Boosted optimal weighted least-squares for the approximation of high-dimensional functions in tree tensor networks

Cécile Haberstich[1], **Anthony Nouy**[2], **Guillaume Perrin**[3]

Workshop on *Optimal Sampling for Approximation* | 10 Mars 2022

[1] CEA,DAM,DIF, F-91297, Arpajon, France
[2] Centrale Nantes, LMJL, UMR CNRS 6629, France
[3] Univ Gustave Eiffel, COSYS-LISIS, F-77454 Marne-la-Vallée, France, guillaume.perrin@univ-eiffel.fr

An increasing role for simulation in our society

- Taking advantage of always increasing computational resources, the importance of simulation keeps increasing.
- It is now completely integrated in most of the decision making processes of our society.
- Thus, simulation has not only to be descriptive, but needs to be **predictive**.
- In the following, let us focus on a system whose design (dimensions, materials, initial conditions...) is characterized by $d \geq 1$ parameters gathered in a vector $\boldsymbol{x}$, and whose behavior is analyzed through the real-valued response function y.
- $\boldsymbol{x}$ is supposed to live in the space $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d \subset \mathbb{R}^d$, which is equipped with a product measure $\mu := \mu_1 \times \cdots \times \mu_d$.

$$y : \begin{cases} \mathcal{X} \to \mathbb{R} \\ \boldsymbol{x} \mapsto y(\boldsymbol{x}) \end{cases}, \quad \|y\|^2 := \int_{\mathcal{X}} y(\boldsymbol{x})^2 d\mu(\boldsymbol{x}) < +\infty.$$

High dimensional approximation

Objective : based on n couples gathered in $\mathcal{S}_n := (\boldsymbol{x}^{(i)}, y(\boldsymbol{x}^{(i)}))_{i=1}^n$, construct a predictor $\widehat{y}$ such that $\|\widehat{y} - y\|$ is minimal.

Context

- y is in $L^2_\mu(\mathcal{X})$ the Hilbert space of **square-integrable real-valued functions defined on** $\mathcal{X} \subset \mathbb{R}^d$.
- y is modeled by a deterministic black-box code (point-wise approach), whose response is supposed to be **costly** → **constraint on the maximal budget**.
- d may be high → we need additional assumptions on y to avoid the **curse of dimensionality**.

→ y has a (more or less known) **low-dimensional structure**.

→ class of **tree-based tensor formats** to exploit this low-rank structure.

High dimensional approximation

Objective : based on n couples gathered in $\mathcal{S}_n := (\boldsymbol{x}^{(i)}, y(\boldsymbol{x}^{(i)}))_{i=1}^n$, construct a predictor $\widehat{y}$ such that $\|\widehat{y} - y\|$ is minimal.

Context

- y is in $L^2_\mu(\mathcal{X})$ the Hilbert space of **square-integrable real-valued functions defined on** $\mathcal{X} \subset \mathbb{R}^d$.
- y is modeled by a deterministic black-box code (point-wise approach), whose response is supposed to be **costly** → **constraint on the maximal budget**.
- d may be high → we need additional assumptions on y to avoid the **curse of dimensionality**.

→ y has a (more or less known) **low-dimensional structure**.

→ class of **tree-based tensor formats** to exploit this low-rank structure.

⇒ This structured approximation class implies a highly structured learning design !

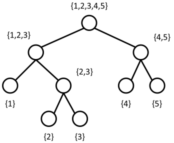
Outline

Tree-based formats

- → The structure of y is characterized by a dimension tree T defined over $D := \{1, \ldots, d\}$.
- → **Approximation class** ↔ the set $\mathcal{F}^T$ of functions written as a series of compositions of functions defined on subspaces of $\mathcal{X}$ controlled by T.

