Principal component analysis for the approximation of high-dimensional functions in tree-based tensor formats

Anthony Nouy

Ecole Centrale de Nantes, LMJL

Supported by the ANR (CHORUS project)

Motivation

We consider the problem of constructing an approximation of a random variable Y by a function of a set of random variables $X = (X_1, \ldots, X_d)$, using samples of (X, Y), when

- the samples of (X, Y) can be generated from adaptively chosen samples of X (active learning),
- there exists a deterministic function u such that

$$Y = u(X).$$

Motivation

In practice, Y could be the output of a numerical model (computer code) and X a set of input parameters modelling uncertainties on the model, with known probability distribution.

The approximation can then be used as a predictive surrogate model.

When the generation of one sample requires a costly numerical simulation (or experiment), only a few samples are available.

Low-dimensional structures of functions have to be exploited (e.g., low effective dimensionality, anisotropy, sparsity, low rank).

Outline

1. Tree-based tensor formats
2. Principal component analysis in tree-based tensor formats

Outline

1. Tree-based tensor formats
2. Principal component analysis in tree-based tensor formats

Tensor spaces of multivariate functions

Assume X_ν has probability law μ_ν with support $\mathcal{X}_\nu$.

Let $\mathcal{H}_\nu$ a Hilbert space of functions defined on $\mathcal{X}_\nu$, typically $L^2_{\mu_\nu}(\mathcal{X}_\nu)$ or a reproducing kernel Hilbert space (RKHS) in $L^2_{\mu_\nu}(\mathcal{X}_\nu)$.

We consider multivariate functions defined on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ that are elements of the tensor Hilbert space

$$\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_d := \mathcal{H}$$

equipped with the canonical norm.

α-rank of higher-order tensors

For a non-empty subset α of $D = \{1, \dots, d\}$ and $\alpha^c = D \setminus \alpha$, a tensor $u \in \mathcal{H}$ can be identified with an order-two tensor

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

where

$$\mathcal{H}_\beta = \bigotimes_{\nu \in \beta} \mathcal{H}_\nu \subset \mathbb{R}^{\mathcal{X}_\beta}.$$

The α-rank of u is defined by

$$\mathrm{rank}_\alpha(u) = \mathrm{rank}(\mathcal{M}_\alpha(u)),$$

which is the minimal integer r_α such that

$$u(x) = \sum_{k=1}^{r_\alpha} v_k^\alpha(x_\alpha) w_k^{\alpha^c}(x_{\alpha^c})$$

α-ranks and related tensor formats

- For T a collection of subsets of D, we define the T-rank of a tensor v as the tuple

$$\text{rank}_T(v) = \{\text{rank}_\alpha(v)\}_{\alpha \in T}.$$

α-ranks and related tensor formats

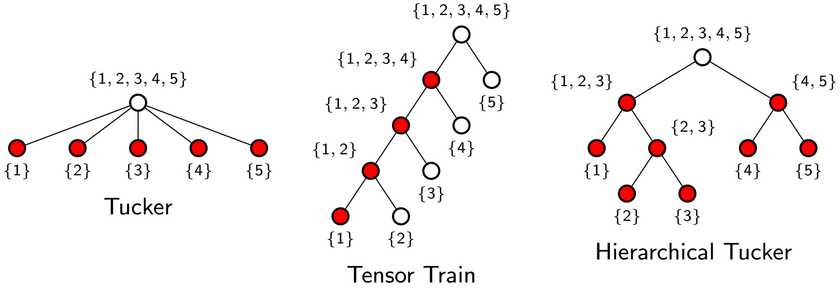
- For T a collection of subsets of D, we define the T-rank of a tensor v as the tuple
$$\mathrm{rank}_T(v) = \{\mathrm{rank}_\alpha(v)\}_{\alpha \in T}.$$

- The set of tensors with T-rank bounded by $r = (r_\alpha)_{\alpha \in T}$ is
$$\mathcal{T}_r^T = \{v \in \mathcal{H} : \mathrm{rank}_T(v) \leq r\} = \bigcap_{\alpha \in T} \{v \in \mathcal{H} : \mathrm{rank}_\alpha(v) \leq r_\alpha\},$$
and is called a tensor format.

