Statistical methods for high-dimensional function approximation in tree-based tensor formats

Anthony Nouy

Ecole Centrale Nantes

Supported by the ANR (CHORUS project)

High-dimensional problems in uncertainty quantification

Parameter-dependent models

$$\mathcal{M}(u(X); X) = 0$$

where $X = (X_1, \ldots, X_d)$ are random variables.

- Forward problem: evaluation of statistics, probability of events, sensitivity indices...

$$\mathbb{E}(f(u(X))) = \int_{\mathbb{R}^d} f(u(x))p(x)dx$$

- Inverse problem: from (partial) observations of u, estimate the density of X

$$p(x)$$

- Solving forward and inverse problems requires the evaluation of the model for many instances of X.

High-dimensional problems in uncertainty quantification

- In practice, we rely on approximations of the solution map

$$x \mapsto u(x)$$

  which are used as surrogate models.

- Strategies depend on the available information on the model : from equations to simple evaluations.

- Complexity issues:
  - For complex models, only a few evaluations of the function are available.
  - High-dimensional function

$$u(x_1, \ldots, x_d)$$

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

Outline

Outline

Tensor spaces of multivariate functions

Let $\mathcal{H}_\nu$ be a Hilbert space of functions defined on $\mathcal{X}_\nu$.

The elementary tensor product $v = v^{(1)} \otimes \ldots \otimes v^{(d)}$ of functions $v^{(\nu)} \in \mathcal{H}_\nu$ is a multivariate function defined on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ such that

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

A Hilbert tensor space is then defined by

$$\mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_d = \overline{\text{span}}\{v^{(1)} \otimes \ldots \otimes v^{(d)}\}$$

equipped with the canonical norm such that

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

Consider $\mathcal{H}_\nu = L^2_{\mu_\nu}(\mathcal{X}_\nu)$ where $\mathcal{X}_\nu$ is equipped with a probability measure μ_ν, and the Hilbert tensor space

$$L^2_{\mu_1}(\mathcal{X}_1) \otimes \ldots \otimes L^2_{\mu_d}(\mathcal{X}_d) = L^2_\mu(\mathcal{X})$$

with $\mu = \mu_1 \otimes \ldots \otimes \mu_d$ and where $\|\cdot\|$ is the natural norm on $L^2_\mu(\mathcal{X})$.

Canonical rank

The canonical rank of a tensor $v \in \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_d$ is the minimal integer r such that

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

For $d = 2$, it is the unique notion of rank and

$$\mathcal{R}_r = \{v : \text{rank}(v) \leq r\}$$

is a proximinal set and a union of smooth manifolds of tensors with fixed rank.

An order-two tensor u in the Hilbert tensor space $\mathcal{H}_1 \otimes \mathcal{H}_2$ admits a singular value decomposition

$$u(x_1, x_2) = \sum_{k \geq 1} \sigma_k v_k^{(1)}(x_1) v_k^{(2)}(x_2)$$

An element of best approximation of u from $\mathcal{R}_r$ is given by the truncated singular value decomposition where we retain the r largest singular values.

Canonical rank

For $d \geq 3$, the set $\mathcal{R}_r$ looses many of the favorable properties of the case $d = 2$.

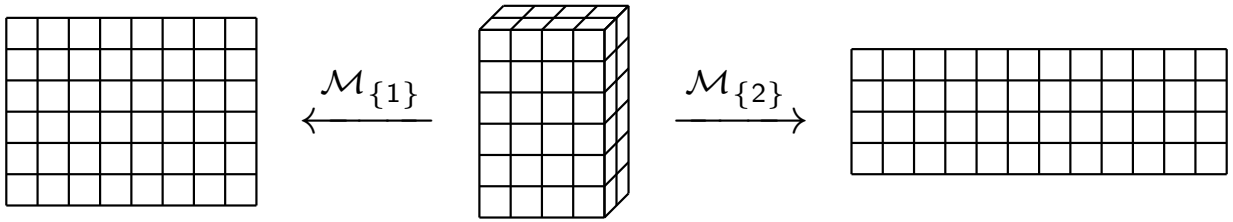
- Determining the rank of a given tensor is a NP-hard problem.
- $\mathcal{R}_r$ is not closed. The consequence is that for most problems involving approximation in canonical format $\mathcal{R}_r$, there is no robust method when $d > 2$.
- The set $\mathcal{R}_r$ is not an algebraic variety.
- No notion of singular value decomposition.

α-rank

For a non-empty subset α of $D = \{1, \ldots, d\}$, a tensor $u \in \mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_d$ can be identified with an order-two tensor

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

where $\mathcal{H}_\alpha = \bigotimes_{\nu \in \alpha} \mathcal{H}_\nu$, and $\alpha^c = D \setminus \alpha$.