Ex. for $d = 5$: $y \in \mathcal{F}^T$ if $\exists$ $f_{\{1,2,3,4,5\}}, f_{\{1,2,3\}}, f_{\{2,3\}}, f_{\{4,5\}}$ s.t.

$$y(\boldsymbol{x}) = f_{\{1,2,3,4,5\}}(f_{\{1,2,3\}}(x_1, f_{\{2,3\}}(x_2, x_3)), f_{\{4,5\}}(x_4, x_5))$$



Objective : construct $\widehat{y}$ as a projection (empirical) of y on $\mathcal{F}^T$ (using $\mathcal{S}_n$).

Preamble : the α-principal subspaces

- For each tuple $\alpha \subset D := \{1, \ldots, d\}$, we note $\boldsymbol{x}_\alpha = (x_i)_{i \in \alpha}$ and $\boldsymbol{x}_{\alpha^c} = (x_i)_{i \notin \alpha}$.
- For each $\alpha \subset D$, function y can be **identified with the bivariate function** $y(\boldsymbol{x}_\alpha, \boldsymbol{x}_{\alpha^c})$, whose truncated SVD can be written :

$$y(\boldsymbol{x}) \approx \sum_{j=1}^{r_\alpha} \sigma_\alpha^j v_j^\alpha(\boldsymbol{x}_\alpha) v_j^{\alpha^c}(\boldsymbol{x}_{\alpha^c}).$$

→ $r_\alpha \geq 1$ is a **chosen** truncation parameter,

→ $\sigma_\alpha^1 \geq \sigma_\alpha^2 \geq \cdots$ are the **singular values**,

→ v_j^α and $v_j^{\alpha^c}$ are respectively the **left and right singular functions**,

→ $U_\alpha = \text{span}\{v_1^\alpha, \ldots, v_{r_\alpha}^\alpha\}$ is the **α-principal subspace of y** solution of

$$\min_{\dim(U_\alpha) = r_\alpha} \|y - \mathcal{P}_{U_\alpha} y\|$$

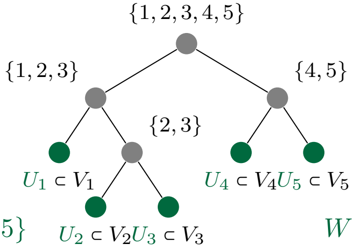
where $\mathcal{P}_{U_\alpha} y$ is the **orthogonal projection** of y onto $U_\alpha \otimes \mathbb{H}_{\alpha^c}$.

(Optimal) Leave-to-root strategy

Desired predictor : $\widehat{y} \leftrightarrow$ orthogonal projection of y on $W^{(1)}$,

with $W^{(L)} \supset W^{(L-1)} \supset \cdots \supset W^{(1)}$ a **nested sequence of tensor product subspaces with decreasing dimensions**, associated with the tree T, from the leaves to the root, and V_i a finite dimensional subspace of $L^2_{\mu_i}(\mathcal{X}_i)$.

$$V := V_1 \otimes V_2 \otimes V_3 \otimes V_4 \otimes V_5 \supset W^{(3)}$$



$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$

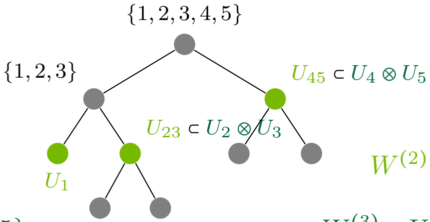
$$W^{(3)} = U_1 \otimes U_2 \otimes U_3 \otimes U_4 \otimes U_5$$

(Optimal) Leave-to-root strategy

Desired predictor : $\widehat{y} \leftrightarrow$ orthogonal projection of y on $W^{(1)}$,

with $W^{(L)} \supset W^{(L-1)} \supset \cdots \supset W^{(1)}$ a **nested sequence of tensor product subspaces with decreasing dimensions**, associated with the tree T, from the leaves to the root, and V_i a finite dimensional subspace of $L^2_{\mu_i}(\mathcal{X}_i)$.