α-ranks and related tensor formats

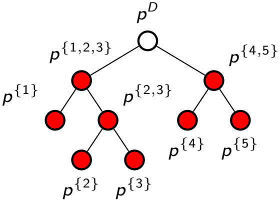
- For T a collection of subsets of D, we define the T-rank of a tensor v as the tuple
$$\text{rank}_T(v) = \{\text{rank}_\alpha(v)\}_{\alpha\in T}.$$

- The set of tensors with T-rank bounded by $r = (r_\alpha)_{\alpha\in T}$ is
$$\mathcal{T}_r^T = \{v \in \mathcal{H} : \text{rank}_T(v) \le r\} = \bigcap_{\alpha\in T} \{v \in \mathcal{H} : \text{rank}_\alpha(v) \le r_\alpha\},$$
and is called a tensor format.
- Tree-based tensor formats correspond to a tree-structured T



Tree-based tensor formats

- A tensor in $\mathcal{T}_r^T$ admits a multilinear parametrization with parameters $\{p^\alpha\}_{\alpha \in T \cup \{D\}}$ forming a tree network of low order tensors.



- Storage complexity scales as $O(dR^{s+1})$ where R is the maximal α-rank and s is the arity of the tree.

- Corresponds to a deep (sum-product) network with depth bounded by $d-1$.

- $\mathcal{T}_r^T$ is weakly closed (and therefore proximinal).

Outline

Principal subspaces

For a subset of variables α, a multivariate function $u(x_1, \dots, x_d)$ is identified with a bivariate function $u \in \mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ which admits a singular value decomposition

$$u(x_\alpha, x_{\alpha^c}) = \sum_{k=1}^{\text{rank}_\alpha(u)} \sigma_k^\alpha v_k^\alpha(x_\alpha) v_k^{\alpha^c}(x_{\alpha^c})$$

The subspace of α-principal components

$$U_\alpha = span\{v_1^\alpha, \dots, v_{r_\alpha}^\alpha\}$$

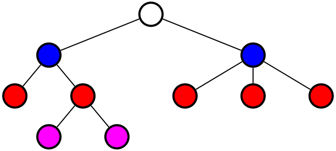
is solution of

$$\min_{\dim(U_\alpha)=r_\alpha} \|u - \mathcal{P}_{U_\alpha} u\| = \min_{\text{rank}_\alpha(v) \le r_\alpha} \|u - v\|$$

where $\mathcal{P}_{U_\alpha} = P_{U_\alpha} \otimes id_{\alpha^c}$ is the orthogonal projection onto $U_\alpha \otimes \mathcal{H}_{\alpha^c}$.

Higher-order principal component analysis for tree-based formats

Let T be a tree-structured collection of subsets of 2^D



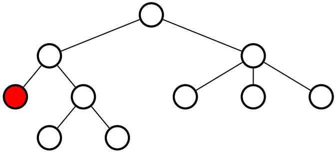
For all α in T, we will determine subspaces U_α that are approximations of α-principal subspaces of u in low-dimensional subspaces V_α of functions defined on $\mathcal{X}_\alpha$.

Higher-order principal component analysis for tree-based formats

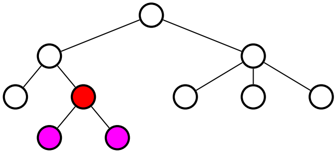
For each $\alpha \in T$, U_α is defined as the r_α-dimensional α-principal subspace of an approximation of u

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

- for $S(\alpha) = \emptyset$ (leaf node), V_α is a given approximation space in $\mathcal{H}_\alpha$ (e.g., polynomials, wavelets, ...),



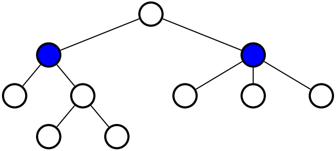
- for $S(\alpha) \neq \emptyset$ (interior node), $V_\alpha = \bigotimes_{\beta \in S(\alpha)} U_\beta$.