The α-rank of u is the rank of the order-two tensor $\mathcal{M}_\alpha(u)$,

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

which is the minimal integer r_α such that

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

for some functions $v_k^\alpha(x_\alpha)$ and $w_k^{\alpha^c}(x_{\alpha^c})$ of groups of variables

$$x_\alpha = \{x_\nu\}_{\nu \in \alpha} \quad \text{and} \quad x_{\alpha^c} = \{x_\nu\}_{\nu \in \alpha^c}.$$

α-rank

The motivation behind the definition of tensor formats based on α-ranks is to benefit from the nice properties of the two dimensional case.

- The set
$$\mathcal{T}_{r_\alpha}^{\{\alpha\}} = \{v \in \mathcal{H} : \text{rank}_\alpha(v) \leq r_\alpha\}$$
of tensors with α-rank bounded by r_α is weakly closed (and therefore proximinal).

- For a given tensor u, $\mathcal{M}_\alpha(u)$ admits a singular value decomposition.

- The determination of the α-rank of a tensor is feasible.

- $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$ is a union of smooth manifolds of tensors with fixed α-rank.

α-ranks and related low-rank formats

For T a collection of subsets of D, we define the T-rank of a tensor v, denoted $\text{rank}_T(u)$, as the tuple

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

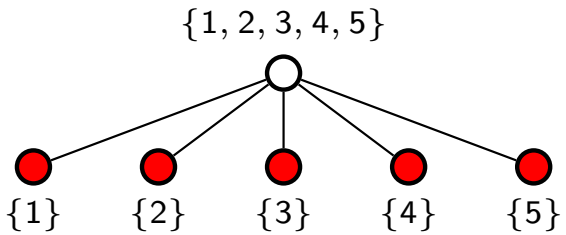
The subset of tensors with T-rank bounded by $r = (r_\alpha)_{\alpha \in T}$ is

$$\mathcal{T}_r^T = \{v \in \mathcal{H} : \text{rank}_T(v) \leq r\} = \bigcap_{\alpha \in T} \mathcal{T}_{r_\alpha}^{\{\alpha\}}.$$

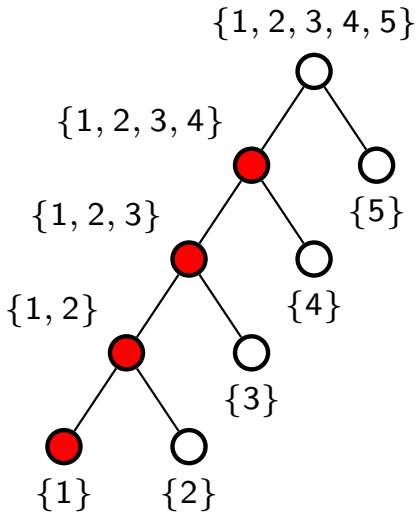
Tree-based formats

Tree-based formats correspond to a tree-structured subset T of 2^D:

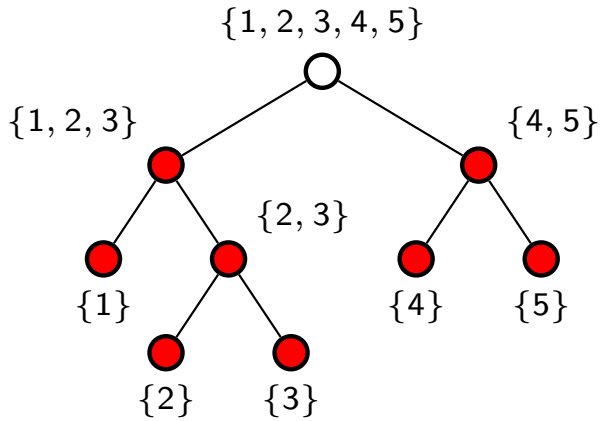
- Tucker format for $T = \{\{1\}, \ldots, \{d\}\}$
- Tensor Train format [Oseledets-Tyrtyshnikov'09] for $T = \{\{1\}, \{1,2\}, \ldots, \{1, \ldots, d-1\}\}$
- more general tree-based (or hierarchical) Tucker formats [Hackbusch-Kuhn'09]



Tucker



Tensor Train



Tree-based Tucker

Tree-based tensor formats

- A tensor v in tree-based tensor format $\mathcal{T}_r^T$ admits a representation

$$v(x_1, \ldots, x_d) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_L=1}^{r_L} \prod_{\nu=1}^{d} p^{(\nu)}\left(x_\nu, (k_i)_{i \in S_\nu}\right) \prod_{\nu=d+1}^{M} p^{(\nu)}\left((k_i)_{i \in S_\nu}\right)$$

  where parameter $p^{(\nu)}$ is a tensor depending on summation variables $(k_i)_{i \in S_\nu}$.