$$V \supset W^{(3)} \supset W^{(2)}.$$



$\{1\}, \{2,3\}, \{4,5\}$ $\quad W^{(2)} = U_1 \otimes U_{23} \otimes U_{45}$

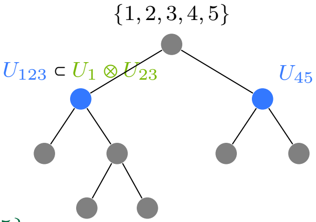
$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ $\quad W^{(3)} = U_1 \otimes U_2 \otimes U_3 \otimes U_4 \otimes U_5$

(Optimal) Leave-to-root strategy

Desired predictor : $\widehat{y} \leftrightarrow$ orthogonal projection of y on $W^{(1)}$,

with $W^{(L)} \supset W^{(L-1)} \supset \cdots \supset W^{(1)}$ a **nested sequence of tensor product subspaces with decreasing dimensions**, associated with the tree T, from the leaves to the root, and V_i a finite dimensional subspace of $L^2_{\mu_i}(\mathcal{X}_i)$.

$$V \supset W^{(3)} \supset W^{(2)} \supset W^{(1)}.$$



$\{1,2,3\},\{4,5\}$ $\quad W^{(1)} = U_{123} \otimes U_{45}$

$\{1\},\{2,3\},\{4,5\}$ $\quad W^{(2)} = U_1 \otimes U_{23} \otimes U_{45}$

$\{1\},\{2\},\{3\},\{4\},\{5\}$ $\quad W^{(3)} = U_1 \otimes U_2 \otimes U_3 \otimes U_4 \otimes U_5$

(Empirical) Leave-to-root strategy

Proposed predictor : $\widehat{y} \leftrightarrow$ empirical projection of y on $\widehat{W}^{(1)}$,

with $\widehat{W}^{(L)} \supset \widehat{W}^{(L-1)} \supset \cdots \supset \widehat{W}^{(1)}$ a nested sequence of products of **approximated** α-principal subspace $\widehat{U}_\alpha$, such that :

$$\frac{1}{n_{\alpha^c}} \sum_{k=1}^{n_{\alpha^c}} \| Q_{V_\alpha} y(\cdot, x_{\alpha^c}^k) - P_{\widehat{U}_\alpha} Q_{V_\alpha} y(\cdot, x_{\alpha^c}^k) \|^2_{L^2_{\mu_\alpha}} \quad (1)$$

is minimum, where :

- → $\{x_{\alpha^c}^k\}_{k=1}^{n_{\alpha^c}}$ are n_{α^c} i.i.d samples of the variables $X_{\alpha^c} \sim \mu_{\alpha^c}$.
- → Q_{V_α} is an empirical projector onto the space V_α based on n_α (potentially chosen) values of $\boldsymbol{x}_\alpha$,
- → $P_{\widehat{U}_\alpha}$ is the orthogonal projector on $\widehat{U}_\alpha$.

Remark : the problem associated with Eq. (1) can be solved by an SVD.

Theoretical bound for the leave-to-root strategy

Theorem

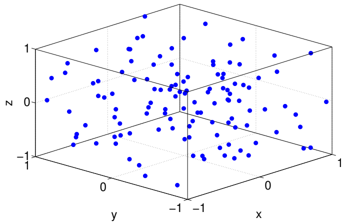
- *Assume that for all $\alpha \in T$, Q_{V_α} verifies quasi-optimality properties in expectation (more details in the next section).*
- *Assume that for all $\alpha \in T \setminus D$, the reconstruction error of the empirical α-principal subspace of $Q_{V_\alpha} y$ is controlled by the one associated to the α-principal subspace of $Q_{V_\alpha} y$ (in expectation).*

Then, the error of approximation is bounded as follows

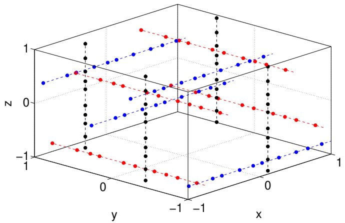
$$\mathbb{E}(\|y - \widehat{y}\|^2) \leq C_1 \varepsilon_{svd}^2 + C_2 \varepsilon_{dis}^2$$

