates to $v \in V$ the unique solution $P_U^W v \in U$ of

$$\langle w, P_U^W v - v \rangle = 0, \ \forall w \in W.$$

For $W = R_V U$, where $R_V : V \to V'$ is the Riesz map, the projection P_U^W coincides with the orthogonal projection P_U. A non orthogonal projection is called an *oblique projection*. If $W \subset V'$, then P_U^W is a projection from V onto U parallel to $Ker(P_U^W) = Z^\perp$, where $Z = R_V^{-1}W$. If $W \subset \tilde{U}'$, with $\tilde{U}$ a closed subspace of V, then $P_U^W|_{\tilde{U}}$ is a projection from $\tilde{U}$ onto U parallel to $Ker(P_U^W) \cap \tilde{U} = Z^\perp \cap \tilde{U}$, where $Z = R_{\tilde{U}}^{-1}W$, with $R_{\tilde{U}}$ the Riesz map from $\tilde{U}$ to $\tilde{U}'$.

Proposition 1 *Let $\tilde{U}$ be a closed subspace of V and assume that $U \subset \tilde{U}$ and $W \subset \tilde{U}'$.[2] Then P_U^W is a continuous operator from $\tilde{U}$ to V.*

Proof Let us equip W with the norm $\|w\|_W = \|w\|_{\tilde{U}'} = \max_{v \in \tilde{U}} \langle w, v \rangle / \|v\|_V$, such that for all $v \in \tilde{U}$, $\langle w, v \rangle \le \|w\|_W \|v\|_V$. Let

$$\alpha = \min_{0 \neq u \in U} \max_{0 \neq w \in W} \frac{\langle w, u \rangle}{\|u\|_V \|w\|_W}.$$

Assumption (1b) implies that $\alpha > 0$. Then for all $v \in \tilde{U}$, we have

$$\|P_U^W v\|_V \le \alpha^{-1} \max_{0 \neq w \in W} \frac{\langle w, P_U^W v \rangle}{\|w\|_W} = \alpha^{-1} \max_{0 \neq w \in W} \frac{\langle w, v \rangle}{\|w\|_W} \le \alpha^{-1} \|v\|_V,$$

which ends the proof.   □

Proposition 2 *Let P and $\tilde{P}$ be projections onto subspaces U and $\tilde{U}$ respectively and assume $U \subset \tilde{U}$. Then*

$$\tilde{P} P = P.$$

---

[1] Uniqueness comes from (1b) while existence comes from (1a) and (1b).

[2] Note that $V' \subset \tilde{U}'$ and we may have $W \not\subset V'$.

Moreover, *if P and $\tilde{P}$ are projections along W and $\tilde{W}$ respectively, with $W \subset \tilde{W}$,
then*

$$\tilde{P}P = P\tilde{P} = P.$$

Proof For all $v \in V$, $Pv \in U \subset \tilde{U}$, and therefore $\tilde{P}Pv = Pv$, which proves the first statement. For the second statement, by definition of the projection P, we have that $\langle \phi, P\tilde{P}v - \tilde{P}v \rangle = 0$ for all $\phi \in W$. Since $W \subset \tilde{W}$ and by definition of $\tilde{P}$, this implies that $\langle \phi, P\tilde{P}v - v \rangle = 0$ for all $\phi \in W$. By definition of Pv and since $P\tilde{P}v \in U$, this implies $P\tilde{P}v = Pv = \tilde{P}Pv$. $\qquad\square$

Proposition 3 *Let U and $\tilde{U}$ be two closed subspaces of V, with U of finite dimension. Let P_U be the orthogonal projection onto U and let P_U^W be the projection onto U along $W \subset \tilde{U}'$. For all $v \in \tilde{U}$,*

$$\|P_U^W v - P_U v\|_V \le \|P_U^W - P_U\|_{\tilde{U} \to V} \|v - P_U v\|_V,$$

with

$$\|P_U^W - P_U\|_{\tilde{U} \to V} = \|P_U^W\|_{(id - P_U)\tilde{U} \to V} \le \|P_U^W\|_{\tilde{U} \to V}.$$

Also, for all $v \in \tilde{U}$,

$$\|v - P_U^W v\|_V^2 \le (1 + \|P_U^W - P_U\|_{\tilde{U} \to V}^2)\|v - P_U v\|_V^2.$$

Proof For $v \in \tilde{U}$, $\|P_U^W v - P_U v\|_V = \|P_U^W(v - P_U v)\|_V = \|(P_U^W - P_U)(v - P_U v)\|_V \le \|P_U^W - P_U\|_{(id - P_U)\tilde{U} \to V} \|v - P_U v\|_V$, with $\|P_U^W - P_U\|_{(id - P_U)\tilde{U} \to V} = \|P_U^W - P_U\|_{\tilde{U} \to V} = \|P_U^W\|_{(id - P_U)\tilde{U} \to V}$. This proves the first statement. The second statement directly follows from $\|v - P_U^W v\|_V^2 = \|v - P_U v\|_V^2 + \|P_U v - P_U^W v\|_V^2$. $\square$

2.2 Projection of functions using point evaluations

Let V be a Hilbert space of functions defined on a set X. For $x \in X$, the point evaluation functional $\delta_x \in V^*$ is defined by $\langle \delta_x, v \rangle = v(x)$.

2.2.1 Interpolation

Let U be a n-dimensional subspace of V and let $\Gamma = \{x^k\}_{k=1}^n$ be a set of n interpolation points in X. The set of interpolation points Γ is assumed to be unisolvent for U, i.e. for any $(a_k)_{k=1}^n \in \mathbb{R}^n$, there exists a unique $u \in U$ such that $u(x^k) = a_k$ for all $1 \le k \le n$. The interpolation operator I_U associated with Γ is a linear operator from V to U such that for $v \in V$, $I_U v$ is the unique element of U such that

$$\langle \delta_x, I_U v - v \rangle = I_U v(x) - v(x) = 0 \quad \forall x \in \Gamma.$$

The interpolation operator I_U is an oblique projection P_U^W onto U along $W = \mathrm{span}\{\delta_x : x \in \Gamma\}$. Note that the condition that Γ is unisolvent for U is equivalent to the condition (1b) on U and W, which ensures that I_U is well defined. From Proposition 2, we deduce the following property.

Proposition 4 *Let U and $\tilde{U}$ be two subspaces associated with sets of interpolation points Γ and $\tilde{\Gamma}$ respectively. If $U \subset \tilde{U}$ and $\Gamma \subset \tilde{\Gamma}$, then*

$$I_U I_{\tilde{U}} = I_{\tilde{U}} I_U = I_U.$$

Magic points. For a given basis $\{\varphi_i\}_{i=1}^n$ of U, a set of interpolation points $\Gamma = \{x^k\}_{k=1}^n$, called *magic points*, can be determined with a greedy algorithm proposed in [32, Remark 2]. The procedure for selecting the set Γ in a subset $\Gamma_\star$ in X is as follows. We first determine a point $x^1 \in \Gamma_\star$ and an index i_1 such that

$$|\varphi_{i_1}(x^1)| = \max_{x \in \Gamma_\star} \max_{1 \leq i \leq n} |\varphi_i(x)|.$$

Then for $k \geq 1$, we define $\psi_i^{(k)}(x) = \varphi_i(x) - \sum_{m=1}^{k} \sum_{p=1}^{k} \varphi_{i_m}(x) a_{m,p}^{(k)} \varphi_i(x^p)$, with the matrix $(a_{m,p}^{(k)})_{1 \leq m, p \leq k}$ being the inverse of the matrix $(\varphi_{i_m}(x^p))_{1 \leq p \leq k, 1 \leq m \leq k}$, such that $\psi_{i_m}^{(k)}(x) = 0$ for all $1 \leq m \leq k$ and $x \in X$, and $\psi_i^{(k)}(x^p) = 0$ for all $1 \leq p \leq k$ and $1 \leq i \leq n$. Then, we determine the point $x^{k+1} \in \Gamma_\star$ and an index i_{k+1} such that

$$|\psi_{i_{k+1}}^{(k)}(x^{k+1})| = \max_{x \in \Gamma_\star} \max_{1 \leq i \leq n} |\psi_i^{(k)}(x)|.$$

2.2.2 Discrete least-squares projection

Let U be a n-dimensional subspace of V and let $\Gamma = \{x^k\}_{k=1}^m$ be a set of m points in X, $m \geq n$, such that $\|v\|_\Gamma = (\sum_{x \in \Gamma} v(x)^2)^{1/2}$ defines a norm on U. The discrete least-squares projection Q_U is the linear operator from V to U such that for $v \in V$, $Q_U v$ is the unique element in U which minimizes $\|v - u\|_\Gamma^2$ over all $u \in U$, or equivalently

$$(u, v - Q_U v)_\Gamma = \sum_{x \in \Gamma} u(x) \langle \delta_x, v - Q_U v \rangle = 0 \quad \forall u \in U,$$

where $(\cdot, \cdot)_\Gamma$ is the inner product associated with the norm $\| \cdot \|_\Gamma$ on U. The discrete least-squares projection Q_U is an oblique projection onto U along $W = \{\sum_{x \in \Gamma} u(x) \delta_x : u \in U\}$. If $\#\Gamma = \dim(U)$ and Γ is unisolvent for U, then Q_U coincides with the interpolation operator I_U.

Proposition 5 *Let U and $\tilde{U}$ be two finite-dimensional subspaces such that $U \subset \tilde{U}$. Let Q_U be the discrete least-squares projection onto U associated with a set of points*

Γ in X, and let $Q_{\tilde{U}}$ be the discrete least-squares projection onto $\tilde{U}$ associated with a set of points $\tilde{\Gamma}$ in X. If either $\Gamma = \tilde{\Gamma}$ or $\Gamma \subset \tilde{\Gamma}$ and $\tilde{\Gamma}$ is unisolvent for $\tilde{U}$, then

$$Q_U Q_{\tilde{U}} = Q_{\tilde{U}} Q_U = Q_U.$$

Proof Q_U is the projection onto U along $W = \{\sum_{x \in \Gamma} u(x)\delta_x : u \in U\}$, and $Q_{\tilde{U}}$ is the projection onto $\tilde{U}$ along $\tilde{W} = \{\sum_{x \in \tilde{\Gamma}} \tilde{u}(x)\delta_x : \tilde{u} \in \tilde{U}\}$. If we prove that $W \subset \tilde{W}$, then the result follows from Proposition 2. Let $w = \sum_{x \in \Gamma} u(x)\delta_x \in W$, with $u \in U$. If $\Gamma = \tilde{\Gamma}$, then since $u \in \tilde{U}$, we clearly have $w \in \tilde{W}$. If $\Gamma \subset \tilde{\Gamma}$ and $\tilde{\Gamma}$ is unisolvent for $\tilde{U}$, there exists a function $\tilde{u} \in \tilde{U}$ such that $\tilde{u}(x) = u(x)$ for all $x \in \Gamma$ and $\tilde{u}(x) = 0$ for all $x \in \tilde{\Gamma}\backslash\Gamma$. Therefore, $w = \sum_{x \in \tilde{\Gamma}} \tilde{u}(x)\delta_x$ is an element of $\tilde{W}$, which ends the proof. $\qquad\square$

3 Tensors

Let $\mathcal{H}_\nu$ be Hilbert spaces of real-valued functions defined on sets $\mathcal{X}_\nu$ equipped with probability measures μ_ν, $1 \le \nu \le d$. We denote by $\| \cdot \|_{\mathcal{H}_\nu}$ the norm on $\mathcal{H}_\nu$ and by $(\cdot, \cdot)_{\mathcal{H}_\nu}$ the associated inner product. Let $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$ and $\mu = \mu_1 \otimes \cdots \otimes \mu_d$. The tensor product of d functions $v^\nu \in \mathcal{H}_\nu$, $1 \le \nu \le d$, denoted $v^1 \otimes \cdots \otimes v^d$, is a multivariate function defined on $\mathcal{X}$ such that $(v^1 \otimes \cdots \otimes v^d)(x) = v^1(x_1) \ldots v^d(x_d)$ for $x = (x_1, \ldots, x_d) \in \mathcal{X}$. Such a function is called an *elementary tensor*. The algebraic tensor space $\mathcal{H}_1 \otimes_a \ldots \otimes_a \mathcal{H}_d$ is defined as the linear span of all elementary tensors, which is a pre-Hilbert space when equipped with the canonical inner product $(\cdot, \cdot)$ defined for elementary tensors by

$$(v^1 \otimes \cdots \otimes v^d, w^1 \otimes \cdots w^d) = (v^1, w^1)_{\mathcal{H}_1} \ldots (v^d, w^d)_{\mathcal{H}_d},$$

and then extended by linearity to the whole algebraic tensor space. We denote by $\| \cdot \|$ the norm associated with inner product $(\cdot, \cdot)$. A Hilbert tensor space $\mathcal{H} = \overline{\mathcal{H}_1 \otimes_a \ldots \otimes_a \mathcal{H}_d}^{\|\cdot\|}$ is then obtained by the completion of the algebraic tensor space, which we simply denote

$$\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_d = \bigotimes_{\nu=1}^{d} \mathcal{H}_\nu.$$

Example 1 Consider finite sets $\mathcal{X}_\nu$ and $\mathcal{H}_\nu = \mathbb{R}^{\mathcal{X}_\nu}$ equipped with the norm $\|v\|_{\mathcal{H}_\nu}^2 = \sum_{x_\nu \in \mathcal{X}_\nu} \mu_\nu(\{x_\nu\})|v(x_\nu)|^2$. Then, $\mathcal{H}$ is the space of multidimensional arrays $\mathbb{R}^{\mathcal{X}_1} \otimes \cdots \otimes \mathbb{R}^{\mathcal{X}_d}$ and $\|v\|^2 = \sum_{x \in \mathcal{X}} \mu(\{x\})|v(x)|^2$, where $\mu(\{x_1, \ldots, x_d\}) = \prod_{\nu=1}^{d} \mu_\nu(\{x_\nu\})$.

Example 2 Consider $\mathcal{X}_\nu = \mathbb{R}$, μ_ν a finite measure on $\mathbb{R}$, and $\mathcal{H}_\nu = L_{\mu_\nu}^2(\mathcal{X}_\nu)$ equipped with the natural norm $\|v\|_{\mathcal{H}_\nu}^2 = \int |v(x_\nu)|^2 \mu_\nu(dx_\nu)$. Then $\mathcal{H}$ is identified with $L_\mu^2(\mathcal{X})$, where $\mu = \mu_1 \otimes \cdots \otimes \mu_d$, and $\|v\|^2 = \int |v(x)|^2 \mu(dx)$.

Example 3 Consider for $\mathcal{H}_\nu$ a reproducing kernel Hilbert space (RKHS) with reproducing kernel $k_\nu : \mathcal{X}_\nu \times \mathcal{X}_\nu \to \mathbb{R}$. Then $\mathcal{H}$ is a RKHS with reproducing kernel $k(x, x') = k_1(x_1, x'_1) \dots k_d(x_d, x'_d)$.

For a non-empty subset α in $\{1, \dots, d\} := D$, we let $\mathcal{X}_\alpha$ be the set $\times_{\nu \in \alpha} \mathcal{X}_\nu$ equipped with the product measure $\mu_\alpha = \bigotimes_{\nu \in \alpha} \mu_\nu$. We denote by $\mathcal{H}_\alpha = \bigotimes_{\nu \in \alpha} \mathcal{H}_\nu$ the Hilbert tensor space of functions defined on $\mathcal{X}_\alpha$, equipped with the canonical norm $\| \cdot \|_{\mathcal{H}_\alpha}$ such that

$$\| \bigotimes_{\nu \in \alpha} v^\nu \|_{\mathcal{H}_\alpha} = \prod_{\nu \in \alpha} \| v^\nu \|_{\mathcal{H}_\nu}$$

for $v^\nu \in \mathcal{H}_\nu$, $1 \leq \nu \leq d$. We have $\mathcal{H}_D = \mathcal{H}$ and we use the convention $\mathcal{H}_\emptyset = \mathbb{R}$.
Matricisations and α-ranks. Let $\alpha \subset D$, with $\alpha \notin \{\emptyset, D\}$, and let $\alpha^c = D \backslash \alpha$ be its complement in D. For $x \in \mathcal{X}$, we denote by x_α the subset of variables $(x_\nu)_{\nu \in \alpha}$. A tensor $v \in \mathcal{H}$ can be identified with an order-two tensor

$$\mathcal{M}_\alpha(v) \in \mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c},$$

where $\mathcal{M}_\alpha$ is the *matricisation operator* associated with α, which defines a linear isometry between $\mathcal{H}$ and $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$. We use the conventions $\mathcal{M}_\emptyset(v) = \mathcal{M}_D(v) = v$ and $\mathcal{H}_\emptyset \otimes \mathcal{H}_D = \mathcal{H}_D \otimes \mathcal{H}_\emptyset = \mathcal{H}$.

The α-*rank* of a tensor $v \in \mathcal{H}$, denoted $\mathrm{rank}_\alpha(v)$, is defined as the rank of the order-two tensor $\mathcal{M}_\alpha(v)$, which is uniquely defined as the minimal integer such that

$$\mathcal{M}_\alpha(v) = \sum_{k=1}^{\mathrm{rank}_\alpha(v)} v_k^\alpha \otimes v_k^{\alpha^c}, \quad \text{or equivalently} \quad v(x) = \sum_{k=1}^{\mathrm{rank}_\alpha(v)} v_k^\alpha(x_\alpha) v_k^{\alpha^c}(x_{\alpha^c}), \quad (2)$$

for some functions $v_k^\alpha \in \mathcal{H}_\alpha$ and $v_k^{\alpha^c} \in \mathcal{H}_{\alpha^c}$ of complementary subsets of variables x_α and x_{α^c} respectively. By convention, we have $\mathrm{rank}_\emptyset(v) = \mathrm{rank}_D(v) = 1$. From now on, when there is no ambiguity, $\mathcal{M}_\alpha(v)$ and $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ will be identified with v and $\mathcal{H}$ respectively.
Minimal subspaces. The *minimal subspace* $U_\alpha^{min}(v)$ of v is defined as the smallest closed subspace in $\mathcal{H}_\alpha$ such that

$$v \in U_\alpha^{min}(v) \otimes \mathcal{H}_{\alpha^c},$$

and we have $\mathrm{rank}_\alpha(v) = \dim(U_\alpha^{min}(v))$ (see [13]). If v admits the representation (2), then $U_\alpha^{min}(v)$ is the closure of $\mathrm{span}\{v_k^\alpha\}_{k=1}^{\mathrm{rank}_\alpha(v)}$. For any partition $S(\alpha)$ of α, we have

$$U_\alpha^{min}(v) \subset \bigotimes_{\beta \in S(\alpha)} U_\beta^{min}(v).$$

We have $U_D^{min}(v) = \mathbb{R}v$ and for any partition $S(D)$ of D,

$$v \in \bigotimes_{\beta \in S(D)} U_\beta^{min}(v).$$

3.1 Operators on tensor spaces

Let consider the Hilbert tensor space $\mathcal{H} = \bigotimes_{v=1}^d \mathcal{H}_v$ equipped with the canonical norm $\|\cdot\|$. For linear operators from $\mathcal{H}$ to $\mathcal{H}$, we also denote by $\|\cdot\|$ the operator norm $\|\cdot\|_{\mathcal{H} \to \mathcal{H}} = \|\cdot\|_{\mathcal{H}}$.