- Multilinear parametrization with storage complexity scaling as $O(dR^s)$ where $\#S_\nu \leq s$, $r_\nu \leq R$.

- As a finite intersection of subsets $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$, $\mathcal{T}_r^T$ inherits from nice geometrical and topological properties:
  - $\mathcal{T}_r^T$ is a union of smooth manifolds of tensors with fixed T-rank.
  - $\mathcal{T}_r^T$ is weakly closed.

- Possible extensions of the notion of singular value decomposition for higher-order tensors u.

Outline

Singular value decomposition

Consider a subset of variables α and its complementary subset $\alpha^c = D \setminus \alpha$.

A multivariate function $u(x_1, \ldots, x_d)$ is identified with a bivariate function $u \in \mathcal{H}_\alpha \otimes \mathcal{H}_{\alpha^c}$ which admits a singular value decomposition

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

The problem of best approximation of u by a function with α-rank r_α,

$$\min_{\text{rank}_\alpha(v) \leq r_\alpha} \|u - v\|^2,$$

admits as a solution the truncated singular value decomposition u_{r_α} of u

$$u_{r_\alpha}(x_\alpha, x_{\alpha^c}) = \sum_{k=1}^{r_\alpha} \sigma_i v_k^\alpha(x_\alpha) v_k^{\alpha^c}(x_{\alpha^c})$$

where $\{v_1^\alpha, \ldots, v_{r_\alpha}^\alpha\}$ are the r_α α-principal components of u.

α-principal components and associated projections

The subspace

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

is such that

$$u_{r_\alpha} = \mathcal{P}_{U_\alpha} u$$

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

It is a solution of

$$\min_{\dim(U_\alpha)=r_\alpha} \|u - \mathcal{P}_{U_\alpha} u\|^2 = \min_{\dim(U_\alpha)=r_\alpha} \int_{\mathcal{X}_{\alpha^c}} \|u(\cdot, x_{\alpha^c}) - \mathcal{P}_{U_\alpha} u(\cdot, x_{\alpha^c})\|^2_{\mathcal{H}_\alpha} \mu_{\alpha^c}(x_{\alpha^c})$$

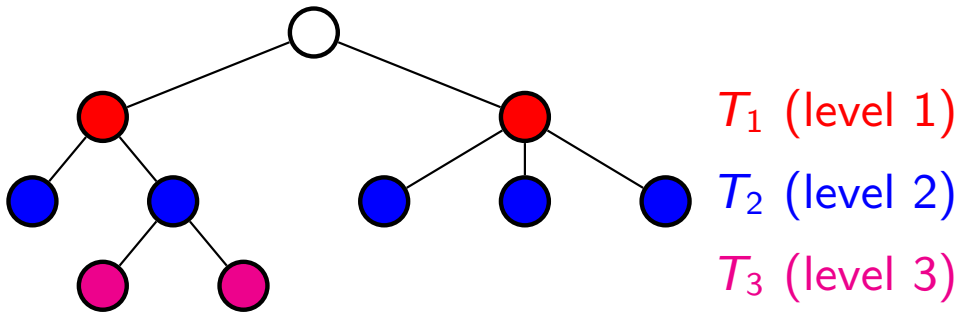
The best approximation error $\|u - \mathcal{P}_{U_\alpha} u\|$ measures how well the set

$$\{u(\cdot, x_{\alpha^c}) : x_{\alpha^c} \in \mathcal{X}_{\alpha^c}\} \subset \mathcal{H}_\alpha$$

can be approximated by a r_α dimensional space U_α. It quantifies the ideal performance of a reduced basis method in a mean-square sense.

Let T be a dimension partition tree.

$$T = \bigcup_{\ell=1}^{L} T_\ell, \quad T_\ell = \{\alpha \in T : \mathit{level}(\alpha) = \ell\}$$