- ε_{svd}^2 *is the error due to the SVD computed at each intermediate node,*
- ε_{dis}^2 *is the discretization error due to the introduction of finite-dimensional subspaces in the leaves (the spaces $V_1, \dots, V_d$).*
- C_1 *and* C_2 *are in* $\mathcal{O}(dC^{l(d)})$*, with* $l(d)$ *the depth of the tree.*

Model class adapted designs vs. space-filling designs



(a) SF design for stand. surr. modeling



(b) Adapted design for TBTA

- Focusing on the leaves, we notice an important structure in the positions where function y is evaluated .
- Even if d is high, the idea is to exploit the tree structure to carry out approximations in **small dimensional spaces** only.

Outline

Least-squares context

- The former strategy strongly relies on the empirical projector Q_α.
- This operator must be able to take advantage of the nested spaces identified at each step → **least squares** approaches seem particularly adapted.

Context

- We focus on the node $\alpha = (\beta, \beta^c)$.
- We have access to a m-dimensional orthonormal basis of $V_m := U_\beta \otimes U_{\beta^c}$ (by tensorization of their finite-dimensional bases), noted $\varphi_1, \ldots, \varphi_m$.
- For each $\boldsymbol{x}^k_{\alpha^c}$, we can define $Q_{V_\alpha} y(\cdot, \boldsymbol{x}^k_{\alpha^c}) = Q_{V_m} y(\cdot, \boldsymbol{x}^k_{\alpha^c}) = \sum_{j=1}^m c^\star_j \varphi_j$, with

$$(c^\star_1, \ldots, c^\star_m) \in \arg\min_{(c_1,\ldots,c_m)} \sum_{i=1}^n w(\boldsymbol{x}^i_\alpha) \left(y(\boldsymbol{x}^i_\alpha, \boldsymbol{x}^k_{\alpha^c}) - \sum_{j=1}^m c_j \varphi_j(\boldsymbol{x}^i_\alpha) \right)^2 .$$

Problematics : how to choose n, the weight function $w \geq 0$, and the $(\boldsymbol{x}^i_\alpha)_{i=1}^n$?

Stability of least-squares methods

- The **stability of Q_{V_m}** is measured by the properties of the **empirical Gram matrix $\hat{\boldsymbol{G}}_n$** (which depends on w).
- The empirical Gram matrix $\hat{\boldsymbol{G}}_n$ associated to the sample $\{\boldsymbol{x}_\alpha^i\}_{i=1}^n$ is given by

$$(\hat{\boldsymbol{G}}_n)_{k,l} = \frac{1}{n}\sum_{i=1}^{n} w(\boldsymbol{x}_\alpha^i)\varphi_k(\boldsymbol{x}_\alpha^i)\varphi_l(\boldsymbol{x}_\alpha^i).$$

- The **smaller** $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$ is, the **more stable** is the projection.
- The minimum projection error is written $\left\|y(\cdot,\boldsymbol{x}_{\alpha^c}^k) - P_{V_m} y(\cdot,\boldsymbol{x}_{\alpha^c}^k)\right\|$.

Optimal least-squares [Cohen and Migliorati., 2017]

Theorem (Optimal weighted least-squares)

Let $d\rho(x) = w(x)^{-1} d\mu(x)$ *with* $w(x)^{-1} = \frac{1}{m}\sum_{j=1}^{m} \varphi_j(x)^2$. *Let* $\eta \in (0,1)$ *and* $\delta \in (0,1)$, *and* $\boldsymbol{x}_\alpha^1, \cdots, \boldsymbol{x}_\alpha^n$ *be i.i.d from* ρ.
For $n \geq \delta^{-2} m \log(2m\eta^{-1})$, *it holds :*

$$\mathbb{P}(\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| \leq \delta) \geq 1 - \eta.$$