Higher-order principal component analysis for tree-based formats

We finally obtain an approximation $u^\star$ by projecting u onto the tensor space $\bigotimes_{\alpha \in S(D)} U_\alpha$

$$u^\star = \prod_{\alpha \in S(D)} \mathcal{P}_{U_\alpha} u$$



Learning algorithm based on principal component analysis

For a feasible algorithm using samples,

1. orthogonal projections $\mathcal{P}_{W_\alpha}$ on subspaces W_α (U_α or V_α) are replaced by oblique projections $\mathcal{I}_{W_\alpha}$ using samples (e.g. interpolation or least-squares projection),

2. principal subspaces U_α of $u_\alpha = \mathcal{I}_{V_\alpha} u$ are estimated using samples of the V_α-valued random variable

$$u_\alpha(\cdot, X_{\alpha^c})$$

With interpolation, this requires the evaluation of u at the $dim(V_\alpha) \times N_\alpha$ points

$$\{(x_\alpha, x_{\alpha^c}^k) : x_\alpha \in \Gamma_{V_\alpha}, 1 \leq k \leq N_\alpha\}$$

where $\Gamma_{V_\alpha} \subset \mathcal{X}_\alpha$ is a unisolvent set of points for V_α (magic points), and the $x_{\alpha^c}^k$ are i.i.d. samples of X_{α^c}.

Properties of the algorithm

Theorem (Prescribed rank)

For a given T-rank, if the subspaces U_α are such that

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

holds with probability higher than $1-\eta$, then we obtain an approximation $u^\star$ such that

$$\|u^\star - u\|^2 \leq \Lambda^2 C^2 \# T \min_{v \in \mathcal{T}_r^T} \|v - u\|^2 + \tilde{\Lambda}^2 \max_{1 \leq \nu \leq d} \|u - \mathcal{P}_{V_\nu} u\|^2$$

holds with probability higher than $1 - \eta \# T$, with Λ and $\tilde{\Lambda}$ depending on the properties of the oblique projection operators.

About complexity: If $N_\alpha = r_\alpha$ for all $\alpha \in T$, then the total number of evaluations N is equal to the storage complexity S of the resulting approximation $u^\star \in \mathcal{T}_r^T$.

Properties of the algorithm

Theorem (Fixed precision)

Let $\epsilon, \tilde{\epsilon} \geq 0$. If the subspaces U_α are determined such that

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

holds with probability higher than $1 - \eta$, and if the approximation spaces V_ν, $1 \leq \nu \leq d$, are such that

$$\|\mathcal{P}_{V_\nu} u - u\| \leq \tilde{\epsilon}\|u\|,$$

then we obtain an approximation $u^\star$ such that

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

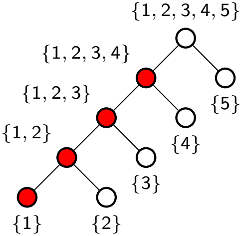
holds with probability higher than $1 - \eta \#T$, with Λ and $\tilde{\Lambda}$ depending on the properties of the oblique projection operators.