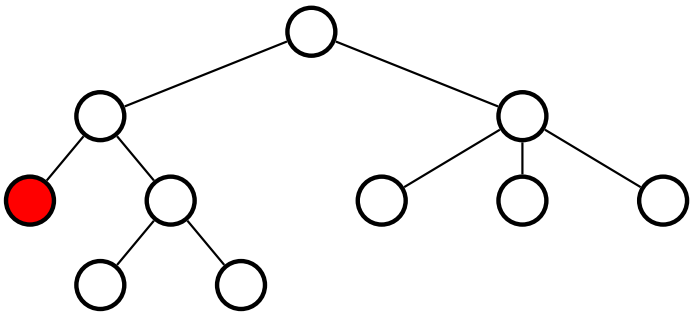


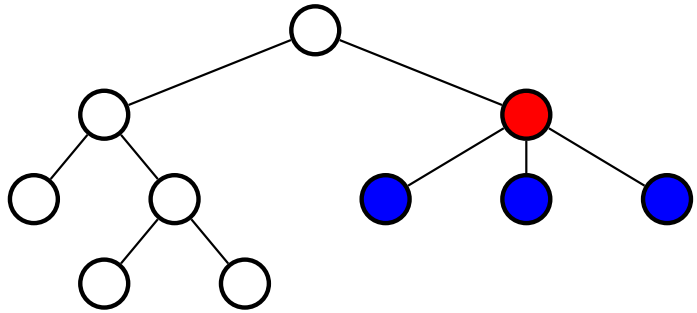
Higher-order principal component analysis for tree-based formats

Start with $u^{L+1} = u$.

At level ℓ, for each $\alpha \in T_\ell$, we determine the subspace U_α of α-principal components of u_α, with

$$u_\alpha = u \text{ if } S(\alpha) = \emptyset, \qquad u_\alpha = \prod_{\beta \in S(\alpha)} \mathcal{P}_{U_\beta} u \text{ if } S(\alpha) \neq \emptyset$$





Then we define

$$u^\ell = \mathcal{P}_\ell u^{\ell+1}, \quad \text{with} \quad \mathcal{P}_\ell = \prod_{\alpha \in T_\ell} \mathcal{P}_{U_\alpha} \quad \text{the orthogonal projection on} \bigotimes_{\alpha \in T_\ell} U_\alpha.$$

Finally, we obtain

$$u^\star = \mathcal{P}_1 \dots \mathcal{P}_L u \in \mathcal{T}_r^T$$

The approximation $u^\star$ is such that

$$\|u^\star - u\|^2 = \sum_{\ell=1}^{L} \|u^\ell - u^{\ell+1}\|^2 = \sum_{\ell=1}^{L} \| \prod_{\alpha \in T_\ell} \mathcal{P}_{U_\alpha} u^{\ell+1} - u^{\ell+1}\|^2$$

$$= \sum_{\ell=1}^{L} \sum_{\alpha \in T_\ell} \|\mathcal{P}_{U_\alpha} u^{\ell+1} - u^{\ell+1}\|^2 \leq \sum_{\ell=1}^{L} \sum_{\alpha \in T_\ell} \|\mathcal{P}_{U_\alpha} u_\alpha - u_\alpha\|^2$$

Theorem

For a desired precision ϵ, if the α-ranks are determined such that

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

we obtain a controlled approximation $u^\star$ such that

$$\|u^\star - u\| \leq \epsilon \|u\|.$$

Higher-order principal component analysis for tree-based formats

The subspace U_α of principal components of u_α is such that

$$\|\mathcal{P}_{U_\alpha} u_\alpha - u_\alpha\| = \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|v - u_\alpha\| \leq \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|v - u\| \leq \min_{v \in \mathcal{T}_r^T} \|v - u\|$$

Theorem

For a given T-rank, we obtain a quasi-optimal approximation $u^\star \in \mathcal{T}_r^T$ such that

$$\|u^\star - u\| \leq \sqrt{\#T} \min_{v \in \mathcal{T}_r^T} \|v - u\|$$

Outline

Interpolation

For a subspace U_α, we construct a unisolvent set of points Γ_{U_α} in $\mathcal{X}_\alpha$ and an associated interpolation operator $\mathcal{I}_{U_\alpha}$.

- For α a leaf of the tree, we introduce an approximation space V_α and a corresponding unisolvent set Γ_{V_α}, and construct

$$U_\alpha \subset V_\alpha \quad \text{and} \quad \Gamma_{U_\alpha} \subset \Gamma_{V_\alpha}$$

- For $\alpha \in T$ with sons $S(\alpha) \neq \emptyset$, we construct

$$U_\alpha \subset \bigotimes_{\beta \in S(\alpha)} U_\beta \quad \text{and} \quad \Gamma_{U_\alpha} \subset \bigtimes_{\beta \in S(\alpha)} \Gamma_{U_\beta}$$

The algorithm becomes

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

where $\mathcal{I}_\ell$ is an interpolation operator onto $\bigotimes_{\alpha \in T_\ell} U_\alpha$ with associated unisolvent tensorized grid $\times_{\alpha \in T_\ell} \Gamma_{U_\alpha}$, and we finally obtain an approximation

$$u^\star = \mathcal{I}_1 \ldots \mathcal{I}_L u = \mathcal{I}_L u.$$

Statistical estimation of principal components

For $\alpha \in T$, consider $u_\alpha = \mathcal{I}_{V_\alpha} u$ for α a leaf of T, and $u_\alpha = \prod_{\beta \in S(\alpha)} \mathcal{I}_{U_\beta} u$ otherwise.