The approximation $Q^C_{V_m} y$ *defined by* $Q_{V_m} y$ *if* $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| < \delta$ *and* 0 *otherwise satisfies*

$$\mathbb{E}(\|y(\cdot, \boldsymbol{x}^k_{\alpha^c}) - Q^C_{V_m} y(\cdot, \boldsymbol{x}^k_{\alpha^c})\|^2) \leq (1-\delta)^{-1} \|y(\cdot, \boldsymbol{x}^k_{\alpha^c}) - P_{V_m} y(\cdot, \boldsymbol{x}^k_{\alpha^c})\|^2 + \eta \|y(\cdot, \boldsymbol{x}^k_{\alpha^c})\|^2.$$

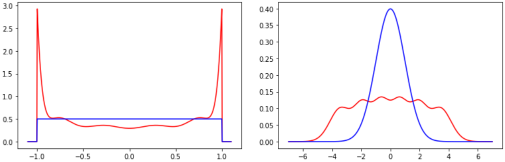


Figure – Evolution of ρ in the uniform and Gaussian cases (with $d = 1, m = 6$).

Optimal least-squares [Cohen and Migliorati., 2017]

Theorem (Optimal weighted least-squares)

Let $d\rho(x) = w(x)^{-1}d\mu(x)$ *with* $w(x)^{-1} = \frac{1}{m}\sum_{j=1}^{m}\varphi_j(x)^2$. *Let* $\eta \in (0,1)$ *and* $\delta \in (0,1)$, *and* $\boldsymbol{x}_\alpha^1, \cdots, \boldsymbol{x}_\alpha^n$ *be i.i.d from* ρ.
For $n \geq \delta^{-2} m \log(2m\eta^{-1})$, *it holds :*

$$\mathbb{P}(\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| \leq \delta) \geq 1 - \eta.$$

The approximation $Q_{V_m}^C y$ *defined by* $Q_{V_m} y$ *if* $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| < \delta$ *and* 0 *otherwise satisfies*

$$\mathbb{E}(\|y(\cdot,\boldsymbol{x}_{\alpha^c}^k) - Q_{V_m}^C y(\cdot,\boldsymbol{x}_{\alpha^c}^k)\|^2) \leq (1-\delta)^{-1}\|y(\cdot,\boldsymbol{x}_{\alpha^c}^k) - P_{V_m} y(\cdot,\boldsymbol{x}_{\alpha^c}^k)\|^2 + \eta\|y(\cdot,\boldsymbol{x}_{\alpha^c}^k)\|^2.$$

More stability / more chance to be stable $\Rightarrow$ lower δ, η $\Rightarrow$ much higher n.

$\Rightarrow$ next, another measure is proposed :

- $\rightarrow$ to impose $\eta = 0$ (to recover quasi-optimality properties in expectation),
- $\rightarrow$ to make $n \sim m$ without too much increasing δ (costly evaluations).

Boosted optimal least-squares [Haberstich et al., 2022b]

1. **Resampling :** draw M independent n-samples $\{\boldsymbol{x}^{n,i}\}_{i=1}^{M}$, with $\boldsymbol{x}^{n,i} = (x^{1,i}, \cdots, x^{n,i})$, where all $x^{j,i} \sim \rho$ are i.i.d. and choose the one which minimizes $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$.

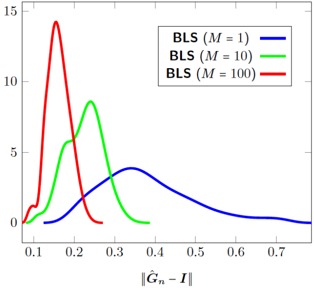


Figure – Distribution of $\|\hat{G}_n - I\|$ for $\delta = 0.9$: resampling improves the stability for a given probability η.