We denote by id the identity operator on $\mathcal{H}$. For a non-empty subset $\alpha \subset D$, we denote by id_α the identity operator on $\mathcal{H}_\alpha$. For A_α in $L(\mathcal{H}_\alpha)$, we define the linear operator $A_\alpha \otimes id_{\alpha^c}$ such that for $v^\alpha \in \mathcal{H}_\alpha$ and $v^{\alpha^c} \in \mathcal{H}_{\alpha^c}$,

$$(A_\alpha \otimes id_{\alpha^c})(v^\alpha \otimes v^{\alpha^c}) = (A_\alpha v^\alpha) \otimes v^{\alpha^c},$$

and we extend this definition by linearity to the whole algebraic tensor space $\mathcal{H}_\alpha \otimes_a \mathcal{H}_{\alpha^c}$. For a finite dimensional tensor space $\mathcal{H}$, this completely characterizes a linear operator on $\mathcal{H}$. For an infinite dimensional tensor space $\mathcal{H}$, if $A_\alpha \in \mathcal{L}(U_\alpha, \mathcal{H}_\alpha)$, with $U_\alpha \subset \mathcal{H}_\alpha$, then $A_\alpha \otimes id_{\alpha^c}$ can be extended by continuity to $U_\alpha \otimes \mathcal{H}$.

We denote by $\mathcal{A}_\alpha$, using calligraphic font style, the linear operator in $L(\mathcal{H})$ associated with an operator A_α in $L(\mathcal{H}_\alpha)$, defined by $\mathcal{A}_\alpha = \mathcal{M}_\alpha^{-1}(A_\alpha \otimes id_{\alpha^c})\mathcal{M}_\alpha$, and simply denoted

$$\mathcal{A}_\alpha = A_\alpha \otimes id_{\alpha^c}$$

when there is no ambiguity. If $A_\alpha \in \mathcal{L}(\mathcal{H}_\alpha)$, then $\mathcal{A}_\alpha \in \mathcal{L}(\mathcal{H})$ and the two operators have the same operator norm $\|\mathcal{A}_\alpha\| = \|A_\alpha\|_{\mathcal{H}_\alpha}$. Also, we have the following more general result.

Proposition 6 *If $A_\alpha \in \mathcal{L}(U_\alpha, \mathcal{H}_\alpha)$, with $U_\alpha \subset \mathcal{H}_\alpha$, then $\mathcal{A}_\alpha \in \mathcal{L}(U_\alpha \otimes \mathcal{H}_{\alpha^c}, \mathcal{H})$ and the two operators have the same operator norm*

$$\|\mathcal{A}_\alpha\|_{U_\alpha \otimes \mathcal{H}_{\alpha^c} \to \mathcal{H}} = \|A_\alpha\|_{U_\alpha \to \mathcal{H}_\alpha}.$$

Corollary 1 *For a tensor $v \in \mathcal{H}$ and an operator $A_\alpha \in \mathcal{L}(U_\alpha^{min}(v), \mathcal{H}_\alpha)$,*

$$\|\mathcal{A}_\alpha v\| \leq \|A_\alpha\|_{U_\alpha^{min}(v) \to \mathcal{H}_\alpha} \|v\|.$$

Let $S = \{\alpha_1, \ldots, \alpha_K\}$ be a collection of disjoint subsets of D and let $A_\alpha \in L(\mathcal{H}_\alpha)$ be linear operators, $\alpha \in S$. Then we can define a linear operator $A_{\alpha_1} \otimes \cdots \otimes A_{\alpha_K} := \bigotimes_{\alpha \in S} A_\alpha$ on $\mathcal{H}_{\alpha_1} \otimes_a \ldots \otimes_a \mathcal{H}_{\alpha_K}$ such that

$$\left(\bigotimes_{\alpha \in S} A_\alpha \right) \left(\bigotimes_{\alpha \in S} v^\alpha \right) = \bigotimes_{\alpha \in S} (A_\alpha v^\alpha)$$

for $v^\alpha \in \mathcal{H}_\alpha$, $\alpha \in S$. The operator $\bigotimes_{\alpha \in S} A_\alpha$ can be identified with an operator

$$\mathcal{A} = \prod_{\alpha \in S} \mathcal{A}_\alpha,$$

defined on the algebraic tensor space $\mathcal{H}_1 \otimes_a \ldots \otimes_a \mathcal{H}_d$. The definition of $\mathcal{A}$ is independent of the ordering of the elements of S. If the operators A_α are continuous, then $\mathcal{A}$ defines a continuous operator from $\mathcal{H}$ to $\mathcal{H}$ and since $\| \cdot \|$ is a uniform crossnorm (see [20, Proposition 4.127]), the operator $\mathcal{A}$ has for operator norm

$$\| \mathcal{A} \| = \prod_{\alpha \in S} \| \mathcal{A}_\alpha \| = \prod_{\alpha \in S} \| A_\alpha \|_{\mathcal{H}_\alpha}.$$

Also, we have the following more general result.

Proposition 7 *Let S be a collection of disjoint subsets of D and let $\beta \subset D$ such that $\beta \cup (\cup_{\alpha \in S}\alpha) = D$. Let U_α be a subspace of $\mathcal{H}_\alpha$ and $A_\alpha \in \mathcal{L}(U_\alpha, \mathcal{H}_\alpha)$, for $\alpha \in S$. Then $\mathcal{A} = \prod_{\alpha \in S} A_\alpha$ is a continuous operator from $\mathcal{U} := (\bigotimes_{\alpha \in S} U_\alpha) \otimes \mathcal{H}_\beta$ to $\mathcal{H}$ such that*

$$\| \mathcal{A} \|_{\mathcal{U} \to \mathcal{H}} = \prod_{\alpha \in S} \| \mathcal{A}_\alpha \|_{U_\alpha \otimes \mathcal{H}_{\alpha^c} \to \mathcal{H}} = \prod_{\alpha \in S} \| A_\alpha \|_{U_\alpha \to \mathcal{H}_\alpha}.$$

Corollary 2 *Let S be a collection of disjoint subsets of D. For a tensor $v \in \mathcal{H}$ and operators A_α, $\alpha \in S$, such that $A_\alpha \in \mathcal{L}(U_\alpha^{min}(v), \mathcal{H}_\alpha)$, the operator $\mathcal{A} = \prod_{\alpha \in S} A_\alpha$ is such that*

$$\| \mathcal{A}v \| \leq \|v\| \prod_{\alpha \in S} \| A_\alpha \|_{U_\alpha^{min}(v) \to \mathcal{H}_\alpha}.$$

3.2 Partial evaluations of tensors

Let α be a non-empty subset of D. For a linear form $\psi_\alpha \in \mathcal{H}_\alpha^*$, $\psi_\alpha \otimes id_{\alpha^c}$ is a linear operator from $\mathcal{H}_\alpha \otimes_a \mathcal{H}_{\alpha^c}$ to $\mathcal{H}_{\alpha^c}$ such that $(\psi_\alpha \otimes id_{\alpha^c})(v^\alpha \otimes v^{\alpha^c}) = \psi_\alpha(v^\alpha)v^{\alpha^c}$. If $\psi_\alpha \in \mathcal{H}_\alpha'$, the definition of $\psi_\alpha \otimes id_{\alpha^c}$ can be extended by continuity to $\mathcal{H}$. Then $\psi_\alpha \otimes id_{\alpha^c}$ is a continuous operator from $\mathcal{H}$ to $\mathcal{H}_{\alpha^c}$ with operator norm $\| \psi_\alpha \otimes id_{\alpha^c} \|_{\mathcal{H} \to \mathcal{H}_{\alpha^c}} = \| \psi_\alpha \|_{\mathcal{H}_\alpha'}$. Also, we have the following result.

Proposition 8 *If $\psi_\alpha \in U_\alpha'$, with U_α a subspace of $\mathcal{H}_\alpha$, then $\psi_\alpha \otimes id_{\alpha^c} \in \mathcal{L}(U_\alpha \otimes \mathcal{H}_{\alpha^c}, \mathcal{H}_{\alpha^c})$ and*

$$\| \psi_\alpha \otimes id_{\alpha^c} \|_{U_\alpha \otimes \mathcal{H}_{\alpha^c} \to \mathcal{H}_{\alpha^c}} = \| \psi_\alpha \|_{U_\alpha'}.$$

Corollary 3 *For a tensor $v \in \mathcal{H}$ and $\psi_\alpha \in U_\alpha^{min}(v)'$, we have*

$$\| (\psi_\alpha \otimes id_{\alpha^c})v \| \leq \| \psi_\alpha \|_{(U_\alpha^{min}(v))'} \|v\|.$$

For a point $x_\alpha \in \mathcal{X}_\alpha$, we denote by $\delta_{x_\alpha} \in \mathcal{H}_\alpha^*$ the point evaluation functional at x_α, defined by $\langle \delta_{x_\alpha}, v^\alpha \rangle = v^\alpha(x_\alpha)$ for $v^\alpha \in \mathcal{H}_\alpha$. Then $\delta_{x_\alpha} \otimes id_{\alpha^c}$ defines a partial evaluation functional, which is a linear operator from $\mathcal{H}$ to $\mathcal{H}_{\alpha^c}$ such that

$$(\delta_{x_\alpha} \otimes id_{\alpha^c})(v^\alpha \otimes v^{\alpha^c}) = v^\alpha(x_\alpha) v^{\alpha^c}.$$

From Corollary 3, we deduce that for a given tensor $v \in \mathcal{H}$, if $\delta_{x_\alpha} \in U_\alpha^{min}(v)'$, then the definition of $\delta_{x_\alpha} \otimes id_{\alpha^c}$ can be extended by continuity to $U_\alpha^{min}(v) \otimes \mathcal{H}_{\alpha^c}$ and the partial evaluation

$$v(x_\alpha, \cdot) = (\delta_{x_\alpha} \otimes id_{\alpha^c})v$$

is an element of $\mathcal{H}_{\alpha^c}$ such that

$$\|v(x_\alpha, \cdot)\|_{\mathcal{H}_{\alpha^c}} = \|(\delta_{x_\alpha} \otimes id_{\alpha^c})v\| \leq \|\delta_{x_\alpha}\|_{U_\alpha^{min}(v)'}\|v\|.$$

3.3 Projection of tensors

Let α be a non-empty and strict subset of D and let U_α be a finite-dimensional subspace of $\mathcal{H}_\alpha$. If P_α is a projection from $\mathcal{H}_\alpha$ onto U_α, then $P_\alpha \otimes id_{\alpha^c}$ is a projection from $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ onto $U_\alpha \otimes \mathcal{H}_{\alpha^c}$.

Proposition 9 *Let $v \in \mathcal{H}$ and $\alpha, \beta \subset D$. Let P_β be a projection from $\mathcal{H}_\beta$ to a subspace U_β and let $\mathcal{P}_\beta$ be the corresponding projection onto $U_\beta \otimes \mathcal{H}_{\beta^c}$. If $\beta \subset \alpha$ or $\beta \subset D \backslash \alpha$, we have*

$$\text{rank}_\alpha(\mathcal{P}_\beta v) \leq \text{rank}_\alpha(v).$$

Proof A tensor v admits a representation $v = \sum_{k=1}^{\text{rank}_\alpha(v)} v_k^\alpha \otimes w_k^{\alpha^c}$. If $\beta \subset \alpha$, then $\mathcal{P}_\beta = (P_\beta \otimes id_{\alpha \backslash \beta}) \otimes id_{D \backslash \alpha}$ and $\mathcal{P}_\beta v = \sum_{k=1}^{\text{rank}_\alpha(v)}((P_\beta \otimes id_{\alpha \backslash \beta})v_k^\alpha) \otimes w_k^{\alpha^c}$. If $\beta \subset D \backslash \alpha$, then $\mathcal{P}_\beta = id_\alpha \otimes (P_\beta \otimes id_{D \backslash \{\alpha \cup \beta\}})$ and $\mathcal{P}_\beta v = \sum_{k=1}^{\text{rank}_\alpha(v)} v_k^\alpha \otimes ((P_\beta \otimes id_{D \backslash \{\alpha \cup \beta\}})w_k^{\alpha^c})$. The result follows from the definition of the α-rank. $\square$

If P_{U_α} is the orthogonal projection from $\mathcal{H}_\alpha$ onto U_α, then $P_{U_\alpha} \otimes id_{\alpha^c}$ coincides with the orthogonal projection $P_{U_\alpha \otimes \mathcal{H}_{\alpha^c}}$ from $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ onto $U_\alpha \otimes \mathcal{H}_{\alpha^c}$, and is identified with the orthogonal projection $\mathcal{P}_{U_\alpha} = P_{U_\alpha} \otimes id_{\alpha^c}$ in $\mathcal{L}(\mathcal{H})$. If $P_{U_\alpha}^{W_\alpha}$ is the oblique projection onto U_α along $W_\alpha \subset \mathcal{H}_\alpha^*$, then $\mathcal{P}_{U_\alpha}^{W_\alpha} := P_{U_\alpha}^{W_\alpha} \otimes id_{\alpha^c}$ is the oblique projection from $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ onto $U_\alpha \otimes \mathcal{H}_{\alpha^c}$ along $W_\alpha \otimes \mathcal{H}_{\alpha^c}'$. If $W_\alpha \subset \mathcal{H}_\alpha'$, then $P_{U_\alpha}^{W_\alpha}$ and $\mathcal{P}_{U_\alpha}^{W_\alpha}$ are continuous operators with equal norms $\|\mathcal{P}_{U_\alpha}^{W_\alpha}\| = \|P_{U_\alpha}^{W_\alpha}\|_{\mathcal{H}_\alpha}$.

Proposition 10 *Let U_α be a finite-dimensional subspace of $\mathcal{H}_\alpha$ and let $P_{U_\alpha}^{W_\alpha}$ be the projection onto U_α along W_α. For a tensor $v \in \mathcal{H}$ such that $W_\alpha \subset U_\alpha^{min}(v)'$, $\mathcal{P}_{U_\alpha}^{W_\alpha}v$ is an element of $U_\alpha \otimes \mathcal{H}_{\alpha^c}$ such that*

$$\|\mathcal{P}_{U_\alpha}^{W_\alpha}v\| \leq \|P_{U_\alpha}^{W_\alpha}\|_{U_\alpha^{min}(v) \to \mathcal{H}_\alpha}\|v\|,$$

and

$$\|\mathcal{P}^{W_\alpha}_{U_\alpha} v - \mathcal{P}_{U_\alpha} v\| \le \|P^{W_\alpha}_{U_\alpha} - P_{U_\alpha}\|_{U^{min}_\alpha(v) \to \mathcal{H}_\alpha} \|v\|,$$

with

$$\|P^{W_\alpha}_{U_\alpha} - P_{U_\alpha}\|_{U^{min}_\alpha(v) \to \mathcal{H}_\alpha} = \|P^{W_\alpha}_{U_\alpha}\|_{(id_\alpha - P_{U_\alpha})U^{min}_\alpha(v) \to \mathcal{H}_\alpha} \le \|P^{W_\alpha}_{U_\alpha}\|_{U^{min}_\alpha(v) \to \mathcal{H}_\alpha}.$$

Also,

$$\|v - \mathcal{P}^{W_\alpha}_{U_\alpha} v\|^2 \le (1 + \|P^{W_\alpha}_{U_\alpha} - P_{U_\alpha}\|^2_{U^{min}_\alpha(v) \to \mathcal{H}_\alpha}) \|v - \mathcal{P}_{U_\alpha} v\|^2.$$

Proof We have $v \in U^{min}_\alpha(v) \otimes \mathcal{H}_{\alpha^c}$. Noting that $\|\mathcal{P}^{W_\alpha}_{U_\alpha}\|_{U^{min}_\alpha(v) \otimes \mathcal{H}_{\alpha^c} \to \mathcal{H}} = \|P^{W_\alpha}_{U_\alpha}\|_{U^{min}_\alpha(v) \to \mathcal{H}_\alpha}$ and $\|\mathcal{P}^{W_\alpha}_{U_\alpha} - \mathcal{P}_{U_\alpha}\|_{U^{min}_\alpha(v) \otimes \mathcal{H}_{\alpha^c} \to \mathcal{H}} = \|P^{W_\alpha}_{U_\alpha} - P_{U_\alpha}\|_{U^{min}_\alpha(v) \to \mathcal{H}_\alpha}$, the results directly follow from Proposition 3. □

Now, let α be a non-empty subset of D and let $S(\alpha)$ be a partition of α. Let $P^{W_\beta}_{U_\beta}$ be oblique projections onto subspaces U_β of $\mathcal{H}_\beta$ along $W_\beta \subset \mathcal{H}^*_\beta$, $\beta \in S(\alpha)$. Then $\bigotimes_{\beta \in S(\alpha)} P^{W_\beta}_{U_\beta} := P^{W_{S(\alpha)}}_{U_{S(\alpha)}}$ is the oblique projection from $\mathcal{H}_{S(\alpha)} = \bigotimes_{\beta \in S(\alpha)} \mathcal{H}_\beta$ onto $\bigotimes_{\beta \in S(\alpha)} U_\beta := U_{S(\alpha)}$ along $\bigotimes_{\beta \in S(\alpha)} W_\beta := W_{S(\alpha)}$, and $\mathcal{P}^{W_{S(\alpha)}}_{U_{S(\alpha)}} = P^{W_{S(\alpha)}}_{U_{S(\alpha)}} \otimes id_{\alpha^c}$ is the oblique projection from $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ to $U_{S(\alpha)} \otimes \mathcal{H}_{\alpha^c}$ along $W_{S(\alpha)} \otimes \mathcal{H}'_{\alpha^c}$. From Proposition 2, we directly obtain the following result.

Proposition 11 *If $U_\alpha \subset \bigotimes_{\beta \in S(\alpha)} U_\beta$ and $W_\alpha \subset \bigotimes_{\beta \in S(\alpha)} W_\beta$, then*

$$\mathcal{P}^{W_\alpha}_{U_\alpha} \left(\prod_{\beta \in S(\alpha)} \mathcal{P}^{W_\beta}_{U_\beta} \right) = \left(\prod_{\beta \in S(\alpha)} \mathcal{P}^{W_\beta}_{U_\beta} \right) \mathcal{P}^{W_\alpha}_{U_\alpha} = \mathcal{P}^{W_\alpha}_{U_\alpha}.$$

4 Tree-based tensor formats

Let $T \subset 2^D \setminus \emptyset$ be a dimension partition tree over D, with root D. The elements of T are called the nodes of the tree. Every node $\alpha \in T$ with $\#\alpha \ge 2$ has a set of sons $S(\alpha)$ which form a partition of α, i.e. $\bigcup_{\beta \in S(\alpha)} \beta = \alpha$. A node $\alpha \in T$ with $\#\alpha = 1$ is such that $S(\alpha) = \emptyset$ and is called a leaf of the tree. The set of leaves of T is denoted $\mathcal{L}(T)$ (see an example on Fig. 1). For $\alpha \in T$, we denote by level(α) the level of α in T, such that level(D) $= 0$ and level(β) $=$ level(α) $+ 1$ if $\beta \in S(\alpha)$. We let $L = \text{depth}(T) = \max_{\alpha \in T} \text{level}(\alpha)$ be the depth of T, which is the maximum level of the nodes in T, and $T_\ell = \{\alpha \in T : \text{level}(\alpha) = \ell\}$ be the subset of nodes with level ℓ, $0 \le \ell \le L$. We let $t_\ell = \bigcup_{\alpha \in T_\ell} \alpha$. We have $t_{\ell+1} \subset t_\ell$ and $t_\ell \setminus t_{\ell+1} \subset \mathcal{L}(T)$ (see example on Fig. 2).