Illustration of tensor recovery: Henon-Heiles potential

$$u(X) = \frac{1}{2}\sum_{i=1}^{d} X_i^2 + 0.2\sum_{i=1}^{d-1}(X_i X_{i+1}^2 - X_i^3) + \frac{0.2^2}{16}\sum_{i=1}^{d-1}(X_i^2 + X_{i+1}^2)^2, \quad X_i \sim U(-1,1),$$

$\text{rank}_\alpha(u) = 3$ for all α in

$T = \{\{1\},\{1,2\},\ldots,\{1,\ldots,d-1\}\}$



Then u can be exactly represented in the tensor train format $\mathcal{T}_r^T$ with T-rank $r = (3,\ldots,3)$

$$u = \sum_{k_1=1}^{3}\sum_{k_2=1}^{3}\ldots\sum_{k_{d-1}=1}^{3} v_{1,k_1}^{(1)}(x_1) v_{k_1,k_2}^{(1,2)}(x_2) v_{k_2,k_3}^{(1,2,3)}(x_3)\ldots v_{k_{d-1},1}^{(1,\ldots,d)}(x_d)$$

with univariate polynomial functions of degree 4.

Illustration of tensor recovery: Henon-Heiles potential

Table: Approximation with prescribed T-rank $r = (3, \dots, 3)$ and polynomial degree 4 for different values of d and $\gamma = N_\alpha / r_\alpha$.

$\gamma = 1$					
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]
$S = N$	165	390	840	2190	4440

$\gamma = 10$					
d	5	10	20	50	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
N	1515	3765	8265	21765	44265

Illustration for approximation: Borehole function

The Borehole function models water flow through a borehole:

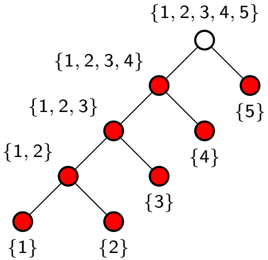
$$u(X) = \frac{2\pi T_u(H_u - H_l)}{\ln(r/r_w)\left(1 + \frac{2LT_u}{\ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right)}, \quad X = (r_w, \log(r), T_u, H_u, T_l, H_l, L, K_w)$$

r_w	radius of borehole (m)	$N(\mu = 0.10, \sigma = 0.0161812)$
r	radius of influence (m)	$LN(\mu = 7.71, \sigma = 1.0056)$
T_u	transmissivity of upper aquifer (m^2/yr)	$U(63070, 115600)$
H_u	potentiometric head of upper aquifer (m)	$U(990, 1110)$
T_l	transmissivity of lower aquifer (m^2/yr)	$U(63.1, 116)$
H_l	potentiometric head of lower aquifer (m)	$U(700, 820)$
L	length of borehole (m)	$U(1120, 1680)$
K_w	hydraulic conductivity of borehole (m/yr)	$U(9855, 12045)$

Illustration for approximation: Borehole function

Approximation in hierarchical Tucker format with a linearly structured tree:

$$T = \{\{1\}, \dots, \{d\}, \{1,2\}, \dots, \{1, \dots, d-1\}\}$$



$$u^\star = \sum_{i_1=1}^{r_1} \dots \sum_{i_d=1}^{r_d} \sum_{k_2=1}^{r_{1,2}} \dots \sum_{k_{d-1}=1}^{r_{1,\dots,d-1}} v_{i_1}^{(1)}(x_1) \dots v_{i_d}^{(d)}(x_d) C^{(1,2)}_{i_1,i_2,k_2} C^{(1,2,3)}_{k_2,i_3,k_3} \dots C^{(1,\dots,d-1)}_{k_{d-2},i_{d-1},k_{d-1}} C^{(1,\dots,d)}_{k_{d-1},i_d}$$

with polynomial functions $v_{i_\nu}^{(\nu)} \in V_\nu = \mathbb{P}_q$.

Illustration for approximation: Borehole function

Table: Approximation with prescribed precision ϵ, adaptive degree $p(\epsilon) = \log_{10}(\epsilon^{-1})$, and $N_\alpha = \dim(V_\alpha)$. Confidence intervals for relative error $\varepsilon(u^\star)$, storage complexity S and number of evaluations M for different ϵ, and average ranks.