The function $u_\alpha(x_\alpha, x_{\alpha^c})$ is interpreted as a random variable

$$x_{\alpha^c} \in \mathcal{X}_{\alpha^c} \mapsto u_\alpha(\cdot, x_{\alpha^c}) \in \mathcal{H}_\alpha$$

The subspace U_α as a statistical estimation of the subspace of principal components of u_α, solution of

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

where $\{x_{\alpha^c}^k\}_{k=1}^{N_\alpha}$ are i.i.d. samples of the group of variables x_{α^c}.

This is equivalent to computing the r_α principal components of

$$\{u_\alpha(\cdot, x_{\alpha^c}^1), \ldots, u_\alpha(\cdot, x_{\alpha^c}^{N_\alpha})\} \in \mathcal{H}_\alpha \otimes \mathbb{R}^{N_\alpha}$$

which requires a number of evaluations of u equal to $N_\alpha \times (\times_{\beta \in S(\alpha)} r_\beta)$ if $S(\alpha) \neq \emptyset$, and $N_\alpha \times \dim(V_\alpha)$ otherwise.

PCA-based algorithm using sampling

Theorem (Fixed precision)

For a desired precision ϵ, if the subspaces U_α are determined such that

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

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

$$\|u^\star - u\| \leq \gamma\epsilon\|u\|$$

holds with probability higher than $1-\eta\#T$, with γ depending on the properties of the interpolation operators $\mathcal{I}_{U_\alpha}$.

PCA-based algorithm using sampling

Theorem (Prescribed rank)

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

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

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

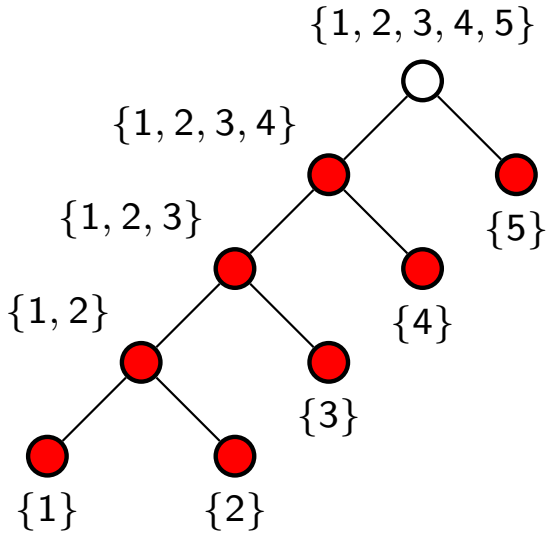
$$\|u^\star - u\| \leq C\gamma\sqrt{\#T} \min_{v \in \mathcal{T}_r^T \cap V} \|v - u\|$$

holds with probability higher than $1 - \eta\#T$, with γ depending on the properties of the interpolation operators.

Illustrations

Approximation in tensor-train Tucker format:

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



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

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

Number of parameters $\text{storage}(u^\star) \leq (q+1)dR + (d-1)R^3$, with $R = \max_\alpha r_\alpha$.

Illustration : Henon-Heiles potential

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

We run the algorithm with

- polynomial degree $q = 3$,
- prescribed T-rank $(3, \ldots, 3)$,
- a number of samples $N_\alpha = r_\alpha$ for the estimation of α-principal components, so that the required number of evaluations N of the function is equal to storage($u^\star$).

d	N	test error
10	345	$3.4e-15$
20	735	$1.9e-14$
40	1515	$3.3e-14$
80	3075	$8.7e-14$

Figure: Number fo samples and test error for different d

Illustration : 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)	$U(9855, 12045)$

Approximation with given T-rank

- Prescribed T-rank $(r, \ldots, r)$,
- Polynomial approximation with degree $q = 15$,
- $N_\alpha = \gamma r_\alpha$ samples for the estimation of U_α, so that the total number N of evaluations of the function is $\boxed{N = \gamma \text{ storage}(u^\star)}$,