We introduce a subset of active nodes $A \subset T \setminus \{D\}$ such that $T \setminus A \subset \{D\} \cup \mathcal{L}(T)$, which means that the set of non active nodes in $T \setminus \{D\}$ is a subset of the leaves (see

Fig. 1 A dimension partition tree T over $D = \{1, 2, 3, 4, 5, 6\}$ and its leaves (blue nodes) (color figure online)
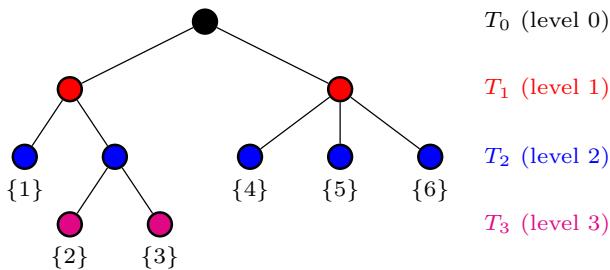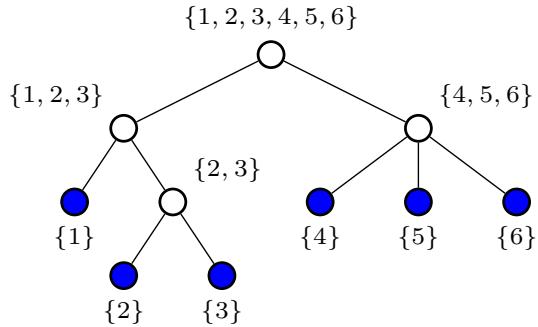

Fig. 2 A dimension partition tree T over $D = \{1, \ldots, 6\}$ with depth $L = 3$ and the corresponding subsets $T_\ell, 0 \leq \ell \leq L$. Here $t_3 = \{2, 3\}$ and $t_2 = t_1 = t_0 = D$ (color figure online)

Fig. 3 A dimension partition tree T over $D = \{1, 2, 3, 4, 5, 6\}$ and an admissible subset of active nodes A (red nodes) (color figure online)
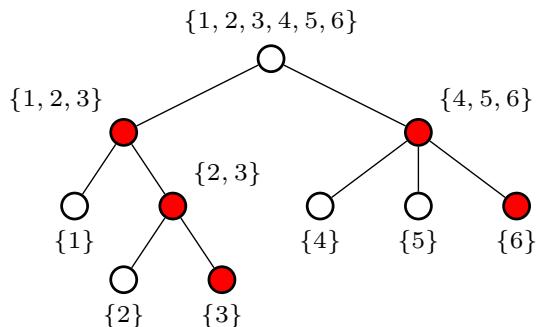
Fig. 3). A set A is admissible if for any $\alpha \in A$, the parent node of α is in $A \cup \{D\}$. We let $\mathcal{L}(A) = A \cap \mathcal{L}(T)$, $A_\ell = A \cap T_\ell$ for $1 \leq \ell \leq L$, and $a_\ell = \cup_{\alpha \in A_\ell} \alpha$. We define the A-rank of a tensor $v \in \mathcal{H}$ as the tuple $\mathrm{rank}_A(v) = \{\mathrm{rank}_\alpha(v)\}_{\alpha \in A}$.

Now we consider a tensor $v \in \mathcal{H}$ with $\mathrm{rank}_A(v) = (r_\alpha)_{\alpha \in A}$. We let $r_D = \mathrm{rank}_D(v) = 1$. For all $\alpha \in A \cup \{D\}$, we denote by $\{v_{k_\alpha}^\alpha\}_{k_\alpha=1}^{r_\alpha}$ a basis of the minimal subspace $U_\alpha^{min}(v) \subset \mathcal{H}_\alpha$, and we let $v_1^D = v$. For $\alpha \in A \cup \{D\}$ such that $\emptyset \neq S(\alpha) \subset A$, since $U_\alpha^{min}(v) \subset \bigotimes_{\beta \in S(\alpha)} U_\beta^{min}(v)$, the tensor $v_{k_\alpha}^\alpha$ admits a representation

$$v_{k_\alpha}^\alpha(x_\alpha) = \sum_{\substack{1 \le k_\beta \le r_\beta \\ \beta \in S(\alpha)}} C_{k_\alpha,(k_\beta)_{\beta \in S(\alpha)}}^\alpha \prod_{\beta \in S(\alpha)} v_{k_\beta}^\beta(x_\beta),$$

with a tensor of coefficients $C^\alpha \in \mathbb{R}^{r_\alpha \times \times_{\beta \in S(\alpha)} r_\beta}$. For $\alpha \in A \cup \{D\}$ such that $\emptyset \ne S(\alpha) \not\subset A$, we have $U_\alpha^{min}(v) \subset (\bigotimes_{\beta \in S(\alpha) \cap A} U_\beta^{min}(v)) \otimes (\bigotimes_{\beta \in S(\alpha) \setminus A} \mathcal{H}_\beta)$, and therefore the tensor $v_{k_\alpha}^\alpha$ admits a representation

$$v_{k_\alpha}^\alpha(x_\alpha) = \sum_{\substack{1 \le k_\beta \le r_\beta \\ \beta \in S(\alpha) \cap A}} C_{k_\alpha,(k_\beta)_{\beta \in S(\alpha) \cap A}}^\alpha ((x_\beta)_{\beta \in S(\alpha) \setminus A}) \prod_{\beta \in S(\alpha) \cap A} v_{k_\beta}^\beta(x_\beta),$$

with $C^\alpha \in \mathbb{R}^{r_\alpha \times \times_{\beta \in S(\alpha) \cap A} r_\beta} \otimes (\bigotimes_{\beta \in S(\alpha) \setminus A} \mathcal{H}_\beta)$. Finally, a tensor v such that $\text{rank}_A(v) = (r_\alpha)_{\alpha \in A}$ admits a representation

$$v = \sum_{\substack{1 \le k_\alpha \le r_\alpha \\ \alpha \in A \cup \{D\}}} \prod_{\alpha \in (A \cup \{D\}) \setminus \mathcal{L}(A)} C_{k_\alpha,(k_\beta)_{\beta \in S(\alpha) \cap A}}^\alpha ((x_\beta)_{\beta \in S(\alpha) \setminus A}) \prod_{\alpha \in \mathcal{L}(A)} v_{k_\alpha}^\alpha(x_\alpha) \quad (3)$$

For a tuple $r = (r_\alpha)_{\alpha \in A}$, we define the subset $\mathcal{T}_r^A(\mathcal{H})$ of tensors in $\mathcal{H}$ with A-rank bounded by r,

$$\mathcal{T}_r^A(\mathcal{H}) = \{v \in \mathcal{H} : \text{rank}_\alpha(v) \le r_\alpha, \alpha \in A\} = \bigcap_{\alpha \in A} \mathcal{T}_{r_\alpha}^{\{\alpha\}}(\mathcal{H}).$$

Remark 1 A tensor $v \in \mathcal{T}_r^A(\mathcal{H})$ admits a representation as a composition of functions. For $\alpha \in A$, let $v^\alpha(x_\alpha) = (v_1^\alpha, \ldots, v_{r_\alpha}^\alpha) \in \mathbb{R}^{r_\alpha}$. If $\emptyset \ne S(\alpha) \subset A$, the tensor C^α can be identified with a multilinear function $f^\alpha : \times_{\beta \in S(\alpha)} \mathbb{R}^{r_\beta} \to \mathbb{R}^{r_\alpha}$, and $v^\alpha(x_\alpha)$ admits the representation

$$v^\alpha(x_\alpha) = f^\alpha((v^\beta(x_\beta))_{\beta \in S(\alpha)}).$$

For $\alpha \in A \cup \{D\}$ such that $\emptyset \ne S(\alpha) \not\subset A$, the tensor $C^\alpha((x_\beta)_{\beta \in S(\alpha) \setminus A})$ can be identified with a multilinear function $f^\alpha(\cdot, (x_\beta)_{\beta \in S(\alpha) \setminus A}) : \times_{\beta \in S(\alpha) \cap A} \mathbb{R}^{r_\beta} \to \mathbb{R}^{r_\alpha}$, and $v^\alpha(x_\alpha)$ admits the representation

$$v^\alpha(x_\alpha) = f^\alpha((v^\beta(x_\beta))_{\beta \in S(\alpha) \cap A}, (x_\beta)_{\beta \in S(\alpha) \setminus A}),$$
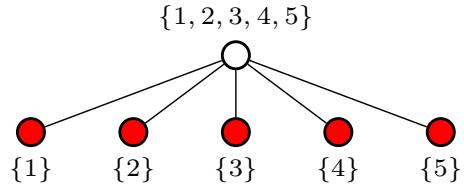
where the f^α is linear in the arguments associated with active nodes $\beta \in S(\alpha) \cap A$. As an example, for the case of Fig. 3, the tensor v admits the representation

$$v(x) = f^{1,2,3,4,5,6}(f^{1,2,3}(x_1, f^{1,2}(x_2, v^3(x_3))), f^{4,5,6}(x_4, x_5, v^6(x_6))).$$

Proposition 12 *Let $V = V_1 \otimes \cdots \otimes V_d \subset \mathcal{H}$, with V_ν a subspace of $\mathcal{H}_\nu$ with dimension $\dim(V_\nu) = n_\nu$, $1 \le \nu \le d$. The storage complexity of a tensor in $\mathcal{T}_r^A(\mathcal{H}) \cap V = \mathcal{T}_r^A(V)$ is*

Fig. 4 Tucker format. Dimension partition tree T over $D = \{1, \ldots, 5\}$ and subset of active nodes A (red nodes) (color figure online)

$$\text{storage}(\mathcal{T}_r^A(V)) = \sum_{\alpha \in (A \cup \{D\}) \backslash \mathcal{L}(A)} r_\alpha \prod_{\beta \in S(\alpha) \cap A} r_\beta \prod_{\beta \in S(\alpha) \backslash A} n_\beta + \sum_{\alpha \in \mathcal{L}(A)} r_\alpha n_\alpha.$$

Example 4 (Tucker format) The Tucker format corresponds to a trivial tree $T = \{\{1, \ldots, d\}, \{1\}, \ldots, \{d\}\}$ with depth $L = 1$, and $A = T \backslash \{D\}$ (see Fig. 4). A tensor v with A-rank bounded by $(r_1, \ldots, r_d)$ admits a representation of the form

$$v(x) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} C_{k_1, \ldots, k_d} v_{k_1}^1(x_1) \ldots v_{k_d}^d(x_d), \tag{4}$$

where $C \in \mathbb{R}^{r_1 \times \cdots \times r_d}$, and $v_{k_\nu}^\nu \in \mathcal{H}_\nu$, $1 \leq \nu \leq d$, or equivalently

$$v(x) = f^{1, \ldots, d}(v^1(x_1), \ldots, v^d(x_d)).$$

Example 5 (Degenerate Tucker format) A degenerate Tucker format corresponds to a trivial tree $T = \{\{1, \ldots, d\}, \{1\}, \ldots, \{d\}\}$ with depth $L = 1$, and an active set of nodes A strictly included in $T \backslash \{D\}$. Up to a permutation of dimensions, this corresponds to $A = \{\{1\}, \ldots, \{p\}\}$, with $p < d$. A tensor v with A-rank bounded by $(r_1, \ldots, r_p)$ admits a representation of the form

$$v(x) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_p=1}^{r_p} C_{k_1, \ldots, k_p}(x_{p+1}, \ldots, x_d) v_{k_1}^1(x_1) \ldots v_{k_p}^p(x_p), \tag{5}$$
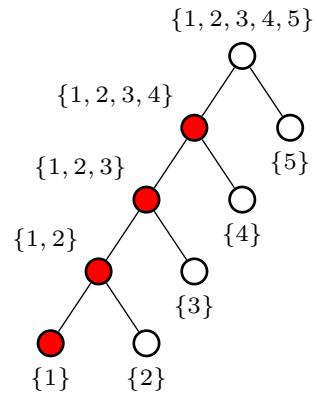
where $C \in \mathbb{R}^{r_1 \times \cdots \times r_p} \otimes \mathcal{H}_{\{p+1, \ldots, d\}}$, and $v_{k_\nu}^\nu \in \mathcal{H}_\nu$, $1 \leq \nu \leq p$, or equivalently

$$v(x) = f^{1, \ldots, d}(v^1(x_1), \ldots, v^p(x_p), x_{p+1}, \ldots, x_d).$$

Example 6 (Tensor train format) The tensor train (TT) format corresponds to a linear tree $T = \{\{1\}, \{2\}, \ldots, \{d\}, \{1, 2\}, \ldots, \{1, \ldots, d\}\}$ and $A = \{\{1\}, \{1, 2\}, \ldots, \{1, \ldots, d-1\}\}$ (see Fig. 5). Here, A is a strict subset of $T \backslash \{D\}$. The nodes $\{2\}, \ldots, \{d\}$ in T are not active.[3] A tensor v with A-rank bounded by $(r_1, \ldots, r_{d-1})$ admits a rep-

---

[3] Note that since $\text{rank}_{\{d\}}(v) = \text{rank}_{\{1, \ldots, d-1\}}(v)$, adding the node $\{d\}$ in the set of active nodes A would yield an equivalent tensor format.

Fig. 5 Tensor train format.
Dimension partition tree T over
$D = \{1, \ldots, 5\}$ and active nodes
A (red nodes) (color figure
online)

resentation of the form

$$v(x) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{d-1}} v_{k_1}^1(x_1) C_{k_1,k_2}^2(x_2) \ldots C_{k_{d-2},k_{d-1}}^{d-1}(x_{d-1}) C_{k_{d-1},1}^d(x_d),$$

where $v^1 \in \mathbb{R}^{r_1} \otimes \mathcal{H}_1$, $C^\nu \in \mathbb{R}^{r_{\nu-1} \times r_\nu} \otimes \mathcal{H}_\nu$ for $2 \le \nu \le d$, with the convention
$r_d = 1$. Here $L = d - 1$, and for $1 \le \ell \le L$, $T_\ell = \{\{1, \ldots, d - \ell\}, \{d - \ell + 1\}\}$,
$t_\ell = \{1, \ldots, d - \ell + 1\}$, $A_\ell = \{\{1, \ldots, d - \ell\}\}$ and $a_\ell = \{1, \ldots, d - \ell\}$. The tensor
v admits the equivalent representation

$$v(x) = f^{1,\ldots,d}\left(f^{1,\ldots,d-1}\left(\ldots f^{1,2}(v^1(x_1), x_2) \ldots, x_{d-1} \right), x_d \right).$$
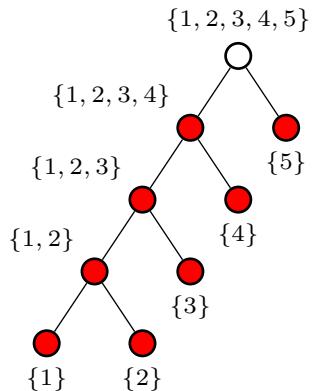
Example 7 (Tensor train Tucker format) The tensor train Tucker (TTT) format corre-
sponds to a linear tree $T = \{\{1\}, \ldots, \{d\}, \{1, 2\}, \ldots, \{1, \ldots, d\}\}$ and $A = T \backslash \{D\}$
(see Fig. 6). A tensor v having a A-rank bounded by $(r_1, \ldots, r_d, s_2, \ldots, s_{d-1})$ admits
a representation of the form (4) with a tensor $C \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ such that

$$C_{k_1,\ldots,k_d} = \sum_{i_2=1}^{s_2} \cdots \sum_{i_{d-1}=1}^{s_{d-1}} C_{k_1,k_2,i_2}^2 C_{i_2,k_3,i_3}^3 \ldots C_{i_{d-2},k_{d-1},i_{d-1}}^{d-1} C_{i_{d-1},k_d,1}^d,$$

where $C^2 \in \mathbb{R}^{r_1 \times r_2 \times s_2}$ and $C^k \in \mathbb{R}^{s_{k-1} \times r_k \times s_k}$ for $3 \le k \le d$, with the convention
$s_d = 1$. Here $L = d - 1$, $T_\ell = A_\ell = \{\{1, \ldots, d - \ell\}, \{d - \ell + 1\}\}$ and $t_\ell = a_\ell = \{1, \ldots, d - \ell + 1\}$ for $1 \le \ell \le L$. The tensor v admits the equivalent representation

$$v(x) = f^{1,\ldots,d}\left(f^{1,\ldots,d-1}\left(\ldots f^{1,2}(v^1(x_1), v^2(x_2)) \ldots, v^{d-1}(x_{d-1}) \right), v^d(x_d) \right).$$

Fig. 6 Tensor train Tucker format. Dimension partition tree T over $D = \{1, \ldots, 5\}$ and active nodes A (red nodes) (color figure online)

5 Principal component analysis for tree-based tensor format

5.1 Principal component analysis of multivariate functions

Here we introduce the notion of principal component analysis for multivariate functions. We consider a given non-empty and strict subset α of D. Any tensor in $\mathcal{H}$ is identified (through the linear isometry $\mathcal{M}_\alpha$) with its α-matricisation in $\mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$. A tensor u with α-rank $\mathrm{rank}_\alpha(u) \in \mathbb{N} \cup \{+\infty\}$ admits a *singular value decomposition* (see [20, Section 4.4.3])

$$u = \sum_{k=1}^{\mathrm{rank}_\alpha(u)} \sigma_k^\alpha u_k^\alpha \otimes u_k^{\alpha^c}, \tag{6}$$

where $\{u_k^\alpha\}_{k=1}^{\mathrm{rank}_\alpha(u)}$ and $\{u_k^{\alpha^c}\}_{k=1}^{\mathrm{rank}_\alpha(u)}$ are orthonormal vectors in $\mathcal{H}_\alpha$ and $\mathcal{H}_{\alpha^c}$ respectively, and where the σ_k^α are the α-*singular values* of u which are supposed to be arranged by decreasing values. The minimal subspace $U_\alpha^{min}(u)$ of u is given by

$$U_\alpha^{min}(u) = \overline{\mathrm{span}\{u_k^\alpha\}_{k=1}^{\mathrm{rank}_\alpha(u)}}^{\|\cdot\|_{\mathcal{H}_\alpha}}.$$

For $r_\alpha < \mathrm{rank}_\alpha(u)$, the truncated singular value decomposition

$$u_{r_\alpha} = \sum_{k=1}^{r_\alpha} \sigma_k^\alpha u_k^\alpha \otimes u_k^{\alpha^c},$$

is such that

$$\|u - u_{r_\alpha}\|^2 = \min_{\mathrm{rank}_\alpha(v) \le r_\alpha} \|u - v\|^2 = \sum_{k=r_\alpha+1}^{\mathrm{rank}_\alpha(u)} (\sigma_k^\alpha)^2.$$

The functions $\{u_k^\alpha\}_{k=1}^{r_\alpha}$ are the r_α principal components of u associated with dimensions α, hereafter called α-*principal components*. The corresponding subspace $U_\alpha^\star = \mathrm{span}\{u_k^\alpha\}_{k=1}^{r_\alpha}$, which is a subspace of $U_\alpha^{min}(u)$, is hereafter called a α-*principal subspace* of dimension r_α. Denoting $\mathcal{P}_{U_\alpha^\star} = P_{U_\alpha^\star} \otimes id_{\alpha^c}$ the orthogonal projection from $\mathcal{H}$ to $U_\alpha^\star \otimes \mathcal{H}_{\alpha^c}$, we have $u_{r_\alpha} = \mathcal{P}_{U_\alpha^\star} u$,[4] and

$$\|u - \mathcal{P}_{U_\alpha^\star} u\| = \min_{\mathrm{rank}_\alpha(v) \leq r_\alpha} \|u - v\| = \min_{\dim(U_\alpha) = r_\alpha} \|u - \mathcal{P}_{U_\alpha} u\|. \tag{7}$$

Remark 2 The optimization problem (7) over subspaces of dimension r_α in $\mathcal{H}_\alpha$ admits a unique solution $U_\alpha^\star$ if and only if $\sigma_{r_\alpha+1}^\alpha > \sigma_{r_\alpha}^\alpha$.

5.2 Principal component analysis for tree-based tensor format

Here, we propose and analyse an algorithm for the construction of an approximation $u^\star$ of a function u in tree-based format $\mathcal{T}_r^A(\mathcal{H})$. It is based on the construction of a hierarchy of subspaces U_α, $\alpha \in A$, from principal component analyses of approximations of u in low-dimensional spaces in $\mathcal{H}_\alpha$. This is a variant of the leaves-to-root higher-order singular value decomposition method proposed in [17] (see also [20, Section 11.4.2.3]).