ϵ	$\varepsilon(u^\star)$	N	S	$[r_{\{1\}}, \dots, r_{\{d\}}, r_{\{1,2\}}, \dots, r_{\{1,\dots,d-1\}}]$
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}	$[0.3; 4.0] \times 10^{-2}$	$[88, 100]$	$[41, 46]$	$[1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1]$
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}	$[1.1; 7.1] \times 10^{-8}$	$[1567, 1567]$	$[512, 512]$	$[4, 3, 4, 7, 6, 3, 3, 3, 2, 2, 3, 2, 3, 3]$
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}	$[0.3; 1.9] \times 10^{-9}$	$[2482, 2828]$	$[774, 838]$	$[5, 4, 6, 10, 7, 4, 3, 3, 2, 2, 3, 2, 3, 3]$

Conclusions

The proposed algorithm

- provides an approximation of a function in tree-based format using evaluations of the function at a structured and adapted set of points,
- provides a stable approximation with prescribed rank, with a number of samples N equal to (or of the order of) the number of parameters,
- provides an approximation with almost the desired precision.

What should be done:

- Control norms of projections and statistical estimations of principal subspaces for obtaining a certified approximation.
- Provide a priori estimations of the complexity for certain classes of functions.

References

A. Nouy.
Higher-order principal component analysis for the approximation of tensors in tree-based low-rank formats.
ArXiv e-prints, 2017.

A. Nouy.
Low-rank methods for high-dimensional approximation and model order reduction.
In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox (eds.), Model Reduction and Approximation: Theory and Algorithms. SIAM, Philadelphia, PA, 2016.

Tree-based tensor formats I

The minimal subspace $U_\alpha^{min}(u)$ of u is the smallest subspace such that

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

and $\text{rank}_\alpha(u) = \dim(U_\alpha^{min}(u))$.

- Any tensor v is such that

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

  with $S(D)$ a partition of D, and

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

  for any $\alpha \subsetneq D$ with non trivial partition $S(\alpha)$.

Tree-based tensor formats II

- For a tensor $v \in \mathcal{T}_r^T$ with $\mathrm{rank}_T(v) = r$, let $\{\varphi^{(\alpha)}_{k_\alpha}\}_{k_\alpha=1}^{r_\alpha}$ be bases of the minimal subspace $U_\alpha^{min}(v)$. The tensor v then admits a hierarchical representation

$$v = \sum_{\substack{1\leq k_\alpha \leq r_\alpha \\ \alpha\in S(D)}} p^D_{(k_\alpha)_{\alpha\in S(D)}} \bigotimes_{\alpha\in S(D)} \varphi^\alpha_{k_\alpha},$$

with

$$\varphi^\alpha_{k_\alpha} = \sum_{\substack{1\leq k_\beta \leq r_\beta \\ \beta\in S(\alpha)}} p^\alpha_{k_\alpha,(k_\beta)_{\beta\in S(\alpha)}} \bigotimes_{\beta\in S(\alpha)} \varphi^\beta_{k_\beta}$$

Partial interpolation of tensors

For a subspace W_α in $\mathcal{H}_\alpha$, we define a unisolvent set of points Γ_{W_α} in $\mathcal{X}_\alpha$ (magic points) and the associated interpolation operator I_{W_α} onto W_α defined for $v \in \mathbb{R}^{\mathcal{X}_\alpha}$ by

$$I_{W_\alpha} v(x_\alpha) = v(x_\alpha) \quad \forall x_\alpha \in \Gamma_{W_\alpha}.$$

We then define the corresponding partial interpolation operator $\mathcal{I}_{W_\alpha} = I_{W_\alpha} \otimes id_{\alpha^c}$ defined for $u \in \mathbb{R}^{\mathcal{X}}$ by

$$\mathcal{I}_{W_\alpha} u(x_\alpha, \cdot) = u(x_\alpha, \cdot) \quad \forall x_\alpha \in \Gamma_{W_\alpha}.$$