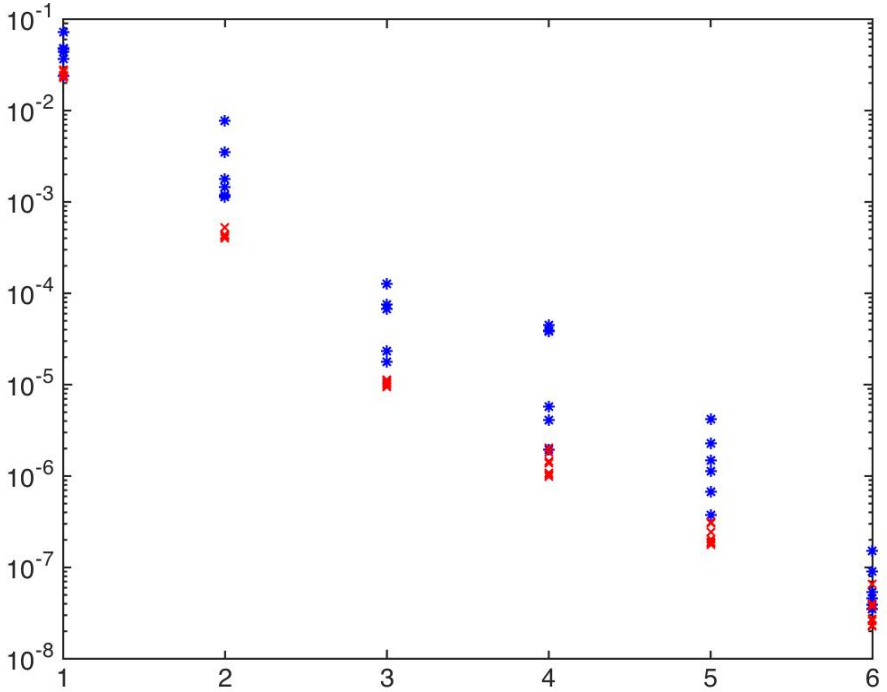


Figure: Test error with respect to rank r for $\gamma = 1$ (blue), $\gamma = 100$ (red)

T-rank	N/γ
$(1 \ldots 1)$	135
$(2 \ldots 2)$	308
$(3 \ldots 3)$	555
$(4 \ldots 4)$	912
$(5 \ldots 5)$	1415
$(6 \ldots 6)$	2100

Figure: N/γ for different values of the rank.

Approximation with prescribed precision

For the estimation of principal components, we use $N_\alpha = \beta \, \mathrm{rank}_\alpha(u_\alpha)$.

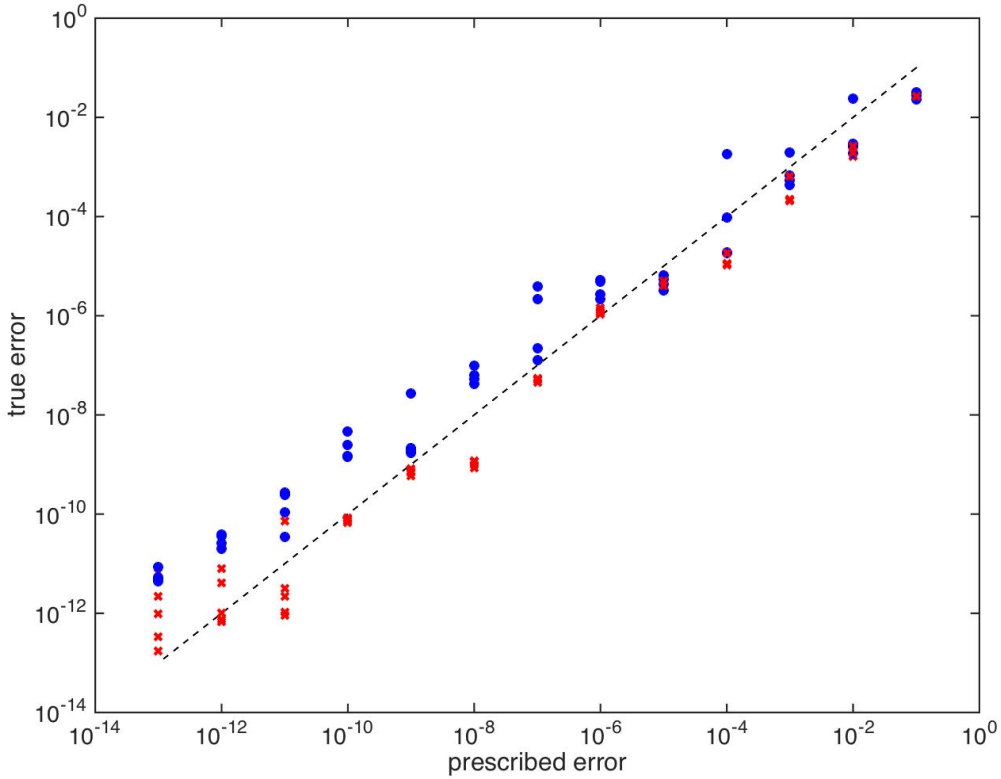


Figure: True error with respect to prescribed error ϵ for $\beta = 1$ (blue), $\beta = 100$ (red)

ϵ	N
$1e-1$	2055
$1e-3$	2118
$1e-5$	2204
$1e-7$	2803
$1e-9$	2967
$1e-11$	4720

Table: Number of required evaluations N for different precisions ϵ, for $\beta = 1$

Conclusions

The proposed algorithm

- provides an approximation of a function in tree-based format using point evaluations of the function on a structured sample set,
- 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.