For each leaf node $\alpha \in \mathcal{L}(T)$, we introduce a finite-dimensional approximation space $V_\alpha \subset \mathcal{H}_\alpha$ with dimension $\dim(V_\alpha) = n_\alpha$, and we let $V = \bigotimes_{\alpha \in \mathcal{L}(T)} V_\alpha \subset \mathcal{H}$. For each non active node $\alpha \in \mathcal{L}(T) \backslash A$, we let $U_\alpha = V_\alpha$. The algorithm then goes through all active nodes of the tree, going from the leaves to the root. For each $\alpha \in A$, we let

$$u_\alpha = \mathcal{P}_{V_\alpha} u,$$

where for $\alpha \notin \mathcal{L}(T)$, V_α is defined by

$$V_\alpha = \bigotimes_{\beta \in S(\alpha)} U_\beta,$$

where the U_β, $\beta \in S(\alpha)$, have been determined at a previous step. Then we determine the r_α-dimensional α-principal subspace U_α of u_α, which is solution of

$$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\| = \min_{\mathrm{rank}_\alpha(v) \leq r_\alpha} \|u_\alpha - v\|. \tag{8}$$

Finally, we define

$$u^\star = \mathcal{P}_{V_D} u, \tag{9}$$

where $\mathcal{P}_{V_D}$ is the orthogonal projection from $\mathcal{H}$ onto $V_D = \bigotimes_{\beta \in S(D)} U_\beta$.

---

[4] For all $m \geq r_\alpha$, we have $\mathcal{P}_{U_\alpha^\star} u_m = \sum_{k=1}^m \sigma_k^\alpha (P_{U_\alpha^\star} u_k^\alpha) \otimes u_k^{\alpha^c} = \sum_{k=1}^{r_\alpha} \sigma_k^\alpha u_k^\alpha \otimes u_k^{\alpha^c} = u_{r_\alpha}$. Then using the continuity of $\mathcal{P}_{U_\alpha^\star}$ and taking the limit with m, we obtain $\mathcal{P}_{U_\alpha^\star} u = u_{r_\alpha}$.

5.3 Analysis of the algorithm

Lemma 1 *For $\alpha \in \mathcal{L}(A)$, $U_\alpha \subset V_\alpha$. For $\alpha \in A \backslash \mathcal{L}(A)$,*

$$U_\alpha \subset \bigotimes_{\beta \in S(\alpha)} U_\beta.$$

Proof For $\alpha \in A$, we have $U_\alpha \subset U_\alpha^{min}(u_\alpha)$. If $\alpha \in \mathcal{L}(A)$, we have $U_\alpha^{min}(u_\alpha) \subset V_\alpha$ since $u_\alpha = \mathcal{P}_{V_\alpha} u$. If $\alpha \in A \backslash \mathcal{L}(A)$, we have $U_\alpha^{min}(u_\alpha) \subset \bigotimes_{\beta \in S(\alpha)} U_\beta^{min}(u_\alpha)$, and $U_\beta^{min}(u_\alpha) \subset U_\beta$ since $u_\alpha = \prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\beta} u$. $\qquad\square$

Proposition 13 *The approximation $u^\star$ is an element of $\mathcal{T}_r^A(\mathcal{H}) \cap V = \mathcal{T}_r^A(V)$.*

Proof Since $u^\star = \mathcal{P}_{V_D} u$, we have $u^\star \in \bigotimes_{\alpha \in S(D)} U_\alpha$. Then using Lemma 1, we prove by recursion that $u^\star \in \bigotimes_{\alpha \in \mathcal{L}(T)} V_\alpha = V$. Also, for any $\beta \in A$, Lemma 1 implies that $u^\star \in U_\beta \otimes \mathcal{H}_{\beta^c}$. Therefore, $U_\beta^{min}(u^\star) \subset U_\beta$, and $\text{rank}_\beta(u^\star) \leq \dim(U_\beta) = r_\beta$. This proves that $u^\star \in \mathcal{T}_r^A(\mathcal{H})$. $\qquad\square$

For any level ℓ, $1 \leq \ell \leq L$, let $\mathcal{P}_{T_\ell} = \prod_{\alpha \in T_\ell} \mathcal{P}_{U_\alpha}$ be the orthogonal projection from $\mathcal{H}$ onto $U_{T_\ell} \otimes \mathcal{H}_{t_\ell^c}$, with $U_{T_\ell} = \bigotimes_{\alpha \in T_\ell} U_\alpha$, and let

$$u^\ell = \mathcal{P}_{T_\ell} u^{\ell+1},$$

with the convention $u^{L+1} = u$.

Lemma 2 *For all $1 \leq \ell < \ell' \leq L$, we have*

$$\mathcal{P}_{T_{\ell'}} \mathcal{P}_{T_\ell} = \mathcal{P}_{T_\ell} = \mathcal{P}_{T_\ell} \mathcal{P}_{T_{\ell'}}.$$

Proof For $1 \leq \ell < L$, we deduce from Lemma 1 that

$$U_{T_\ell} = \bigotimes_{\alpha \in T_\ell} U_\alpha \subset \left(\bigotimes_{\beta \in T_{\ell+1}} U_\beta \right) \otimes \left(\bigotimes_{\substack{\alpha \in T_\ell \\ S(\alpha) = \emptyset}} U_\alpha \right) \subset U_{T_{\ell+1}} \otimes \mathcal{H}_{t_\ell \backslash t_{\ell+1}},$$

and then $U_{T_\ell} \otimes \mathcal{H}_{t_\ell^c} \subset U_{T_{\ell+1}} \otimes \mathcal{H}_{t_{\ell+1}^c}$. Therefore, for $1 \leq \ell < \ell' \leq L$, we have $U_{T_\ell} \otimes \mathcal{H}_{t_\ell^c} \subset U_{T_{\ell'}} \otimes \mathcal{H}_{t_{\ell'}^c}$, and the result follows from Proposition 2. $\qquad\square$

From Lemma 2, we have that

$$u^\ell = \mathcal{P}_{T_\ell} u^{\ell+1} = \mathcal{P}_{T_\ell} \ldots \mathcal{P}_{T_L} u = \mathcal{P}_{T_\ell} u,$$

for $1 \leq \ell \leq L$, and

$$u^\star = \mathcal{P}_{T_1} u = u^1.$$

We now state the two main results about the proposed algorithm.

Theorem 1 *For a given r, the approximation $u^\star \in \mathcal{T}_r^A(\mathcal{H}) \cap V$ satisfies*

$$\|u - u^\star\|^2 \le \#A \min_{v \in \mathcal{T}_r^A(\mathcal{H})} \|u - v\|^2 + \sum_{\alpha \in \mathcal{L}(T)} \|u - \mathcal{P}_{V_\alpha} u\|^2.$$

Proof We first note that for all $1 \le \ell < \ell' \le L$, $u^\ell - u^{\ell+1}$ is orthogonal to $u^{\ell'} - u^{\ell'+1}$. Indeed, using Lemma 2, we obtain that

$$
\begin{aligned}
(u^\ell - u^{\ell+1}, u^{\ell'} - u^{\ell'+1}) &= (u^\ell - u^{\ell+1}, \mathcal{P}_{T_{\ell'}} u^{\ell'+1} - \mathcal{P}_{T_{\ell'+1}} u^{\ell'+1}) \\
&= (\mathcal{P}_{T_{\ell'}}(u^\ell - u^{\ell+1}), \mathcal{P}_{T_{\ell'+1}}(\mathcal{P}_{T_{\ell'}} u^{\ell'+1} - u^{\ell'+1})) \\
&= (\mathcal{P}_{T_{\ell'+1}}\mathcal{P}_{T_{\ell'}}(u^\ell - u^{\ell+1}), \mathcal{P}_{T_{\ell'}} u^{\ell'+1} - u^{\ell'+1}) \\
&= (\mathcal{P}_{T_{\ell'}}(u^\ell - u^{\ell+1}), (\mathcal{P}_{T_{\ell'}} - id)u^{\ell'+1}) = 0.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\|u - u^\star\|^2 &= \sum_{\ell=1}^{L} \|u^{\ell+1} - u^\ell\|^2 = \sum_{\ell=1}^{L} \|u^{\ell+1} - \mathcal{P}_{T_\ell} u^{\ell+1}\|^2 \\
&\le \sum_{\ell=1}^{L} \sum_{\alpha \in T_\ell} \|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\|^2.
\end{aligned}
$$

From Lemma 2, we know that $u^{\ell+1} = \mathcal{P}_{T_{\ell+1}} u$, where we use the convention $\mathcal{P}_{T_{L+1}} = id$. For $\alpha \in \mathcal{L}(T_\ell)$, since $\mathcal{P}_{U_\alpha}$ and $\mathcal{P}_{T_{\ell+1}}$ commute, $\|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\| = \|\mathcal{P}_{T_{\ell+1}} u - \mathcal{P}_{U_\alpha}\mathcal{P}_{T_{\ell+1}} u\| = \|\mathcal{P}_{T_{\ell+1}}(u - \mathcal{P}_{U_\alpha} u)\| \le \|u - \mathcal{P}_{U_\alpha} u\|$. Therefore, for $\alpha \in \mathcal{L}(T_\ell)\backslash\mathcal{L}(A)$, we have $\|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\| \le \|u - \mathcal{P}_{V_\alpha} u\|$, and for $\alpha \in \mathcal{L}(A_\ell)$, we have $\|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\|^2 \le \|u - \mathcal{P}_{V_\alpha} u\|^2 + \|\mathcal{P}_{V_\alpha} u - \mathcal{P}_{U_\alpha} u\|^2 = \|u - \mathcal{P}_{V_\alpha} u\|^2 + \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2$.

For $\alpha \in A_\ell\backslash\mathcal{L}(A)$, we have

$$u^{\ell+1} = \mathcal{P}_{T_{\ell+1}} u = \prod_{\delta \in T_{\ell+1}\backslash S(\alpha)} \mathcal{P}_{U_\delta} \prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\beta} u = \prod_{\delta \in T_{\ell+1}\backslash S(\alpha)} \mathcal{P}_{U_\delta} u_\alpha,$$

so that

$$\|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\| = \|\prod_{\delta \in T_{\ell+1}\backslash S(\alpha)} \mathcal{P}_{U_\delta}(u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha)\| \le \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|.$$

Gathering the above results, we obtain

$$\|u - u^\star\|^2 \le \sum_{\alpha \in A} \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2 + \sum_{\alpha \in \mathcal{L}(T)} \|u - \mathcal{P}_{V_\alpha} u\|^2. \tag{10}$$

For $\alpha \in A$, we let $U_\alpha^\star$ be the subspace in $\mathcal{H}_\alpha$ such that

$$\|u - \mathcal{P}_{U_\alpha^\star} u\| = \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|u - v\| \leq \min_{\text{rank}_A(v) \leq r} \|u - v\|.$$

For $\alpha \in \mathcal{L}(A)$, we have $u_\alpha = \mathcal{P}_{V_\alpha} u$. From Proposition 9, we know that $\text{rank}_\alpha(\mathcal{P}_{V_\alpha} \mathcal{P}_{U_\alpha^\star} u) \leq \text{rank}_\alpha(\mathcal{P}_{U_\alpha^\star} u) \leq r_\alpha$. The optimality of U_α then implies that

$$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\| \leq \|\mathcal{P}_{V_\alpha} u - \mathcal{P}_{V_\alpha} \mathcal{P}_{U_\alpha^\star} u\| \leq \|u - \mathcal{P}_{U_\alpha^\star} u\|.$$

Now consider $\alpha \notin A \backslash \mathcal{L}(A)$. We know that $\text{rank}_\alpha(\prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\beta} \mathcal{P}_{U_\alpha^\star} u) \leq \text{rank}_\alpha(\mathcal{P}_{U_\alpha^\star} u)$ $\leq r_\alpha$ from Proposition 9. The optimality of U_α then implies that

$$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\| \leq \|u_\alpha - \prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\beta} \mathcal{P}_{U_\alpha^\star} u\| = \|\prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\beta}(u - \mathcal{P}_{U_\alpha^\star} u)\|$$

$$\leq \|u - \mathcal{P}_{U_\alpha^\star} u\|.$$

Finally, we obtain

$$\sum_{\alpha \in A} \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2 \leq \sum_{\alpha \in A} \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|u - v\|^2 \leq \#A \min_{\text{rank}_A(v)} \|u - v\|^2,$$

which ends the proof. $\qquad\square$

Theorem 2 *For any $\epsilon \geq 0$, if for all $\alpha \in A$, the rank r_α is chosen such that*

$$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\| \leq \frac{\epsilon}{\sqrt{\#A}} \|u_\alpha\|,$$

the approximation $u^\star$ satisfies

$$\|u - u^\star\|^2 \leq \sum_{\alpha \in \mathcal{L}(T)} \|u - \mathcal{P}_{V_\alpha} u\|^2 + \epsilon^2 \|u\|^2.$$

Proof Starting from (10), we obtain

$$\|u - u^\star\|^2 \leq \sum_{\alpha \in \mathcal{L}(T)} \|u - \mathcal{P}_{V_\alpha} u\|^2 + \sum_{\alpha \in A} \frac{\epsilon^2}{\#A} \|u_\alpha\|^2,$$

and the result follows from $\|u_\alpha\| = \|\prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\alpha} u\| \leq \|u\|$ if $\alpha \notin A \backslash \mathcal{L}(A)$, and $\|u_\alpha\| = \|\mathcal{P}_{V_\alpha} u\| \leq \|u\|$ if $\alpha \in \mathcal{L}(A)$. $\qquad\square$

6 Empirical principal component analysis for tree-based tensor format

6.1 Empirical principal component analysis of multivariate functions

Here we present the empirical principal component analysis for the statistical estimation of α-principal subspaces of a multivariate function (see Sect. 5.1). We consider that $\mathcal{H} = L^2_\mu(\mathcal{X})$ or that $\mathcal{H}$ is a separable reproducing kernel Hilbert space compactly embedded in $L^2_\mu(\mathcal{X})$, equipped with the natural norm in $L^2_\mu(\mathcal{X})$. Let (X_α, X_{α^c}) be the random vector with values in $\mathcal{X}_\alpha \times \mathcal{X}_{\alpha^c}$ with probability law $\mu_\alpha \otimes \mu_{\alpha^c}$. The tensor u can be identified with a random variable defined on $\mathcal{X}_{\alpha^c}$ with values in $\mathcal{H}_\alpha$ which associates to $x_{\alpha^c} \in \mathcal{X}_{\alpha^c}$ the function $u(\cdot, x_{\alpha^c}) = (id_\alpha \otimes \delta_{x_{\alpha^c}})u$, this random variable being an element of the Bochner space $L^2_{\mu_{\alpha^c}}(\mathcal{X}_{\alpha^c}; \mathcal{H}_\alpha)$. Then problem (7) is equivalent to find a r_α-dimensional subspace in $\mathcal{H}_\alpha$ solution of

$$\min_{\dim(U_\alpha)=r_\alpha} \mathbb{E}\left(\|u(\cdot, X_{\alpha^c}) - P_{U_\alpha} u(\cdot, X_{\alpha^c})\|^2_{\mathcal{H}_\alpha} \right). \tag{11}$$

Given a set $\{x^k_{\alpha^c}\}^{m_\alpha}_{k=1}$ of m_α samples of X_{α^c}, the α-principal subspace can be estimated by an *empirical α-principal subspace* $\widehat{U}_\alpha$ solution of

$$\|u - \mathcal{P}_{\widehat{U}_\alpha} u\|_{\alpha,m_\alpha} = \min_{\dim(U_\alpha)=r_\alpha} \|u - \mathcal{P}_{U_\alpha} u\|_{\alpha,m_\alpha}, \tag{12}$$

where

$$\|u - \mathcal{P}_{U_\alpha} u\|^2_{\alpha,m_\alpha} = \frac{1}{m_\alpha} \sum_{k=1}^{m_\alpha} \|u(\cdot, x^k_{\alpha^c}) - P_{U_\alpha} u(\cdot, x^k_{\alpha^c})\|^2_{\mathcal{H}_\alpha}.$$

The problem is equivalent to finding the r_α left principal components of $\{u(\cdot, x^k_{\alpha^c})\}^{m_\alpha}_{k=1}$, which is identified with an order-two tensor in $\mathcal{H}_\alpha \otimes \mathbb{R}^{m_\alpha}$. We note that the number of samples m_α must be such that $m_\alpha \geq r_\alpha$ in order to estimate r_α principal components. In the case of i.i.d. samples, the semi-norm $\|\cdot\|_{\alpha,m_\alpha}$ on $\mathcal{H}$ is the natural statistical estimation of the Bochner norm $\|\cdot\|_\alpha$ in $L^2_{\mu_{\alpha^c}}(\mathcal{X}_{\alpha^c}; \mathcal{H}_\alpha)$, defined by $\|v\|^2_\alpha = \mathbb{E}(\|v(X_{\alpha^c})\|_{\mathcal{H}_\alpha})$. This norm $\|\cdot\|_\alpha$ coincides with the norm $\|\cdot\|$ on $\mathcal{H}$ when $\mathcal{H}$ is equipped with the $L^2_\mu(\mathcal{X})$-norm.[5]

For some results on the comparison between $\|u - \mathcal{P}_{\widehat{U}_\alpha} u\|$ and the best approximation error $\|u - \mathcal{P}_{U^\star_\alpha} u\|$, see [3,24,25,41]. Under suitable assumptions on u (e.g., u uniformly bounded), for any $\eta > 0$ and $\epsilon > 0$, there exists a m_α sufficiently large (depending on η, ϵ, r_α and u) such that

$$\|u - \mathcal{P}_{\widehat{U}_\alpha} u\|^2 \leq \|u - \mathcal{P}_{U^\star_\alpha} u\|^2 + \epsilon^2$$

---

[5] Note that when $\mathcal{H}$ is equipped with a norm stronger than the norm in $L^2_\mu(\mathcal{X})$, then $\|\cdot\|_\alpha$ does not coincides with the norm $\|\cdot\|$ on $\mathcal{H}$, so that the subspaces solutions of (7) and (11) will be different in general.

holds with probability higher than $1 - \eta$. Then, for any $\tau > 0$, there exists a m_α sufficiently large (depending on η, τ, r_α and u) such that

$$\|u - \mathcal{P}_{\widehat{U}_\alpha} u\|^2 \leq (1 + \tau^2) \|u - \mathcal{P}_{U_\alpha^\star} u\|^2$$

holds with probability higher than $1 - \eta$.

6.2 Empirical principal component analysis for tree-based format

Now we propose a modification of the algorithm proposed in Sect. 5.2 using only evaluations of the function u at some selected points in $\mathcal{X}$. It is based on the construction of a hierarchy of subspaces $\{U_\alpha\}_{\alpha \in A}$, from empirical principal component analysis, and a corresponding hierarchy of commuting interpolation operators associated with nested sets of points.

For each leaf node $\alpha \in \mathcal{L}(T)$, we introduce a finite-dimensional approximation space $V_\alpha \subset \mathcal{H}_\alpha$ with dimension $\dim(V_\alpha) = n_\alpha$, we introduce a set Γ_{V_α} of points in $\mathcal{X}_\alpha$ which is unisolvent for V_α, we denote by I_{V_α} the corresponding interpolation operator from $\mathcal{H}_\alpha$ to V_α, and we let $\mathcal{I}_{V_\alpha} = I_{V_\alpha} \otimes id_{\alpha^c}$ be the corresponding oblique projection from $\mathcal{H}$ to $V_\alpha \otimes \mathcal{H}_{\alpha^c}$. We let $V = \bigotimes_{\alpha \in \mathcal{L}(T)} V_\alpha \subset \mathcal{H}$. For each non active $\alpha \in \mathcal{L}(T) \backslash A$, we let $U_\alpha = V_\alpha$ and $\Gamma_{U_\alpha} = \Gamma_{V_\alpha}$.

The algorithm then goes through all active nodes of the tree, going from the leaves to the root.