If the minimal subspace $U_\alpha^{min}(u)$ is a RKHS, then I_{W_α} is continuous from $U_\alpha^{min}(u)$ to W_α and $\mathcal{I}_{W_\alpha}$ is continuous from $U_\alpha^{min}(u) \otimes \mathcal{H}_{\alpha^c}$ to $W_\alpha \otimes \mathcal{H}_{\alpha^c}$, so that

$$\mathcal{I}_{W_\alpha} u \in W_\alpha \otimes \mathcal{H}_{\alpha^c}.$$

Statistical estimation of principal components

For $\alpha \in T$, consider $u_\alpha = \mathcal{I}_{V_\alpha} u$.

For $\|\cdot\|$ the $L^2_\mu(\mathcal{X})$-norm, the α-principal subspace of u_α is solution of

$$\min_{\dim(U_\alpha)=r_\alpha} \mathbb{E}\left(\|u_\alpha(\cdot, X_{\alpha^c}) - \mathcal{P}_{U_\alpha} u_\alpha(\cdot, X_{\alpha^c})\|^2_{L^2_{\mu_\alpha}(\mathcal{X}_\alpha)} \right),$$

where $u_\alpha(\cdot, X_{\alpha^c})$ is interpreted as a V_α-valued random variable.

It can be estimated by the solution of

$$\min_{\dim(U_\alpha)=r_\alpha} \frac{1}{N_\alpha} \sum_{k=1}^{N_\alpha} \|u_\alpha(\cdot, x^k_{\alpha^c}) - \mathcal{P}_{U_\alpha} u_\alpha(\cdot, x^k_{\alpha^c})\|^2_{\mathcal{H}_\alpha}$$

where the $x^k_{\alpha^c}$ are i.i.d. samples of X_{α^c}.

Complexity

- The storage complexity (number of parameters) of a tensor in $\mathcal{T}_r^T \cap V$ is

$$S = \sum_{\alpha \in (T \cup \{D\}) \setminus \mathcal{L}(T)} r_\alpha \prod_{\beta \in S(\alpha)} r_\beta + \sum_{\alpha \in \mathcal{L}(T)} r_\alpha \dim(V_\alpha).$$

- The total number of evaluations of the function required by the algorithm is

$$N = \sum_{\alpha \in \mathcal{L}(T)} N_\alpha \dim(V_\alpha) + \sum_{\alpha \in T \setminus \mathcal{L}(T)} N_\alpha \prod_{\beta \in S(\alpha)} r_\beta + \prod_{\beta \in S(D)} r_\beta,$$

  where N_α is the number of samples used for estimating the r_α α-principal components of u_α, taken such that

$$r_\alpha \leq N_\alpha$$

- If $N_\alpha = r_\alpha$ for all α, then

$$N = S.$$

About the constants

If oblique projections I_{U_α} and I_{V_α} were orthogonal projections, the constants Λ and $\tilde{\Lambda}$ would be equal to 1.

These constants Λ and $\tilde{\Lambda}$ depend on

$$\|I_{V_\alpha}\|_{U_\alpha^{min}(u)\to\mathcal{H}_\alpha} \quad \text{and} \quad \|I_{U_\alpha} - P_{U_\alpha}\|_{U_\alpha^{min}(u)\to\mathcal{H}_\alpha}$$

that depend on the properties of oblique projection operators restricted to minimal subspaces of u.

Case of tensor recovery

Assume that $U_\alpha^{min}(u) \subset V_\alpha$ for all leaves α (no discretization error).

If for all $\alpha \in T$, the set of N_α samples $u(\cdot, x_{\alpha^c}^k)$ contains $\mathrm{rank}_\alpha(u)$ linearly independent functions, then $U_\alpha = U_\alpha^{min}(u)$.

The constants $\Lambda = 1$, and $\tilde{\Lambda} = 1$ (i.e. same stability than the ideal algorithm).