Conclusions

The proposed algorithm

- provides an approximation of a function in tree-based format using point evaluations of the function on a structured sample set,
- 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.

Still to be done to further reduce the number required samples...

- Underlying bases adaptation
- Adaptive sampling for estimating principal components
- Exploit sparsity in tensor representations

Conclusions

The proposed algorithm

- provides an approximation of a function in tree-based format using point evaluations of the function on a structured sample set,
- 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.

Still to be done to further reduce the number required samples...

- Underlying bases adaptation
- Adaptive sampling for estimating principal components
- Exploit sparsity in tensor representations

Still to be done to obtain a certified control of the error...

- Control norms of interpolation operators
- Control statistical estimations

Outline

Statistical learning methods for tensor approximation

- Approximation of a function $u(X) = u(X_1, \ldots, X_d)$ from evaluations $\{y_k = u(x^k)\}_{k=1}^N$ on a training set $\{x^k\}_{k=1}^N$ (i.i.d. samples of X)

Statistical learning methods for tensor approximation

- Approximation of a function $u(X) = u(X_1, \ldots, X_d)$ from evaluations $\{y_k = u(x^k)\}_{k=1}^N$ on a training set $\{x^k\}_{k=1}^N$ (i.i.d. samples of X)
- Approximation in subsets of rank-structured functions $\mathcal{M}_r$ by minimization of an empirical risk

$$\widehat{\mathcal{R}}_N(v) = \frac{1}{N} \sum_{k=1}^{N} \ell(u(x^k), v(x^k))$$

where ℓ is a certain loss function.

Statistical learning methods for tensor approximation

- Approximation of a function $u(X) = u(X_1, \ldots, X_d)$ from evaluations $\{y_k = u(x^k)\}_{k=1}^N$ on a training set $\{x^k\}_{k=1}^N$ (i.i.d. samples of X)
- Approximation in subsets of rank-structured functions $\mathcal{M}_r$ by minimization of an empirical risk

$$\widehat{\mathcal{R}}_N(v) = \frac{1}{N} \sum_{k=1}^{N} \ell(u(x^k), v(x^k))$$

where ℓ is a certain loss function.

- For least-squares regression and noise-free observations

$$\widehat{\mathcal{R}}_N(v) = \frac{1}{N} \sum_{k=1}^{N} (u(x^k) - v(x^k))^2 = \widehat{\mathbb{E}}_N((u(X) - v(X))^2)$$

but other loss functions could be used for different objectives than L^2-approximation (e.g. classification).

Alternating minimization algorithm

- Multilinear parametrization of tensor manifolds

$$\mathcal{M}_r = \{v = F(p_1, \ldots, p_L) : p_l \in \mathbb{R}^{m_l}, 1 \leq l \leq L\}$$

so that

$$\min_{v \in \mathcal{M}_r} \widehat{\mathcal{R}}_N(v) = \min_{p_1, \ldots, p_L} \widehat{\mathcal{R}}_N(F(p_1, \ldots, p_L))$$

Alternating minimization algorithm

- Multilinear parametrization of tensor manifolds

$$\mathcal{M}_r = \{v = F(p_1, \ldots, p_L) : p_l \in \mathbb{R}^{m_l}, 1 \leq l \leq L\}$$

so that

$$\min_{v \in \mathcal{M}_r} \widehat{\mathcal{R}}_N(v) = \min_{p_1, \ldots, p_L} \widehat{\mathcal{R}}_N(F(p_1, \ldots, p_L))$$

- Alternating minimization algorithm: Successive minimization problems

$$\min_{p_l \in \mathbb{R}^{m_l}} \widehat{\mathcal{R}}_N(\underbrace{F(p_1, \ldots, p_l, \ldots, p_L)}_{\Psi_l(\cdot)^T p_l})$$

which are standard linear approximation problems

$$\min_{p_l \in \mathbb{R}^{m_l}} \frac{1}{N} \sum_{k=1}^{N} \ell(u(x^k), \Psi_l(x^k)^T p_l)$$

Alternating minimization algorithm

- Regularization

$$\min_{p_l} \frac{1}{N} \sum_{k=1}^{N} \ell(u(x^k), \Psi_l(x^k)^T p_l) + \Omega_l(p_l) \qquad (\star)$$

with regularization functional Ω_l promoting

  - smoothness (of univariate functions),
  - sparsity (e.g. $\Omega_l(p_l) = \lambda_l \|p_l\|_1$ for convex relaxation methods, or a characteristic function for working set algorithms),
  - ...