For each active node $\alpha \in A$, we let

$$u_\alpha = \mathcal{I}_{V_\alpha} u$$

where for $\alpha \notin \mathcal{L}(A)$, the space V_α is defined by

$$V_\alpha = \bigotimes_{\beta \in S(\alpha)} U_\beta,$$

where the U_β, $\beta \in S(\alpha)$, have been determined at a previous step. For $\alpha \notin \mathcal{L}(A)$, $\mathcal{I}_{V_\alpha} = I_{V_\alpha} \otimes id_{\alpha^c}$, where I_{V_α} is the interpolation operator onto $V_\alpha = \bigotimes_{\beta \in S(\alpha)} U_\beta$ associated with the product grid $\Gamma_{V_\alpha} = \times_{\beta \in S(\alpha)} \Gamma_{U_\beta}$, where each Γ_{V_β} have been determined at a previous step. Then we determine a r_α-dimensional empirical α-principal subspace U_α of u_α, which is solution of

$$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|_{\alpha, m_\alpha} = \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|u_\alpha - v\|_{\alpha, m_\alpha}, \tag{13}$$

where

$$\|u_\alpha - v\|_{\alpha, m_\alpha}^2 = \frac{1}{m_\alpha} \sum_{k=1}^{m_\alpha} \|u_\alpha(\cdot, x_{\alpha^c}^k) - v(\cdot, x_{\alpha^c}^k)\|_{\mathcal{H}_\alpha}^2,$$

and where $\{x_{\alpha^c}^k\}_{k=1}^{m_\alpha}$ are m_α random samples of X_{α^c}, with $m_\alpha \geq r_\alpha$. The problem is equivalent to finding the r_α left principal components of $\{u_\alpha(\cdot, x_{\alpha^c}^k)\}_{k=1}^{m_\alpha}$, which is identified with an order two tensor in $V_\alpha \otimes \mathbb{R}^{m_\alpha}$. The number of evaluations of the function u for computing U_α is $m_\alpha \times \dim(V_\alpha)$. We let $\{\varphi_k^\alpha\}_{k=1}^{r_\alpha}$ be the set of principal components, such that $U_\alpha = \mathrm{span}\{\varphi_k^\alpha\}_{k=1}^{r_\alpha}$. We then construct a set of points Γ_{U_α} which is unisolvent for U_α, and such that

$$\Gamma_{U_\alpha} \subset \Gamma_{V_\alpha}. \tag{14}$$

For the practical construction of the set Γ_{U_α}, we use the procedure described in Sect. 2.2.1. We denote by I_{U_α} the interpolation operator from $\mathcal{H}_\alpha$ onto U_α associated with the grid Γ_{U_α}, and we let $\mathcal{I}_{U_\alpha} = I_{U_\alpha} \otimes id_{\alpha^c}$ be the corresponding projection from $\mathcal{H}$ onto $U_\alpha \otimes \mathcal{H}_{\alpha^c}$.

Finally, we compute

$$u^\star = \mathcal{I}_{V_D} u, \tag{15}$$

where $\mathcal{I}_{V_D} = \bigotimes_{\beta \in S(D)} I_{U_\beta}$ is the interpolation operator from $\mathcal{H}$ onto $V_D = \bigotimes_{\beta \in S(D)} U_\beta$, associated with the product grid $\Gamma_{V_D} = \times_{\beta \in S(D)} \Gamma_{U_\beta}$.

6.3 Analysis of the algorithm

Let us first prove that the algorithm produces an approximation $u^\star$ in the desired tensor format.

Lemma 3 *For $\alpha \in \mathcal{L}(T) \backslash A$, $U_\alpha = V_\alpha$. For $\alpha \in \mathcal{L}(A)$, $U_\alpha \subset V_\alpha$. For $\alpha \in A \backslash \mathcal{L}(A)$,*

$$U_\alpha \subset \bigotimes_{\beta \in S(\alpha)} U_\beta.$$

Proof For $\alpha \in A$, we have $U_\alpha \subset U_\alpha^{min}(u_\alpha)$. If $\alpha \in \mathcal{L}(A)$, we have $U_\alpha^{min}(u_\alpha) \subset V_\alpha$ since $u_\alpha = \mathcal{I}_{V_\alpha} u$. If $\alpha \in A \backslash \mathcal{L}(A)$, we have $U_\alpha^{min}(u_\alpha) \subset \bigotimes_{\beta \in S(\alpha)} U_\beta^{min}(u_\alpha)$, and $U_\beta^{min}(u_\alpha) \subset U_\beta$ since $u_\alpha = \prod_{\beta \in S(\alpha)} \mathcal{I}_{U_\beta} u$. □

Proposition 14 *The algorithm produces an approximation*

$$u^\star \in \mathcal{T}_r^A(\mathcal{H}) \cap V = \mathcal{T}_r^A(V).$$

Proof Since $u^\star = \mathcal{I}_{V_D} u$, we have $u^\star \in V_D = \bigotimes_{\alpha \in S(D)} U_\alpha$. Then using Lemma 3, we prove by recursion that $u^\star \in \bigotimes_{\alpha \in \mathcal{L}(T)} V_\alpha = V$. Also, for any $\alpha \in A$, Lemma 3 implies that $u^\star \in U_\alpha \otimes \mathcal{H}_{\alpha^c}$. Therefore, $U_\alpha^{min}(u^\star) \subset U_\alpha$, and $\mathrm{rank}_\alpha(u^\star) \leq \dim(U_\alpha) = r_\alpha$. This proves that $u^\star \in \mathcal{T}_r^A(\mathcal{H})$. □

For all $\alpha \in T$, the operator $\mathcal{I}_{V_\alpha} = I_{V_\alpha} \otimes id_{\alpha^c}$ is a projection from $\mathcal{H}$ onto $V_\alpha \otimes \mathcal{H}_{\alpha^c}$ along $W_\alpha^\star \otimes \mathcal{H}_{\alpha^c}^\star$, with $W_\alpha^\star = \mathrm{span}\{\delta_x : x \in \Gamma_{V_\alpha}\}$. For all $\alpha \in T \backslash \{D\}$, the operator $\mathcal{I}_{U_\alpha} = I_{U_\alpha} \otimes id_{\alpha^c}$ is an oblique projection from $\mathcal{H}$ onto $U_\alpha \otimes \mathcal{H}_{\alpha^c}$ along $W_\alpha \otimes \mathcal{H}_{\alpha^c}^\star$,

with $W_\alpha = \operatorname{span}\{\delta_x : x \in \Gamma_{U_\alpha}\}$. From the property (14) of the grids, we deduce the following result.

Lemma 4 *For $\alpha \in \mathcal{L}(T) \backslash A$, $W_\alpha = W_\alpha^\star$. For $\alpha \in \mathcal{L}(A)$, $W_\alpha \subset W_\alpha^\star$. For $\alpha \in A \backslash \mathcal{L}(A)$,*

$$W_\alpha \subset W_{S(\alpha)} = \bigotimes_{\beta \in S(\alpha)} W_\beta.$$

Remark 3 Note that interpolation operators I_{U_α}, $\alpha \in A$, could be replaced by oblique projections $P_{U_\alpha}^{W_\alpha}$ onto U_α along subspaces W_α in $\mathcal{H}_\alpha^*$, with subspaces W_α satisfying for $\alpha \notin \mathcal{L}(T)$, $W_\alpha \subset \bigotimes_{\beta \in S(\alpha)} W_\beta$. Under this condition, all results of this section remain valid.

For any level ℓ, $1 \leq \ell \leq L$, let

$$\mathcal{I}_{T_\ell} = \prod_{\alpha \in T_\ell} \mathcal{I}_{U_\alpha} = I_{U_{T_\ell}} \otimes id_{t_\ell^c},$$

where $I_{U_{T_\ell}} = \bigotimes_{\alpha \in T_\ell} I_{U_\alpha}$ is the interpolation operator from $\mathcal{H}_{t_\ell}$ to $U_{T_\ell} = \bigotimes_{\alpha \in T_\ell} U_\alpha$ associated with the tensor product grid $\Gamma^{T_\ell} = \times_{\alpha \in T_\ell} \Gamma^\alpha$, and let

$$u^\ell = \mathcal{I}_{T_\ell} u^{\ell+1},$$

with the convention $u^{L+1} = u$. We then prove that operators $\mathcal{I}_{T_\ell}$, $1 \leq \ell \leq L$, are commuting oblique projections.

Lemma 5 *For all $1 \leq \ell \leq L$, the operator $\mathcal{I}_{T_\ell}$ is an oblique projection from $\mathcal{H}$ to $\mathcal{U}_\ell := U_{T_\ell} \otimes \mathcal{H}_{t_\ell^c}$ along $\mathcal{W}_\ell := W_{T_\ell} \otimes \mathcal{H}_{t_\ell^c}^*$. For all $1 \leq \ell < \ell' \leq L$, we have $\mathcal{U}_\ell \subset \mathcal{U}_{\ell'}$ and $\mathcal{W}_\ell \subset \mathcal{W}_{\ell'}$, and therefore*

$$\mathcal{I}_{T_\ell} \mathcal{I}_{T_{\ell'}} = \mathcal{I}_{T_\ell} = \mathcal{I}_{T_{\ell'}} \mathcal{I}_{T_\ell}.$$

Proof For $1 \leq \ell < L$, we have

$$\mathcal{U}_\ell = \left(\bigotimes_{\alpha \in T_\ell \backslash \mathcal{L}(T)} U_\alpha \right) \otimes \left(\bigotimes_{\alpha \in T_\ell \cap \mathcal{L}(T)} U_\alpha \right) \otimes \mathcal{H}_{t_\ell^c} \text{ and } \mathcal{U}_{\ell+1} = \left(\bigotimes_{\beta \in T_{\ell+1}} U_\beta \right) \otimes \mathcal{H}_{t_{\ell+1}^c}.$$

From Lemma 3, we know that $\bigotimes_{\alpha \in T_\ell \backslash \mathcal{L}(T)} U_\alpha$ is a subspace of $\bigotimes_{\beta \in T_{\ell+1}} U_\beta \subset \mathcal{H}_{t_{\ell+1}}$. Therefore, we obtain $\mathcal{U}_\ell \subset \mathcal{U}_{\ell+1}$. In the same way, using Lemma 4, we obtain that $\mathcal{W}_\ell \subset \mathcal{W}_{\ell+1}$. We then deduce $\mathcal{I}_{T_\ell} \mathcal{I}_{T_{\ell+1}} = \mathcal{I}_{T_{\ell+1}} \mathcal{I}_{T_\ell} = \mathcal{I}_{T_\ell}$ from Proposition 2, which ends the proof. $\square$

Lemma 6 *The approximation $u^\star$ satisfies*

$$\|u - u^\star\|^2 \leq (1 + \delta(L-1)) \sum_{\ell=1}^L \|u^{\ell+1} - u^\ell\|^2,$$

where $\delta = \max_\ell \delta_{T_\ell}$ and

$$\delta_{T_\ell} = \|I_{U_{T_\ell}} - P_{U_{T_\ell}}\|_{U_{T_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{t_\ell}},$$

with $t_\ell = \cup_{\alpha \in T_\ell} \alpha$. If $u \in V$, then

$$\delta_{T_\ell} \le \delta_{A_\ell} := \|I_{U_{A_\ell}} - P_{U_{A_\ell}}\|_{U_{A_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{a_\ell}},$$

with $a_\ell = \cup_{\alpha \in A_\ell} \alpha$.

Proof Since $u - u^\star = \sum_{\ell=1}^L (u^{\ell+1} - u^\ell)$, we have

$$\|u - u^\star\|^2 = \sum_{\ell=1}^L \|u^{\ell+1} - u^\ell\|^2 + 2 \sum_{\ell' < \ell} (u^{\ell+1} - u^\ell, u^{\ell'+1} - u^{\ell'}).$$

For $\ell' < \ell$, since $\mathcal{P}_{T_\ell}(u^{\ell'+1} - u^{\ell'}) = u^{\ell'+1} - u^{\ell'}$, we have

$$
\begin{aligned}
(u^{\ell+1} - u^\ell, u^{\ell'+1} - u^{\ell'}) &= (u^{\ell+1} - u^\ell, \mathcal{P}_{T_\ell}(u^{\ell'+1} - u^{\ell'})) \\
&= (\mathcal{P}_{T_\ell}(u^{\ell+1} - u^\ell), u^{\ell'+1} - u^{\ell'}) \\
&= (\mathcal{P}_{T_\ell} u^{\ell+1} - \mathcal{I}_{T_\ell} u^{\ell+1}, u^{\ell'+1} - u^{\ell'}) \\
&= ((\mathcal{P}_{T_\ell} - \mathcal{I}_{T_\ell})(u^{\ell+1} - u^\ell), u^{\ell'+1} - u^{\ell'}) \\
&\le \|(\mathcal{P}_{T_\ell} - \mathcal{I}_{T_\ell})(u^{\ell+1} - u^\ell)\| \|u^{\ell'+1} - u^{\ell'}\|,
\end{aligned}
$$

where we have used the fact that $\mathcal{P}_{T_\ell} \mathcal{I}_{T_\ell} = \mathcal{I}_{T_\ell}$ and $(\mathcal{P}_{T_\ell} - \mathcal{I}_{T_\ell})u^\ell = 0$. Since $\mathcal{P}_{T_\ell} - \mathcal{I}_{T_\ell} = (P_{U_{T_\ell}} - I_{U_{T_\ell}}) \otimes id_{t_\ell^c}$ and $u^{\ell+1} - u^\ell = u^{\ell+1} - \mathcal{I}_{T_\ell} u^{\ell+1} \subset U_{T_\ell}^{min}(u^{\ell+1}) \otimes \mathcal{H}_{t_\ell^c}$, we obtain from Proposition 10 that

$$|(u^{\ell+1} - u^\ell, u^{\ell'+1} - u^{\ell'})| \le \delta_{T_\ell} \|u^{\ell+1} - u^\ell\| \|u^{\ell'+1} - u^{\ell'}\|,$$

for $\ell' < \ell$. We deduce that

$$\|u - u^\star\|^2 \le \sum_{\ell,\ell'=1}^L B_{\ell,\ell'} \|u^{\ell+1} - u^\ell\| \|u^{\ell'+1} - u^{\ell'}\| \le \rho(B) \sum_{\ell=1}^L \|u^{\ell+1} - u^\ell\|^2,$$

where the matrix $B \in \mathbb{R}^{L \times L}$ is such that $B_{\ell,\ell} = 1$ and $B_{\ell,\ell'} = \delta_{T_{\max\{\ell,\ell'\}}}$ if $\ell \ne \ell'$. Using the theorem of Gerschgorin, we have that

$$\rho(B) \le 1 + \max_\ell \sum_{\ell' \ne \ell} B_{\ell,\ell'} = 1 + \max_\ell \left((\ell-1)\delta_{T_\ell} + \sum_{\ell' > \ell} \delta_{T_{\ell'}} \right) \le 1 + \delta(L-1),$$

with $\delta = \max_\ell \delta_{T_\ell}$.

Finally, when $u \in V$, we have $U_\alpha^{min}(u^{\ell+1}) \subset U_\alpha^{min}(u) \subset V_\alpha$ for all $\alpha \in \mathcal{L}(T)$. Therefore, $I_{V_\alpha} v = P_{V_\alpha} v$ for all $v \in U_\alpha^{min}(u^{\ell+1})$ and $\alpha \in \mathcal{L}(T)$, and

$$
\begin{aligned}
\delta_{T_\ell} &= \|I_{U_{A_\ell}} \otimes I_{V_{T_\ell \setminus A_\ell}} - P_{U_{A_\ell}} \otimes P_{V_{T_\ell \setminus A_\ell}}\|_{U_{T_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{t_\ell}} \\
&= \|(I_{U_{A_\ell}} - P_{U_{A_\ell}}) \otimes P_{V_{T_\ell \setminus A_\ell}}\|_{U_{T_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{t_\ell}} \\
&= \|I_{U_{A_\ell}} - P_{U_{A_\ell}}\|_{U_{A_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{a_\ell}} \|P_{V_{T_\ell \setminus A_\ell}}\|_{U_{T_\ell \setminus A_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{t_\ell \setminus a_\ell}} \\
&\leq \|I_{U_{A_\ell}} - P_{U_{A_\ell}}\|_{U_{A_\ell}^{min}(u^{\ell+1}) \to \mathcal{H}_{a_\ell}} = \delta_{A_\ell}.
\end{aligned}
$$

$\square$

Lemma 7 *For* $1 \leq \ell \leq L$,

$$
\begin{aligned}
\|u^{\ell+1} - u^\ell\|^2 \leq (1 + \delta_{T_\ell}^2)\Big(&\sum_{\alpha \in A_\ell} \Lambda_{T_{\ell+1} \setminus S(\alpha)}^2 (1 + a_\alpha) \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2 \\
&+ \sum_{\alpha \in T_\ell \cap \mathcal{L}(T)} \Lambda_{T_{\ell+1}}^2 (1 + 2 a_\alpha \delta_\alpha^2) \|u - \mathcal{P}_{V_\alpha} u\|^2\Big),
\end{aligned}
$$

where for $S \subset T$,

$$
\begin{aligned}
\Lambda_S &= \prod_{\alpha \in S} \Lambda_\alpha(U_\alpha), \quad \Lambda_\alpha(U_\alpha) = \|I_{U_\alpha}\|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha}, \\
a_\alpha &= \mathbf{1}_{\alpha \in \mathcal{L}(A)} \mathbf{1}_{\delta_\alpha \neq 0},
\end{aligned}
\tag{16}
$$

and

$$
\delta_\alpha = \|I_{V_\alpha} - P_{V_\alpha}\|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha}
\tag{17}
$$

for $\alpha \in \mathcal{L}(T)$. *Moreover, if* $u \in V$, *then* $\delta_\alpha = 0$ *for all* $\alpha \in \mathcal{L}(T)$, *and* $a_\alpha = 0$ *for all* $\alpha \in T$.

Proof For all $1 \leq \ell \leq L$, we have

$$
\begin{aligned}
\|u^{\ell+1} - u^\ell\|^2 &= \|u^{\ell+1} - \mathcal{I}_{T_\ell} u^{\ell+1}\|^2 = \|u^{\ell+1} - \mathcal{P}_{T_\ell} u^{\ell+1}\|^2 + \|\mathcal{I}_{T_\ell} u^{\ell+1} - \mathcal{P}_{T_\ell} u^{\ell+1}\|^2 \\
&= \|u^{\ell+1} - \mathcal{P}_{T_\ell} u^{\ell+1}\|^2 + \|(\mathcal{I}_{T_\ell} - \mathcal{P}_{T_\ell})(u^{\ell+1} - \mathcal{P}_{T_\ell} u^{\ell+1})\|^2 \\
&\leq (1 + \delta_{T_\ell}^2)\|u^{\ell+1} - \mathcal{P}_{T_\ell} u^{\ell+1}\|^2 \leq (1 + \delta_{T_\ell}^2) \sum_{\alpha \in T_\ell} \|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\|^2.
\end{aligned}
$$

For $\alpha \in T_\ell \setminus \mathcal{L}(T) = A_\ell \setminus \mathcal{L}(T)$,

$$
u^{\ell+1} = \mathcal{I}_{T_{\ell+1}} u = \prod_{\delta \in T_{\ell+1} \setminus S(\alpha)} \mathcal{I}_{U_\delta} \prod_{\beta \in S(\alpha)} \mathcal{I}_{U_\beta} u = \prod_{\delta \in T_{\ell+1} \setminus S(\alpha)} \mathcal{I}_{U_\delta} u_\alpha,
$$

and since $\mathcal{P}_{U_\alpha}$ and $\prod_{\delta \in T_{\ell+1} \setminus S(\alpha)} \mathcal{I}_{U_\delta}$ commute, we have

$$\|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\| = \|\prod_{\delta \in T_{\ell+1} \setminus S(\alpha)} \mathcal{I}_{U_\delta}(u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha)\| \le \Lambda_{T_{\ell+1} \setminus S(\alpha)} \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|.$$

Now for $\alpha \in T_\ell \cap \mathcal{L}(T)$, we have that $\mathcal{P}_{U_\alpha}$ and $\mathcal{I}_{T_{\ell+1}}$ commute, and therefore

$$\|u^{\ell+1} - \mathcal{P}_{U_\alpha} u^{\ell+1}\| = \|\mathcal{I}_{T_{\ell+1}}(u - \mathcal{P}_{U_\alpha} u)\| \le \Lambda_{T_{\ell+1}} \|u - \mathcal{P}_{U_\alpha} u\|.$$

If $\alpha \in T_\ell \setminus A_\ell$, we have $U_\alpha = V_\alpha$. If $\alpha \in A_\ell \cap \mathcal{L}(T)$, we have

$$\|u - \mathcal{P}_{U_\alpha} u\|^2 = \|u - \mathcal{P}_{U_\alpha} \mathcal{P}_{V_\alpha} u\|^2 = \|u - \mathcal{P}_{V_\alpha} u\|^2 + \|(id - \mathcal{P}_{U_\alpha}) \mathcal{P}_{V_\alpha} u\|^2,$$

so that if $\delta_\alpha = \|I_{V_\alpha} - P_{V_\alpha}\|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha} = 0$, we have $\mathcal{P}_{V_\alpha} u = \mathcal{I}_{V_\alpha} u = u_\alpha$ and

$$\|u - \mathcal{P}_{U_\alpha} u\|^2 \le \|u - \mathcal{P}_{V_\alpha} u\|^2 + \|(id - \mathcal{P}_{U_\alpha}) u_\alpha\|^2, \tag{18}$$

and if $\delta_\alpha \ne 0$, we have

$$
\begin{aligned}
\|u - \mathcal{P}_{U_\alpha} u\|^2 &\le \|u - \mathcal{P}_{V_\alpha} u\|^2 + 2\|(id - \mathcal{P}_{U_\alpha})(\mathcal{P}_{V_\alpha} - \mathcal{I}_{V_\alpha})u\|^2 + 2\|(id - \mathcal{P}_{U_\alpha})\mathcal{I}_{V_\alpha} u\|^2 \\
&= \|u - \mathcal{P}_{V_\alpha} u\|^2 + 2\|(id - \mathcal{P}_{U_\alpha})(\mathcal{P}_{V_\alpha} - \mathcal{I}_{V_\alpha})(u - \mathcal{P}_{V_\alpha} u)\|^2 \\
&\quad + 2\|(id - \mathcal{P}_{U_\alpha}) u_\alpha\|^2 \\
&\le (1 + 2\delta_\alpha^2)\|u - \mathcal{P}_{V_\alpha} u\|^2 + 2\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2, \tag{19}
\end{aligned}
$$

where we have used Proposition 10. We conclude from (18) and (19) that if $\alpha \in A_\ell \cap \mathcal{L}(T)$,

$$\|u - \mathcal{P}_{U_\alpha} u\|^2 \le (1 + 2a_\alpha \delta_\alpha^2)\|u - \mathcal{P}_{V_\alpha} u\|^2 + (1 + a_\alpha)\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2.$$

Gathering the above results, we obtain

$$
\begin{aligned}
\|u^{\ell+1} - u^\ell\|^2 \le \left(1 + \delta_{T_\ell}^2\right) \Big(&\sum_{\alpha \in A_\ell \setminus \mathcal{L}(T)} \Lambda_{T_{\ell+1} \setminus S(\alpha)}^2 \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2 \\
+ &\sum_{\alpha \in A_\ell \cap \mathcal{L}(T)} (1 + a_\alpha)\Lambda_{T_{\ell+1}}^2 \|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2 \\
+ &\sum_{\alpha \in A_\ell \cap \mathcal{L}(T)} (1 + 2a_\alpha \delta_\alpha^2)\Lambda_{T_{\ell+1}}^2 \|u - \mathcal{P}_{V_\alpha} u\|^2 + \sum_{\alpha \in T_\ell \setminus A_\ell} \Lambda_{T_{\ell+1}}^2 \|u - \mathcal{P}_{V_\alpha} u\|^2 \Big),
\end{aligned}
$$

which ends the proof. $\square$

We now state the two main results about the proposed algorithm.

Theorem 3 *Assume that for all $\alpha \in A$, the subspace U_α is such that*

$$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\|^2 \leq (1 + \tau^2) \min_{\mathrm{rank}_\alpha(v) \leq r_\alpha} \|u_\alpha - v\|^2 \tag{20}$$

holds with probability higher than $1 - \eta$, for some $\tau \geq 1$. Then the approximation $u^\star \in \mathcal{T}_r^A(\mathcal{H}) \cap V$ is such that

$$\|u - u^\star\|^2 \leq (1 + \tau^2) C^2 \min_{v \in \mathcal{T}_r^A(\mathcal{H})} \|u - v\|^2 + \sum_{\alpha \in \mathcal{L}(T)} D_\alpha^2 \|u - \mathcal{P}_{V_\alpha} u\|^2 \tag{21}$$

holds with probability higher than $1 - \#A\eta$, where C is defined by

$$C^2 = (1 + \delta(L - 1)) \sum_{\ell=1}^{L} (1 + \delta_{T_\ell}^2) \Lambda_{T_{\ell+1}}^2 \sum_{\alpha \in A_\ell} (1 + a_\alpha) \lambda_\alpha^2, \tag{22}$$

with

$$\lambda_\alpha = \mathbf{1}_{\alpha \notin \mathcal{L}(A)} + \mathbf{1}_{\alpha \in \mathcal{L}(A)} \|I_{V_\alpha}\|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha} \tag{23}$$

and a_α and δ_α defined by (16) and (17) respectively, and where D_α is defined by

$$D_\alpha^2 = (1 + \delta(L - 1))(1 + \delta_{T_\ell}^2) \Lambda_{T_{\ell+1}}^2 (1 + 2a_\alpha \delta_\alpha^2) \tag{24}$$

for $\alpha \in \mathcal{L}(T) \cap T_\ell$.

Proof For $\alpha \in A$, let $\widehat{U}_\alpha$ be a subspace such that

$$\|u_\alpha - \mathcal{P}_{\widehat{U}_\alpha} u_\alpha\| = \min_{\mathrm{rank}_\alpha(v) \leq r_\alpha} \|u_\alpha - v\|,$$

and let $U_\alpha^\star \subset U_\alpha^{min}(u)$ be a subspace such that

$$\|u - \mathcal{P}_{U_\alpha^\star} u\| = \min_{\mathrm{rank}_\alpha(v) \leq r_\alpha} \|u - v\| \leq \min_{v \in \mathcal{T}_r^A(\mathcal{H})} \|u - v\|.$$

For $\alpha \in \mathcal{L}(A)$, we have $u_\alpha = \mathcal{I}_{V_\alpha} u$. We know that $\mathrm{rank}_\alpha(\mathcal{I}_{V_\alpha} \mathcal{P}_{U_\alpha^\star} u) \leq r_\alpha$ from Proposition 9. By the optimality of $\widehat{U}_\alpha$, we obtain

$$\|u_\alpha - \mathcal{P}_{\widehat{U}_\alpha} u_\alpha\| \leq \|u_\alpha - \mathcal{I}_{V_\alpha} \mathcal{P}_{U_\alpha^\star} u\| \leq \|I_{V_\alpha}\|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha} \|u - \mathcal{P}_{U_\alpha^\star} u\|.$$

Now consider $\alpha \in A \backslash \mathcal{L}(A)$. We know that $\mathrm{rank}_\alpha(\prod_{\beta \in S(\alpha)} \mathcal{I}_{U_\beta} \mathcal{P}_{U_\alpha^\star} u) \leq \mathrm{rank}_\alpha(\mathcal{P}_{U_\alpha^\star} u)$ $\leq r_\alpha$ from Proposition 9. By the optimality of $\widehat{U}_\alpha$, we obtain

$$\|u_\alpha - \mathcal{P}_{\widehat{U}_\alpha} u_\alpha\| \leq \|u_\alpha - \prod_{\beta \in S(\alpha)} \mathcal{I}_{U_\beta} \mathcal{P}_{U_\alpha^\star} u\| = \|\prod_{\beta \in S(\alpha)} \mathcal{I}_{U_\beta}(u - \mathcal{P}_{U_\alpha^\star} u)\|$$

$$\leq \Lambda_{S(\alpha)} \| u - \mathcal{P}_{U_\alpha^\star} u \|.$$

Then, using Lemma 7 and assumption (20), we obtain

$$\| u^{\ell+1} - u^\ell \|^2 \leq (1 + \delta_{T_\ell}^2) \Lambda_{T_{\ell+1}}^2 \Big(\sum_{\alpha \in A_\ell} (1 + a_\alpha) \lambda_\alpha^2 (1 + \tau^2) \min_{\text{rank}_\alpha(v) \leq r_\alpha} \| u - v \|^2$$
$$+ \sum_{\alpha \in T_\ell \cap \mathcal{L}(T)} (1 + 2a_\alpha \delta_\alpha^2) \| u - \mathcal{P}_{V_\alpha} u \|^2 \Big).$$

Then, using Lemma 6, we obtain (21). $\qquad \square$

Remark 4 Assume $u \in V$ (no discretization). Then $\delta_\alpha = 0$ and $\| I_{V_\alpha} \|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha} = 1$ for all $\alpha \in \mathcal{L}(T)$, $a_\alpha = 0$ and $\lambda_\alpha = 1$ for all $\alpha \in T$, $\Lambda_{T_\ell} = \Lambda_{A_\ell}$ and $\delta_{T_\ell} = \delta_{A_\ell}$ for all ℓ. Also, the constant C defined by (22) is such that

$$C^2 = (1 + \delta(L - 1)) \sum_{\ell=1}^{L} (1 + \delta_{A_\ell}^2) \Lambda_{A_{\ell+1}}^2 \# A_\ell. \tag{25}$$

Moreover, if $U_\alpha = U_\alpha^{min}(u)$ for all α, then $\Lambda_{T_\ell} = \Lambda_{A_\ell} = 1$ and $\delta_{T_\ell} = \delta_{A_\ell} = 0$ for all ℓ, which implies

$$C^2 = \# A. \tag{26}$$

Theorem 4 *Let $\epsilon, \tilde{\epsilon} \geq 0$. Assume that for all $\alpha \in A$, the subspace U_α is such that*

$$\| u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha \| \leq \epsilon \| u_\alpha \| \tag{27}$$

holds with probability higher than $1 - \eta$, and further assume that the subspaces V_α, $\alpha \in \mathcal{L}(T)$, are such that

$$\| u - \mathcal{P}_{V_\alpha} u \| \leq \tilde{\epsilon} \| u \|. \tag{28}$$

Then the approximation $u^\star$ is such that

$$\| u - u^\star \|^2 \leq (C^2 \epsilon^2 + D^2 \tilde{\epsilon}^2) \| u \|^2$$

holds with probability higher than $1 - \# A \eta$, where C is defined by (22) and where $D^2 = \sum_{\alpha \in \mathcal{L}(T)} D_\alpha^2$, with D_α defined by (24), is such that

$$D^2 = (1 + \delta(L - 1)) \sum_{\ell=1}^{L} (1 + \delta_{T_\ell}^2) \Lambda_{T_{\ell+1}}^2 \sum_{\alpha \in T_\ell \cap \mathcal{L}(T)} (1 + 2a_\alpha \delta_\alpha^2). \tag{29}$$

Proof We first note that for $\alpha \in A \backslash \mathcal{L}(A)$, we have $\|u_\alpha\| \leq \Lambda_{S(\alpha)} \|u\|$. Also, for $\alpha \in \mathcal{L}(T)$, we have $\|u_\alpha\| \leq \lambda_\alpha \|u\|$, with λ_α defined in (23). Using Lemma 7 and assumptions (27) and (28), we then obtain

$$\|u^{\ell+1} - u^\ell\|^2 \leq (1 + \delta_{T_\ell}^2) \Lambda_{T_{\ell+1}}^2 \left(\sum_{\alpha \in A_\ell} (1 + a_\alpha) \lambda_\alpha^2 \epsilon^2 \|u\|^2 \right.$$

$$\left. + \sum_{\alpha \in T_\ell \cap \mathcal{L}(T)} (1 + 2a_\alpha \delta_\alpha^2) \tilde{\epsilon}^2 \|u\|^2 \right).$$

Finally, we obtain the desired result by using Lemma 6. □

Example 8 For the Tucker format described in Example 4, the constants C and D are given by

$$C^2 = (1 + \delta_{T_1}^2) \sum_{\alpha \in \mathcal{L}(T)} (1 + \mathbf{1}_{\delta_\alpha \neq 0}) \|I_{V_\alpha}\|_{U_\alpha^{min}(u) \to \mathcal{H}_\alpha}^2,$$

$$D^2 = (1 + \delta_{T_1}^2) \sum_{\alpha \in \mathcal{L}(T)} (1 + 2\delta_\alpha^2),$$

with

$$\delta_{T_1} = \left\| \bigotimes_{\alpha \in \mathcal{L}(T)} I_{U_\alpha} - \bigotimes_{\alpha \in \mathcal{L}(T)} P_{U_\alpha} \right\|_{U_D^{min}(u) \to \mathcal{H}} = \left\| \bigotimes_{\alpha \in \mathcal{L}(T)} I_{U_\alpha} u - \bigotimes_{\alpha \in \mathcal{L}(T)} P_{U_\alpha} u \right\| / \|u\|.$$

If $u \in V$, then

$$C = (1 + \delta_{T_1}^2)^{1/2} \sqrt{d}.$$

Example 9 For the tensor train format described in Example 6, the constant C and D are given by

$$C^2 = (1 + \delta(d-2)) \left(\sum_{\ell=1}^{d-2} \left(1 + \delta_{T_\ell}^2\right) \Lambda_{\{1,\dots,d-\ell-1\}}^2 \|I_{V_{d-\ell}}\|_{U_{d-\ell}^{min}(u) \to \mathcal{H}_{d-\ell}}^2 \right.$$

$$\left. + \left(1 + \delta_{T_{d-1}}^2\right) \left(1 + \mathbf{1}_{\delta_1 \neq 0}\right) \|I_{V_1}\|_{U_1^{min}(u) \to \mathcal{H}_1}^2 \right),$$

$$D^2 = (1 + \delta(d-2)) \left(\sum_{\ell=1}^{d-2} (1 + \delta_{T_\ell}^2) \Lambda_{\{1,\dots,d-\ell-1\}}^2 \|I_{V_{d-\ell}}\|_{U_{d-\ell}^{min}(u) \to \mathcal{H}_{d-\ell}}^2 \right.$$

$$\left. + \left(1 + \delta_{T_{d-1}}^2\right) \left(2 + 2\delta_1^2\right) \right),$$

with

$$\delta_{T_\ell} = \|I_{U_{\{1,\dots,d-\ell\}}} \otimes I_{V_{\{d-\ell+1\}}} - P_{U_{\{1,\dots,d-\ell\}}} \otimes P_{V_{\{d-\ell+1\}}}\|_{U_{\{1,\dots,d-\ell+1\}}^{min}(u^{\ell+1}) \to \mathcal{H}_{\{1,\dots,d-\ell+1\}}}.$$

If $u \in V$, then

$$C^2 = (1 + \delta(d-2)) \left(\sum_{\ell=1}^{d-2} (1 + \delta_{T_\ell}^2) \Lambda_{\{1,\dots,d-\ell-1\}}^2 + (1 + \delta_{T_{d-1}}^2) \right).$$

Example 10 For the tensor train Tucker format described in Example 7, the constant C and D are given by

$$C^2 = (1 + \delta(d-2))$$

$$\times \left(\sum_{\ell=1}^{d-2} (1 + \delta_{T_\ell}^2) \Lambda_{\{1,\dots,d-\ell-1\}}^2 \Lambda_{\{d-\ell\}}^2 (1 + (1 + \mathbf{1}_{\delta_{d-\ell+1} \neq 0}) \|I_{V_{\{d-\ell+1\}}}\|_{U_{\{d-\ell+1\}}^{min} \to \mathcal{H}_{\{d-\ell+1\}}}^2) \right.$$

$$\left. + (1 + \delta_{T_{d-1}}^2) \left((1 + \mathbf{1}_{\delta_1 \neq 0}) \|I_{V_1}\|_{U_1^{min}(u) \to \mathcal{H}_1}^2 + (1 + \mathbf{1}_{\delta_2 \neq 0}) \|I_{V_2}\|_{U_2^{min}(u) \to \mathcal{H}_2}^2 \right) \right),$$

$$D^2 = (1 + \delta(d-2)) \left(\sum_{\ell=1}^{d-2} (1 + \delta_{T_\ell}^2) \Lambda_{\{1,\dots,d-\ell-1\}}^2 \Lambda_{\{d-\ell\}}^2 (2 + 2\delta_{d-\ell+1}^2) \right.$$

$$\left. + (1 + \delta_{T_{d-1}}^2) (1 + 2\delta_1^2)(1 + 2\delta_2^2) \right).$$

If $u \in V$, then

$$C^2 = (1 + \delta(d-2)) \left(\sum_{\ell=1}^{d-2} 2(1 + \delta_{T_\ell}^2) \Lambda_{\{1,\dots,d-\ell-1\}}^2 \Lambda_{\{d-\ell\}}^2 + (1 + \delta_{T_{d-1}}^2) \right).$$

6.4 Complexity

Here we analyse the complexity of the algorithm in terms of the number of evaluations of the function. Evaluations of the function u are required (i) for the computation of the subspaces $\{U_\alpha\}_{\alpha \in A}$ through empirical principal component analysis of the V_α-valued functions $u_\alpha(\cdot, X_{\alpha^c})$, with V_α a given approximation space if $\alpha \in \mathcal{L}(A)$ or $V_\alpha = \bigotimes_{\beta \in S(\alpha)} U_\beta$ if $\alpha \in A \backslash \mathcal{L}(A)$, and (ii) for the computation of the final interpolation $\mathcal{I}_{V_D} u$.

We then obtain the following result about the number of evaluations of the function required by the algorithm

Proposition 15 *The total number of evaluations of u required by the algorithm for computing an approximation $u^\star$ in the tensor format $\mathcal{T}_r^A(V)$ is*

$$M(A, r, m, n) = \sum_{\alpha \in \mathcal{L}(A)} m_\alpha n_\alpha + \sum_{\alpha \in A \setminus \mathcal{L}(A)} m_\alpha \prod_{\beta \in S(\alpha) \cap A} r_\beta \prod_{\beta \in S(\alpha) \setminus A} n_\beta$$
$$+ \prod_{\beta \in S(D) \cap A} r_\beta \prod_{\beta \in S(D) \setminus A} n_\beta.$$

where $n = (n_\alpha)_{\alpha \in \mathcal{L}(T)}$, with $n_\alpha = \dim(V_\alpha)$, and $m = (m_\alpha)_{\alpha \in A}$, with m_α the number of samples of the Z_α-valued random variable $u_\alpha(\cdot, X_{\alpha^c})$ used for computing U_α.

Proof For $\alpha \in A$, the function u_α is an interpolation of u in $Z_\alpha = V_\alpha$ if $\alpha \in \mathcal{L}(A)$, or in $Z_\alpha = \bigotimes_{\beta \in S(\alpha)} U_\beta = \left(\bigotimes_{\beta \in S(\alpha) \cap A} U_\beta \right) \otimes \left(\bigotimes_{\beta \in S(\alpha) \setminus A} V_\beta \right)$ if $\alpha \notin \mathcal{L}(A)$. Therefore, computing $u_\alpha(\cdot, x_{\alpha^c}^k)$ for one realization $x_{\alpha^c}^k$ of X_{α^c} requires $\dim(V_\alpha) = n_\alpha$ evaluations of u if $\alpha \in \mathcal{L}(A)$ or $\dim(\bigotimes_{\beta \in S(\alpha)} U_\beta) = \prod_{\beta \in S(\alpha) \cap A} r_\beta \prod_{\beta \in S(\alpha) \setminus A} n_\beta$ if $\alpha \notin \mathcal{L}(A)$. Finally, the computation of the interpolation $\mathcal{I}_{T_1} u = \mathcal{I}_{S(D)} u$ requires $\dim(\bigotimes_{\alpha \in S(D)} U_\alpha) = \prod_{\beta \in S(D) \cap A} r_\beta \prod_{\beta \in S(D) \setminus A} n_\beta$ evaluations of u. $\qquad\square$

For computing a r_α-dimensional subspace U_α, the number of samples m_α of $u_\alpha(\cdot, X_{\alpha^c})$ has to be at least r_α.