- $(\star)$ is a standard regularized linear approximation problem.
  - For square-loss and $\Omega_l(p_l) = \lambda_l \|p_l\|_1$, $(\star)$ is a LASSO problem.

- Cross-validation methods for the selection of Ω_l.

Illustration : Borehole function

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

- Approximation in tensor-train format

$$v(x_1, \ldots, x_d) = \sum_{i_1=1}^{r_1} \ldots \sum_{i_{d-1}=1}^{r_{d-1}} v_{1,i_1}^{(1)}(x_1) v_{i_1,i_2}^{(2)}(x_2) \ldots v_{i_{d-1},1}^{(d)}(x_d)$$

- Polynomial approximation of univariate functions with degree $q = 8$.
- Sparse approximation of univariate functions using working set strategies and cross-validation for the selection of polynomial bases.
- Heuristic strategy for rank adaptation: sequence of approximations with increasing T-rank (anisotropic increase of the ranks based on the α-singular values)

Illustration : Borehole function

- Training set of size $N = 1000$

iteration	rank	test error
0	(1 1 1 1 1 1 1)	$1.4\,10^{-2}$
1	(2 2 2 2 2 2 2)	$4.4\,10^{-4}$
2	(2 2 2 3 3 2 2)	$8.1\,10^{-6}$
3	(3 3 3 4 3 2 2)	$6.2\,10^{-6}$
4	(3 3 3 4 4 3 2)	$2.1\,10^{-5}$
5	(3 3 3 4 4 3 3)	$9.6\,10^{-6}$
6	(3 4 4 4 5 4 4)	$1.5\,10^{-5}$

The selected rank is one order of magnitude better than the optimal "isotropic" rank $(r, r, \ldots, r)$.

Illustration : Borehole function

- Training set of size $N = 1000$

iteration	rank	test error
0	(1 1 1 1 1 1 1)	$1.4\,10^{-2}$
1	(2 2 2 2 2 2 2)	$4.4\,10^{-4}$
2	(2 2 2 3 3 2 2)	$8.1\,10^{-6}$
3	(3 3 3 4 3 2 2)	$6.2\,10^{-6}$
4	(3 3 3 4 4 3 2)	$2.1\,10^{-5}$
5	(3 3 3 4 4 3 3)	$9.6\,10^{-6}$
6	(3 4 4 4 5 4 4)	$1.5\,10^{-5}$

The selected rank is one order of magnitude better than the optimal "isotropic" rank $(r, r, \ldots, r)$.

- Different sizes N of training set, selection of optimal ranks.

N	rank	test error
100	(3 4 4 3 3 2 1)	$7.1\,10^{-4}$
1000	(3 3 3 4 4 3 2)	$6.2\,10^{-6}$
10000	(5 6 6 7 7 5 4)	$2.4\,10^{-8}$

Concluding remarks on statistical learning methods

Statistical learning algorithms for tree-based format:

- Multilinear parametrization allows the use of the machinery of statistical learning for linear models.
- Exploit low-rank and sparsity.
- Error estimation and model selection using standard cross-validation methods.

Concluding remarks on statistical learning methods

Statistical learning algorithms for tree-based format:

- Multilinear parametrization allows the use of the machinery of statistical learning for linear models.
- Exploit low-rank and sparsity.
- Error estimation and model selection using standard cross-validation methods.

Still to be done...

- Global optimization algorithms in low-rank manifolds.
- Robust strategies for rank and tree adaptation.
- Convex relaxation of the constraints on the ranks.
- Goal-oriented approximations.

References

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.

A. Falco, W. Hackbusch, and A. Nouy.

Geometric Structures in Tensor Representations (Final Release).
ArXiv e-prints, May 2015.

M. Chevreuil, R. Lebrun, A. Nouy, and P. Rai.

A least-squares method for sparse low rank approximation of multivariate functions.
SIAM/ASA Journal on Uncertainty Quantification, 3(1):897–921, 2015.

M. Chevreuil, L. Giraldi, A. Nouy, and P. Rai.

Learning algorithms for low-rank approximation of multivariate functions in tensor-train tensor format.
Preprint available soon