Corollary 4 *If the number of samples $m_\alpha = r_\alpha$ for all $\alpha \in A$, then the number of evaluations of the function required by the algorithm is*

$$M(A, r, r, n) = \text{storage}(\mathcal{T}_r^A(V)).$$

The above result states that for a prescribed rank $r = (r_\alpha)_{\alpha \in A}$, the algorithm is able to construct an approximation of u using a number os samples equal to the storage complexity of the tensor format $\mathcal{T}_r^A(V)$.

When using the algorithm with a prescribed tolerance ϵ, the rank r_α is not fixed a priori but defined as the minimal integer such that the condition (27) is satisfied. Since samples of $u_\alpha(\cdot, X_{\alpha^c})$ belongs to the subspace $U_\alpha^{min}(u_\alpha) \subset Z_\alpha$ with dimension $\text{rank}_\alpha(u_\alpha) \leq \dim(Z_\alpha)$, the selected rank r_α is at most $\dim(Z_\alpha)$. Therefore, by taking $m_\alpha = \dim(Z_\alpha)$ for all $\alpha \in A$, if we assume that the set of m_α samples of $u(\cdot, X_{\alpha^c})$ contains $\text{rank}_\alpha(u_\alpha)$ linearly independent functions in Z_α, then the algorithm is able to produce an approximation with arbitrary small tolerance ϵ.

Corollary 5 *If the number of samples $m_\alpha = \dim(Z_\alpha)$ for all $\alpha \in A$, then*

$$M(A, r, m, n) = \sum_{\alpha \in \mathcal{L}(A)} n_\alpha^2 + \sum_{\alpha \in A \setminus \mathcal{L}(A)} \prod_{\beta \in S(\alpha) \cap A} r_\beta^2 \prod_{\beta \in S(\alpha) \setminus A} n_\beta^2$$
$$+ \prod_{\beta \in S(D) \cap A} r_\beta \prod_{\beta \in S(D) \setminus A} n_\beta.$$

Remark 5 For numerical experiments, when working with prescribed tolerance, we will use $m_\alpha = \dim(Z_\alpha)$ for al $\alpha \in A$.

7 Numerical examples

In all examples, we consider functions u in the tensor space $L^2_\mu(\mathcal{X})$, with $\mathcal{X} \subset \mathbb{R}^d$, equipped with the natural norm $\|\cdot\|$ (see Example 2).[6] For an approximation $u^\star$ provided by the algorithm, we estimate the relative error $\varepsilon(u^\star) = \|u - u^\star\|/\|u\|$ using Monte-Carlo integration. We denote by M the total number of evaluations of the function u required by the algorithm to provide an approximation $u^\star$, and by S the storage complexity of the approximation $u^\star$. Since the algorithm uses random evaluations of the function u (for the estimation of principal components), we run the algorithm several times and indicate confidence intervals of level 90% for $\varepsilon(u^\star)$, and also for M, S and approximation ranks when these quantities are random.

For the approximation with a prescribed A-rank, we use $m_\alpha = \gamma r_\alpha$ samples for the estimation of principal subspaces U_α, $\alpha \in A$. If $\gamma = 1$, then $M = S$ (see corollary 4).

For the approximation with a prescribed tolerance ϵ, we use $m_\alpha = \dim(Z_\alpha)$ for all $\alpha \in A$ (see corollary 5 for the estimation of M).

In all examples except the last one, we use polynomial approximation spaces $V_\nu = \mathbb{P}_p(\mathcal{X}_\nu)$ over $\mathcal{X}_\nu \subset \mathbb{R}$, $\nu \in D$, with the same polynomial degree p in all dimensions. For each $\nu \in D$, we use an orthonormal polynomial basis of $V_\nu = \mathbb{P}_p(\mathcal{X}_\nu)$ (Hermite polynomials for a Gaussian measure, Legendre polynomials for a uniform measure,...), and associated interpolation grids $\Gamma^\star_\nu$ selected in a set of 1000 random points (drawn from the measure μ_ν) by using the greedy algorithm described in Sect. 2.2.1.

7.1 Henon–Heiles potential

We consider $\mathcal{X} = \mathbb{R}^d$ equipped with the standard Gaussian measure μ and the modified Henon–Heiles potential [29]

$$u(x_1, \ldots, x_d) = \frac{1}{2} \sum_{i=1}^{d} x_i^2 + \sigma_* \sum_{i=1}^{d-1} \left(x_i x_{i+1}^2 - x_i^3 \right) + \frac{\sigma_*^2}{16} \sum_{i=1}^{d-1} \left(x_i^2 + x_{i+1}^2 \right)^2,$$

with $\sigma_\star = 0.2$. We consider approximation in the tensor train format $\mathcal{T}_r^A(V)$ described in Example 6. The function is such that $\mathrm{rank}_\alpha(u) = 3$ for all $\alpha \in A$. We use a polynomial degree $p = 4$, so that there is no discretization error, i.e. $u \in V$.

In Table 1, we observe that the algorithm with a prescribed rank $r = (3, \ldots, 3)$ is able to recover the function at very high precision with high probability with a number of samples equal to the storage complexity of the approximation (when $\gamma = 1$), with no deterioration when the dimension d increases from 5 to 100. The accuracy is slightly improved when $\gamma = 100$ but with a much higher number of evaluations of the function.

---

[6] For the last example, $\mathcal{X}$ is a finite product set equipped with the uniform measure and $L^2_\mu(\mathcal{X})$ then corresponds to the space of multidimensional arrays equipped with the canonical norm.

Table 1 Henon–Heiles potential

	$\gamma = 1$				
d	5	10	20	50	100
$\varepsilon(u^\star) \times 10^{14}$	[1.0; 234.2]	[1.5; 67.5]	[2.5; 79.9]	[6.6; 62.8]	[15.7; 175.1]
$S = M$	165	390	840	2190	4440
	$\gamma = 100$				
$\varepsilon(u^\star) \times 10^{14}$	[0.1; 0.4]	[0.2; 0.4]	[0.3; 0.4]	[0.4; 0.7]	[0.6; 0.8]
S	165	390	840	2190	4440
M	1515	3765	8265	21,765	44,265

Approximation with prescribed rank $r = (3, \ldots, 3)$ and $\gamma = 1$ and $\gamma = 100$, for different values of d

7.2 Sine of a sum

We consider $\mathcal{X} = [-1, 1]^d$ equipped with the uniform measure and the function

$$u(x_1, \ldots, x_d) = \sin(x_1 + \cdots + x_d).$$

We consider approximation in the tensor train Tucker format $\mathcal{T}_r^A(V)$ described in Example 7. The function is such that $\text{rank}_\alpha(u) = 2$ for all $\alpha \in A$. In Table 2, we observe the behavior of the algorithm with a prescribed rank $r = (2, \ldots, 2)$ for different polynomial degrees p and different values of d. We observe a linear dependence of the complexity with respect to d.

In Table 3, we observe the behavior of the algorithm with prescribed tolerance $\epsilon = 10^{-12}$ and fixed polynomial degree $p = 17$, for different values of d. For this value of ϵ, the algorithm always provides an approximation with rank $(2, \ldots, 2)$ with a fixed number of evaluations which is about ten times the storage complexity.

7.3 Sum of bivariate functions

We consider $\mathcal{X} = [-1, 1]^d$ equipped with the uniform measure and the function

$$u(x_1, \ldots, x_d) = g(x_1, x_2) + g(x_3, x_4) + \cdots + g(x_{d-1}, x_d) \tag{30}$$

where g is a bivariate function, and $d = 10$. We consider approximation in the tensor train Tucker format $\mathcal{T}_r^A(V)$ described in Example 7. The function is such that $\text{rank}_{\{v\}}(u) = \text{rank}(g) + 1$ for all $v \in D$, and $\text{rank}_{\{1,\ldots,v\}}(u) = 2$ if v is even, or $\text{rank}_{\{1,\ldots,v\}}(u) = \text{rank}(g) + 1$ if v is odd. Here, we use the algorithm we a prescribed tolerance ϵ.

We first consider the function $g(y, z) = \sum_{j=0}^3 y^j z^j$ whose rank is 4 and we use polynomial spaces of degree $p = 5$, so that there is no discretization error. We observe in Table 4 the behavior of the algorithm for decreasing values of ϵ. For $\epsilon = 10^{-4}$, the algorithm always provides the solution at almost machine precision, with an exact recovery of the rank of the function u. We observe that increasing γ (i.e. the number

Table 2 Sine of a sum

	$d = 10$		$d = 20$		$d = 50$	
	$\varepsilon(u^\star)$	M	$\varepsilon(u^\star)$	M	$\varepsilon(u^\star)$	M
$p = 3$	$[3.2; 3.3] \times 10^{-1}$	148	$[5.2; 5.3] \times 10^{-1}$	308	$[8.8; 8.81] \times 10^{-1}$	788
$p = 5$	$[1.29; 1.31] \times 10^{-2}$	188	$[2.3; 2.33] \times 10^{-2}$	388	$[5.2; 5.3] \times 10^{-2}$	988
$p = 7$	$[1.77; 1.81] \times 10^{-4}$	228	$[2.9; 3.0] \times 10^{-4}$	468	$[6.0; 6.1] \times 10^{-4}$	1188
$p = 9$	$[4.1; 4.2] \times 10^{-6}$	268	$[6.4; 6.6] \times 10^{-6}$	548	$[1.27; 1.29] \times 10^{-5}$	1388
$p = 11$	$[2.17, 2.2] \times 10^{-8}$	308	$[3.7; 3.8] \times 10^{-8}$	628	$[8.2; 8.4] \times 10^{-8}$	1588
$p = 13$	$[7.6, 7.7] \times 10^{-10}$	348	$[1.32; 1.24] \times 10^{-10}$	708	$[3.00; 3.04] \times 10^{-10}$	1788
$p = 15$	$[7.6, 7.8] \times 10^{-12}$	388	$[1.0; 1.1] \times 10^{-12}$	788	$[1.7; 2.5] \times 10^{-12}$	1988
$p = 17$	$[4.1, 13] \times 10^{-14}$	428	$[0.8; 4.9] \times 10^{-14}$	868	$[0.4; 6.7] \times 10^{-13}$	2188

Approximation with prescribed rank $r = (2, \ldots, 2)$ and $\gamma = 1$. Relative error $\varepsilon(u^\star)$ and number of evaluations $M = S$ for different values of d and p

Table 3 Sine of a sum

$d = 10$			$d = 20$			$d = 50$		
$\varepsilon(u^\star)$	S	M	$\varepsilon(u^\star)$	S	M	$\varepsilon(u^\star)$	S	M
$[3.7; 6.3] \times 10^{-13}$	428	3372	$[0.6; 1.3] \times 10^{-14}$	868	6772	$[1.4; 3.2] \times 10^{-14}$	2188	16,972

Approximation with prescribed tolerance $\epsilon = 10^{-12}$, $p = 17$ and $\gamma = 1$ for different values of d

Table 4 Sum of bivariate functions (30) with $g(y, z) = \sum_{j=0}^{3} y^j z^j$

ϵ	$\varepsilon(u^\star)$	M	S
$\gamma = 1$			
10^{-1}	$[1.4 \times 10^{-1}; 2.8 \times 10^{-1}]$	[444, 521]	[160, 192]
10^{-2}	$[0.8 \times 10^{-1}; 1.5 \times 10^{-1}]$	[918, 1034]	[345, 373]
10^{-3}	$[1.7 \times 10^{-15}; 2.6 \times 10^{-2}]$	[1916, 2088]	[530, 560]
10^{-4}	$[1.6 \times 10^{-15}; 7.8 \times 10^{-15}]$	2088	560
$\gamma = 10$			
10^{-1}	$[1.7 \times 10^{-1}; 2.0 \times 10^{-1}]$	[5364, 5484]	[202, 212]
10^{-2}	$[0.9 \times 10^{-2}; 1.1 \times 10^{-2}]$	[16,132, 16,412]	[486, 500]
10^{-3}	$[2.1 \times 10^{-15}; 2.7 \times 10^{-15}]$	20,736	560
10^{-4}	$[1.7 \times 10^{-15}; 2.7 \times 10^{-15}]$	20,736	560

Approximation with prescribed ϵ, degree $p = 5$, and different γ. Confidence intervals for relative error $\varepsilon(u^\star)$, storage complexity S and number of evaluations M

Table 5 Sum of bivariate functions (30) with $g(y, z) = \exp^{-\frac{1}{8}(y-z)^2}$

ϵ	$\varepsilon(u^\star)$	M	S
10^{-1}	$[3.8 \times 10^{-2}; 5.3 \times 10^{-2}]$	[1219, 1222]	[119, 131]
10^{-2}	$[1.8 \times 10^{-2}; 3.8 \times 10^{-2}]$	[1282, 1294]	[252, 256]
10^{-3}	$[1.2 \times 10^{-4}; 2.0 \times 10^{-3}]$	[1813, 1876]	[507, 519]
10^{-4}	$[1.2 \times 10^{-4}; 1.6 \times 10^{-4}]$	[1876, 1876]	[519, 519]
10^{-5}	$[1.6 \times 10^{-5}; 6.9 \times 10^{-5}]$	[3275, 4063]	[821, 935]
10^{-6}	$[1.8 \times 10^{-6}; 7.1 \times 10^{-6}]$	[4135, 4410]	[975, 995]
10^{-7}	$[3.1 \times 10^{-8}; 2.5 \times 10^{-6}]$	[4685, 4960]	[1015, 1035]
10^{-8}	$[2.7 \times 10^{-8}; 1.3 \times 10^{-7}]$	[5048, 6120]	[1056, 1164]
10^{-9}	$[1.2 \times 10^{-8}; 4.8 \times 10^{-8}]$	[9671, 11,595]	[1476, 1578]
10^{-10}	$[1.9 \times 10^{-10}; 1.5 \times 10^{-8}]$	[11,647, 13,117]	[1603, 1659]

Approximation with prescribed ϵ, degree $p = 10$, $\gamma = 1$. Confidence intervals for relative error $\varepsilon(u^\star)$, storage complexity S and number of evaluations M

of evaluations for the estimation of principal components) allows us to obtain a more accurate approximation for a given prescribed tolerance but with a significant increase in the number of evaluations.

We now consider the function $g(y, z) = \exp^{-\frac{1}{8}(y-z)^2}$ with infinite rank. We observe in Tables 5 and 6 the behavior of the algorithm for decreasing values of ϵ, and for a fixed polynomial degree $p = 10$ in Table 5, and an adaptive polynomial degree $p(\epsilon) = \log_{10}(\epsilon^{-1})$ in Table 6. We observe that the relative error of the obtained approximation is below the prescribed tolerance with high probability. Also, we clearly see the interest of adapting the discretization to the desired precision, which yields a lower complexity for small or moderate ϵ.

Table 6 Sum of bivariate functions (30) with $g(y, z) = \exp^{-\frac{1}{8}(y-z)^2}$

ϵ	$\varepsilon(u^\star)$	M	S
10^{-1}	$[1.4 \times 10^{-1}; 3.3 \times 10^{-1}]$	$[52, 70]$	$[32, 42]$
10^{-2}	$[2.9 \times 10^{-2}; 4.2 \times 10^{-2}]$	$[162, 184]$	$[88, 100]$
10^{-3}	$[3.2 \times 10^{-3}; 1.1 \times 10^{-2}]$	$[598, 778]$	$[258, 292]$
10^{-4}	$[1.7 \times 10^{-4}; 2.5 \times 10^{-4}]$	$[916, 916]$	$[339, 339]$
10^{-5}	$[5.7 \times 10^{-5}; 1.5 \times 10^{-4}]$	$[2056, 2759]$	$[562, 622]$
10^{-6}	$[1.1 \times 10^{-6}; 3.5 \times 10^{-5}]$	$[3190, 3465]$	$[758, 778]$
10^{-7}	$[6.9 \times 10^{-8}; 2.1 \times 10^{-7}]$	$[4390, 4390]$	$[885, 885]$
10^{-8}	$[3.2 \times 10^{-8}; 1.2 \times 10^{-7}]$	$[4560, 5319]$	$[935, 998]$
10^{-9}	$[8.3 \times 10^{-9}; 4.1 \times 10^{-8}]$	$[9415, 11{,}385]$	$[1396, 1509]$
10^{-10}	$[1.6 \times 10^{-10}; 1.7 \times 10^{-8}]$	$[11{,}647, 12{,}382]$	$[1603, 1631]$

Approximation with prescribed ϵ, degree $p(\epsilon) = \log_{10}(\epsilon^{-1})$, $\gamma = 1$. Confidence intervals for relative error $\varepsilon(u^\star)$, storage complexity S and number of evaluations M

7.4 Borehole function

We here consider the function

$$f(Y_1, \ldots, Y_8) = \frac{2\pi Y_3(Y_4 - Y_6)}{(Y_2 - \log(Y_1))\left(1 + \frac{2Y_7 Y_3}{(Y_2 - \log(Y_1))Y_1^2 Y_8} + \frac{Y_3}{Y_5}\right)}$$

which models the water flow through a borehole as a function of 8 independent random variables $Y_1 \sim \mathcal{N}(0.1, 0.0161812)$, $Y_2 \sim \mathcal{N}(7.71, 1.0056)$, $Y_3 \sim \mathcal{U}(63070, 115600)$, $Y_4 \sim \mathcal{U}(990, 1110)$, $Y_5 \sim \mathcal{U}(63.1, 116)$, $Y_6 \sim \mathcal{U}(700, 820)$, $Y_7 \sim \mathcal{U}(1120, 1680)$, $Y_8 \sim \mathcal{U}(9855, 12045)$. We then consider the function

$$u(x_1, \ldots, x_d) = f(g_1(x_1), \ldots, g_8(x_8)),$$

where g_ν are functions such that $Y_\nu = g_\nu(X_\nu)$, with $X_\nu \sim \mathcal{N}(0, 1)$ for $\nu \in \{1, 2\}$, and $X_\nu \sim \mathcal{U}(-1, 1)$ for $\nu \in \{3, \ldots, 8\}$. Function u is then defined on $\mathcal{X} = \mathbb{R}^2 \times [-1, 1]^6$. We use polynomial approximation spaces $V_\nu = \mathbb{P}_p(\mathcal{X}_\nu), \nu \in D$. We consider approximation in the tensor train Tucker format $\mathcal{T}_r^A(V)$ described in Example 7.

In Table 7, we observe the behavior of the algorithm with prescribed ranks $(r, \ldots, r)$ and fixed degree $p = 10$. We observe a very fast convergence of the approximation with the rank. Increasing γ (i.e. the number of evaluations for the estimation of principal components) allows us to improve the accuracy for a given rank but it we look at the error as a function of the complexity M, $\gamma = 1$ is much better than $\gamma = 100$.

In Table 8, we observe the behavior of the algorithm for decreasing values of ϵ, and for an adaptive polynomial degree $p(\epsilon) = \log_{10}(\epsilon^{-1})$. We observe that for all ϵ, the relative error of the obtained approximation is below ϵ with high probability.

Table 7 Borehole function

r	S	$\gamma = 1$ $\varepsilon(u^\star)$	$\gamma = 100$ $\varepsilon(u^\star)$
1	88	$[2.4 \times 10^{-2}; 2.7 \times 10^{-2}]$	$[2.3 \times 10^{-2}; 2.4 \times 10^{-2}]$
2	308	$[1.4 \times 10^{-3}; 1.4 \times 10^{-2}]$	$[4.1 \times 10^{-4}; 5.0 \times 10^{-4}]$
3	660	$[1.8 \times 10^{-5}; 4.9 \times 10^{-5}]$	$[9.9 \times 10^{-6}; 2.3 \times 10^{-5}]$
4	1144	$[2.9 \times 10^{-6}; 3.5 \times 10^{-6}]$	$[8.8 \times 10^{-7}; 1.9 \times 10^{-6}]$
5	1760	$[5.2 \times 10^{-7}; 6.1 \times 10^{-7}]$	$[1.8 \times 10^{-7}; 7.4 \times 10^{-7}]$
6	2508	$[9.0 \times 10^{-8}; 1.3 \times 10^{-7}]$	$[1.9 \times 10^{-8}; 5.2 \times 10^{-8}]$
7	3388	$[5.7 \times 10^{-8}; 9.2 \times 10^{-8}]$	$[5.1 \times 10^{-9}; 1.1 \times 10^{-8}]$
8	4400	$[1.6 \times 10^{-9}; 5.1 \times 10^{-9}]$	$[4.3 \times 10^{-10}; 2.0 \times 10^{-9}]$
9	5544	$[1.5 \times 10^{-9}; 2.4 \times 10^{-9}]$	$[3.1 \times 10^{-10}; 8.6 \times 10^{-10}]$
10	6820	$[5.5 \times 10^{-11}; 1.1 \times 10^{-10}]$	$[4.3 \times 10^{-11}; 7.6 \times 10^{-11}]$

Approximation in tensor train Tucker format with prescribed rank $(r, \ldots, r)$, fixed degree $p = 10$. Relative error $\varepsilon(u^\star)$ and storage complexity S for different values of r and γ

We note that the required number of evaluations M is about 2 to 4 times the storage complexity.

7.5 Tensorization of a univariate function

We consider the approximation of the univariate function $f : [0, 1] \to \mathbb{R}$ using tensorization of functions [26,40]. We denote by f_N the piecewise constant approximation of f on a uniform partition $0 = t_0 \le t_1 \le \cdots \le t_N = 1$ with $N = 2^d$ elements, such that $f_N(ih) = f(ih)$ for $0 \le i \le N$ and $h = N^{-1} = 2^{-d}$. We denote by $v \in \mathbb{R}^N$ the vector with components $v(i) = f(ih)$, $0 \le i \le N - 1$. The vector $v \in \mathbb{R}^{2^d}$ can be identified with an order-d tensor $u \in \mathcal{H} = \mathbb{R}^2 \otimes \cdots \otimes \mathbb{R}^2$ such that

$$u(i_1, \ldots, i_d) = v(i), \quad i = \sum_{k=1}^{d} i_k 2^{d-k},$$

where $(i_1, \ldots, i_d) \in \{0, 1\}^d = \mathcal{X}$ is the binary representation of the integer $i \in \{0, \ldots, 2^d - 1\}$. The set $\mathcal{X}$ is equipped with the uniform measure μ. Then we consider approximation of the tensor u in tensor train format. The algorithm evaluates the tensor u at some selected entries $(i_1, \ldots, i_d)$, which corresponds to evaluating the function f at some particular points t_i.

In this finite-dimensional setting, we consider $V = \mathcal{H}$. In all examples, we consider $d = 40$, and $N = 2^d \approx 10^{12}$. This corresponds to a storage complexity of one terabyte for the standard representation of f_N as a vector v of size N.

We observe in Tables 9 and 10 the behavior of the algorithm with prescribed tolerance ϵ applied to the functions $f(t) = t^2$ and $f(t) = t^{1/2}$ respectively. We indicate relative errors in ℓ^2 and ℓ^∞ norms between the tensor u and the approximation $u^\star$.

Table 8 Borehole function

ϵ	$\varepsilon(u^\star)$	M	S	$[r_{\{1\}}, \ldots, r_{\{d\}}, r_{\{1,2\}}, \ldots, r_{\{1,\ldots,d-1\}}]$
10^{-1}	$[1.8; 2.7] \times 10^{-1}$	$[39, 39]$	$[23, 23]$	$[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$
10^{-2}	$[0.3; 4.0] \times 10^{-2}$	$[88, 100]$	$[41, 46]$	$[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1]$
10^{-3}	$[0.8; 1.9] \times 10^{-3}$	$[159, 186]$	$[61, 78]$	$[2, 1, 1, 2, 2, 1, 1, 1, 1, 2, 2, 2, 1, 1]$
10^{-4}	$[2.5; 5.6] \times 10^{-5}$	$[328, 328]$	$[141, 141]$	$[2, 2, 2, 3, 3, 2, 2, 2, 1, 2, 2, 2, 2, 2]$
10^{-5}	$[0.6; 1.6] \times 10^{-5}$	$[444, 472]$	$[166, 178]$	$[2, 2, 2, 4, 4, 2, 2, 2, 1, 2, 2, 2, 2, 2]$
10^{-6}	$[3.1; 5.7] \times 10^{-6}$	$[596, 664]$	$[204, 241]$	$[3, 2, 2, 4, 5, 3, 2, 2, 2, 2, 2, 2, 2, 2]$
10^{-7}	$[1.0; 6.3] \times 10^{-7}$	$[1042, 1267]$	$[374, 429]$	$[4, 3, 4, 6, 5, 3, 3, 3, 2, 2, 3, 2, 2, 2]$
10^{-8}	$[1.1; 7.1] \times 10^{-8}$	$[1567, 1567]$	$[512, 512]$	$[4, 3, 4, 7, 6, 3, 3, 3, 2, 3, 2, 3, 3, 3]$
10^{-9}	$[0.2; 4.9] \times 10^{-8}$	$[1719, 1854]$	$[534, 560]$	$[4, 4, 4, 8, 6, 3, 3, 3, 2, 2, 3, 2, 3, 3]$
10^{-10}	$[0.3; 1.9] \times 10^{-9}$	$[2482, 2828]$	$[774, 838]$	$[5, 4, 6, 10, 7, 4, 3, 3, 2, 2, 3, 2, 3, 3, 3]$

Approximation in tensor train Tucker format with prescribed ϵ, $p(\epsilon) = \log_{10}(\epsilon^{-1})$, $\gamma = 1$. Confidence intervals for relative error $\varepsilon(u^\star)$, storage complexity S and number of evaluations M for different ϵ, and average ranks

Table 9 Tensorization of $f(t) = t^2$, $d = 40$

ϵ	$\varepsilon(u^\star)$	$\varepsilon_\infty(u^\star)$	M	S	$\max_\alpha r_\alpha$
10^{-1}	$[1.9 \times 10^{-2}; 1.2 \times 10^{-1}]$	$[2.2 \times 10^{-2}; 1.8 \times 10^{-1}]$	$[158, 194]$	$[80, 96]$	$[1, 2]$
10^{-2}	$[2.4 \times 10^{-3}; 7.7 \times 10^{-3}]$	$[3.1 \times 10^{-3}; 1.8 \times 10^{-2}]$	$[230, 250]$	$[114, 122]$	$[2, 3]$
10^{-3}	$[2.6 \times 10^{-4}; 3.1 \times 10^{-3}]$	$[3.1 \times 10^{-4}; 7.2 \times 10^{-3}]$	$[274, 326]$	$[134, 160]$	$[3, 3]$
10^{-4}	$[2.7 \times 10^{-5}; 1.2 \times 10^{-4}]$	$[4.2 \times 10^{-5}; 2.5 \times 10^{-4}]$	$[370, 394]$	$[182, 194]$	$[3, 3]$
10^{-5}	$[2.1 \times 10^{-6}; 8.9 \times 10^{-6}]$	$[2.9 \times 10^{-6}; 1.1 \times 10^{-5}]$	$[446, 470]$	$[220, 232]$	$[3, 3]$
10^{-6}	$[2.5 \times 10^{-7}; 7.8 \times 10^{-7}]$	$[3.1 \times 10^{-7}; 1.4 \times 10^{-6}]$	$[514, 546]$	$[254, 270]$	$[3, 3]$
10^{-7}	$[3.0 \times 10^{-8}; 2.4 \times 10^{-7}]$	$[4.0 \times 10^{-8}; 2.6 \times 10^{-7}]$	$[586, 614]$	$[290, 304]$	$[3, 3]$
10^{-8}	$[2.1 \times 10^{-9}; 4.8 \times 10^{-9}]$	$[3.4 \times 10^{-9}; 5.6 \times 10^{-9}]$	$[678, 690]$	$[336, 342]$	$[3, 3]$
10^{-9}	$[2.3 \times 10^{-10}; 4.8 \times 10^{-10}]$	$[2.8 \times 10^{-10}; 7.5 \times 10^{-10}]$	$[746, 766]$	$[370, 380]$	$[3, 3]$
10^{-10}	$[3.1 \times 10^{-11}; 7.5 \times 10^{-11}]$	$[3.9 \times 10^{-11}; 1.0 \times 10^{-10}]$	$[810, 842]$	$[402, 418]$	$[3, 3]$

Approximation in tensor train format with prescribed ϵ, $\gamma = 1$. Confidence intervals for relative ℓ^2-error $\varepsilon(u^\star)$, relative ℓ^∞-error $\varepsilon_\infty(u^\star)$, number of evaluations M, storage complexity S and maximal rank for different ϵ

Table 10 Tensorization of $f(t) = t^{1/2}$, $d = 40$

ϵ	$\varepsilon(u^\star)$	$\varepsilon_\infty(u^\star)$	M	S	$\max_\alpha r_\alpha$
10^{-1}	$[9.3 \times 10^{-3}; 5.5 \times 10^{-2}]$	$[4.1 \times 10^{-2}; 2.7 \times 10^{-1}]$	$[182, 230]$	$[90, 114]$	$[2, 2]$
10^{-2}	$[3.7 \times 10^{-3}; 8.6 \times 10^{-3}]$	$[2.6 \times 10^{-2}; 5.1 \times 10^{-2}]$	$[314, 350]$	$[156, 172]$	$[2, 3]$
10^{-3}	$[5.4 \times 10^{-4}; 9.2 \times 10^{-4}]$	$[3.0 \times 10^{-3}; 8.5 \times 10^{-3}]$	$[514, 606]$	$[252, 300]$	$[3, 3]$
10^{-4}	$[1.3 \times 10^{-4}; 3.3 \times 10^{-3}]$	$[7.9 \times 10^{-4}; 2.4 \times 10^{-2}]$	$[838, 962]$	$[414, 474]$	$[4, 4]$
10^{-5}	$[1.8 \times 10^{-5}; 8.2 \times 10^{-4}]$	$[1.6 \times 10^{-4}; 5.4 \times 10^{-3}]$	$[1270, 1398]$	$[626, 692]$	$[4, 5]$
10^{-6}	$[1.3 \times 10^{-6}; 6.3 \times 10^{-5}]$	$[1.2 \times 10^{-5}; 4.3 \times 10^{-4}]$	$[1900, 2036]$	$[938, 1014]$	$[5, 5]$
10^{-7}	$[4.9 \times 10^{-7}; 1.3 \times 10^{-6}]$	$[3.5 \times 10^{-6}; 1.5 \times 10^{-5}]$	$[2444, 2718]$	$[1218, 1344]$	$[5, 6]$
10^{-8}	$[1.0 \times 10^{-7}; 1.2 \times 10^{-6}]$	$[1.1 \times 10^{-6}; 1.5 \times 10^{-5}]$	$[3304, 3468]$	$[1642, 1722]$	$[6, 6]$
10^{-9}	$[2.2 \times 10^{-8}; 1.3 \times 10^{-7}]$	$[1.7 \times 10^{-7}; 1.2 \times 10^{-6}]$	$[4116, 4328]$	$[2046, 2144]$	$[7, 7]$
10^{-10}	$[8.6 \times 10^{-10}; 6.7 \times 10^{-8}]$	$[8.8 \times 10^{-9}; 4.0 \times 10^{-7}]$	$[5024, 5136]$	$[2490, 2552]$	$[7, 7]$

Approximation in tensor train format with prescribed ϵ, $\gamma = 1$. Confidence intervals for relative ℓ^2-error $\varepsilon(u^\star)$, relative ℓ^∞-error $\varepsilon_\infty(u^\star)$, number of evaluations M, storage complexity S and maximal rank for different ϵ

Let us recall that for $f(t) = t^\alpha$, the approximation error $\|f - f_N\|_{L^\infty} = O(N^{-\beta}) = O(2^{-d\beta})$ with $\beta = \min\{1, \alpha\}$, which is an exponential convergence with respect to d. For the function $f(t) = t^2$, we observe that the relative error in ℓ^2 norm is below the prescribed tolerance with high probability. For the function $f(t) = t^{1/2}$, the probability of obtaining a relative error in ℓ^2 norm below the prescribed tolerance decreases with ϵ but the ratio between the true relative error and the prescribed tolerance remains relatively small (below 100). We note that for $f(t) = t^2$, the approximation ranks are bounded by 3, which is the effective rank of f_N. For $f(t) = t^{1/2}$, the approximation ranks slowly increase with ϵ^{-1}.

In both cases, we observe a very good behavior of the algorithm, which requires a number of evaluations which scales as $\log(\epsilon^{-1})$.

References

1. Bachmayr, M., Schneider, R., Uschmajew, A.: Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. Found. Comput. Math. **16**(6), 1423–1472 (2016)
2. Ballani, J., Grasedyck, L., Kluge, M.: Black box approximation of tensors in hierarchical Tucker format. Linear Algebra Appl. **438**(2), 639–657 (2013). Tensors and Multilinear Algebra
3. Blanchard, G., Bousquet, O., Zwald, L.: Statistical properties of kernel principal component analysis. Mach. Learn. **66**(2–3), 259–294 (2007)
4. Bungartz, H.-J., Griebel, M.: Sparse grids. Acta Numer. **13**, 147–269 (2004)
5. Chevreuil, M., Lebrun, R., Nouy, A., Rai, P.: A least-squares method for sparse low rank approximation of multivariate functions. SIAM/ASA J. Uncertain. Quantif. **3**(1), 897–921 (2015)
6. Cohen, A., DeVore, R.: Approximation of high-dimensional parametric pdes. Acta Numer. **24**, 1–159 (2015)
7. Cohen, N., Sharir, O., Shashua, A.: On the expressive power of deep learning: a tensor analysis. In: Conference on Learning Theory, pp. 698–728 (2016)
8. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000)
9. de Silva, V., Lim, L.-H.: Tensor rank and ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. **30**(3), 1084–1127 (2008)
10. DeVore, R.A.: Nonlinear approximation. Acta Numer. **7**, 51–150 (1998)
11. Doostan, A., Validi, A., Iaccarino, G.: Non-intrusive low-rank separated approximation of high-dimensional stochastic models. Comput. Methods Appl. Mech. Eng. **263**, 42–55 (2013)
12. Espig, M., Grasedyck, L., Hackbusch, W.: Black box low tensor-rank approximation using fiber-crosses. Constr. Approx. **30**, 557–597 (2009)
13. Falcó, A., Hackbusch, W.: On minimal subspaces in tensor representations. Found. Comput. Math. **12**, 765–803 (2012)
14. Falco, A., Hackbusch, W., Nouy, A.: Geometric Structures in Tensor Representations (Final Release). ArXiv e-prints (2015)
15. Falcó, A., Hackbusch, W., Nouy, A.: On the Dirac–Frenkel variational principle on tensor Banach spaces. Found. Comput. Math. (2018). https://doi.org/10.1007/s10208-018-9381-4
16. Falcó, A., Hackbusch, W., Nouy, A.: Tree-based tensor formats. SeMA J. (2018). https://doi.org/10.1007/s40324-018-0177-x
17. Grasedyck, L.: Hierarchical singular value decomposition of tensors. SIAM J. Matrix Anal. Appl. **31**, 2029–2054 (2010)
18. Grasedyck, L., Kressner, D., Tobler, C.: A literature survey of low-rank tensor approximation techniques. GAMM-Mitteilungen **36**(1), 53–78 (2013)
19. Grelier, E., Nouy, A., Chevreuil, M.: Learning with tree-based tensor formats (2018). arXiv e-prints arXiv:1811.04455
20. Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus, Volume 42 of Springer Series in Computational Mathematics. Springer, Heidelberg (2012)

21. Hackbusch, W., Kuhn, S.: A new scheme for the tensor representation. J. Fourier Anal. Appl. **15**(5), 706–722 (2009)
22. Hillar, C., Lim, L.-H.: Most tensor problems are np-hard. J. ACM (JACM) **60**(6), 45 (2013)
23. Holtz, S., Rohwedder, T., Schneider, R.: On manifolds of tensors of fixed tt-rank. Numer. Math. **120**(4), 701–731 (2012)
24. Jirak, M., Wahl, M.: A tight sin Θ theorem for empirical covariance operators. ArXiv e-prints (2018)
25. Jirak, M., Wahl, M.: Relative perturbation bounds with applications to empirical covariance operators. ArXiv e-prints (2018)
26. Khoromskij, B.: O (dlog n)-quantics approximation of nd tensors in high-dimensional numerical modeling. Constr. Approx. **34**(2), 257–280 (2011)
27. Khoromskij, B.: Tensors-structured numerical methods in scientific computing: survey on recent advances. Chemometr. Intell. Lab. Syst. **110**(1), 1–19 (2012)
28. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
29. Kressner, D., Steinlechner, M., Uschmajew, A.: Low-rank tensor methods with subspace correction for symmetric eigenvalue problems. SIAM J. Sci. Comput. **36**(5), A2346–A2368 (2014)
30. Lubich, C., Rohwedder, T., Schneider, R., Vandereycken, B.: Dynamical approximation by hierarchical tucker and tensor-train tensors. SIAM J. Matrix Anal. Appl. **34**(2), 470–494 (2013)
31. Luu, T.H., Maday, Y., Guillo, M., Guérin, P.: A new method for reconstruction of cross-sections using Tucker decomposition. J. Comput. Phys. **345**, 189–206 (2017)
32. Maday, Y., Nguyen, N.C., Patera, A.T., Pau, G.S.H.: A general multipurpose interpolation procedure: the magic points. Commun. Pure Appl. Anal. **8**(1), 383–404 (2009)
33. Megginson, R.E: An Introduction to Banach Space Theory, Vol. 183. Springer, Berlin (2012)
34. Nouy, A.: Low-rank methods for high-dimensional approximation and model order reduction. In: Benner, P., Cohen, A., Ohlberger, M., Willcox, K. (eds.) Model Reduction and Approximation: Theory and Algorithms. SIAM, Philadelphia (2017)
35. Nouy, A.: Low-Rank Tensor Methods for Model Order Reduction, pp. 857–882. Springer, Cham (2017)
36. Orus, R.: A practical introduction to tensor networks: matrix product states and projected entangled pair states. Ann. Phys. **349**, 117–158 (2014)
37. Oseledets, I.: Tensor-train decomposition. SIAM J. Sci. Comput. **33**(5), 2295–2317 (2011)
38. Oseledets, I., Tyrtyshnikov, E.: Breaking the curse of dimensionality, or how to use svd in many dimensions. SIAM J. Sci. Comput. **31**(5), 3744–3759 (2009)
39. Oseledets, I., Tyrtyshnikov, E.: TT-cross approximation for multidimensional arrays. Linear Algebra Appl. **432**(1), 70–88 (2010)
40. Oseledets, I., Tyrtyshnikov, E.: Algebraic wavelet transform via quantics tensor train decomposition. SIAM J. Sci. Comput. **33**(3), 1315–1328 (2011)
41. Reiß, M., Wahl, M.: Non-asymptotic upper bounds for the reconstruction error of PCA (2016). arXiv preprint arXiv:1609.03779
42. Schneider, R., Uschmajew, A.: Approximation rates for the hierarchical tensor format in periodic Sobolev spaces. J. Complex. **30**(2), 56–71 (2014). Dagstuhl 2012
43. Temlyakov, V.: Nonlinear methods of approximation. Found. Comput. Math. **3**(1), 33–107 (2003)
44. Temlyakov, V.: Greedy Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge (2011)
45. Uschmajew, A., Vandereycken, B.: The geometry of algorithms using hierarchical tensors. Linear Algebra Appl. **439**(1), 133–166 (2013)