2. **Conditioning by rejection :** Repeat step 1 until $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| < \delta \to$ output sample $\widetilde{\boldsymbol{x}} = (\widetilde{x}^1, \cdots, \widetilde{x}^n)$. This ensures stability almost surely.

Boosted optimal least-squares [Haberstich et al., 2022b]

3. **Greedy removal of samples :** $\hat{\boldsymbol{G}}_K \leftrightarrow$ emp. Gram matrix associated with $\{\tilde{x}^i : i \in K\}$. Begin with $K = \{1, \cdots, n\}$ and while $\|\hat{\boldsymbol{G}}_K - \boldsymbol{I}\| \leq \delta$, remove sequentially a sample $\{k^\star\}$ that minimizes $\|\hat{\boldsymbol{G}}_{K \setminus \{k^\star\}} - \boldsymbol{I}\|$.

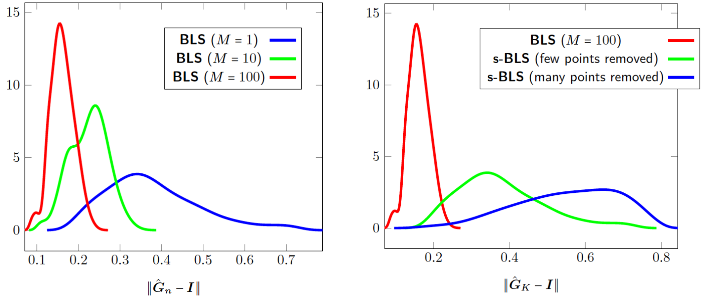


Figure – Distribution of $\|\hat{G}_n - I\|$ for $\delta = 0.9$: removing samples worsens stability while remaining bounded by δ.

Boosted optimal least-squares [Haberstich et al., 2022b]

Theorem (Control of the error bound in expectation)

Let $\eta \in (0,1)$, $\delta \in (0,1)$, and $Q_{V_m} y(\cdot, \boldsymbol{x}^k_{\alpha^c})$ be the s-BLS projection with $n_\alpha \geq \delta^{-2} m \log(2m\eta^{-1})$ and $\#K \geq n_0$. It holds :

$$\mathbb{E}(\|y(\cdot, \boldsymbol{x}^k_{\alpha^c}) - Q_{V_m} y(\cdot, \boldsymbol{x}^k_{\alpha^c})\|^2) \leq C \|y(\cdot, \boldsymbol{x}^k_{\alpha^c}) - P_{V_m} y(\cdot, \boldsymbol{x}^k_{\alpha^c})\|^2$$

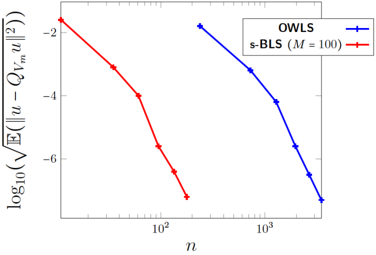
with $C = (1 + \frac{n_\alpha}{n_0}(1-\delta)^{-1}(1-\eta^M)^{-1} M)$ (better bounds can be obtained when adding assumptions on y).

- → If $n_0 = \frac{n}{\beta}$, for some $\beta \geq 1$ → quasi-optimality property (in expectation).
- → C increases with M (number of repetitions) and β (greedy reduction coefficient).
- → When $n_0 = m$, $C \sim \mathcal{O}(\log(m))$

Illustration on a simple example

$$u(x) = \frac{1}{(1 - \frac{0.5}{2d}\sum_{i=1}^{d} x_i)^{d+1}} \text{ defined on } \mathcal{X} = [-1,1]^d,\ \mu \sim U([-1,1]^d)$$

- Hyperbolic cross polynomial approximation spaces with Legendre polynomials for different m with $d = 2$.



- Guaranteed stability with probability greater than 0.99 for the OWLS method and almost surely for the s-BLS method.

→ With subsampling (OWLS → s-BLS) n (or n_α) is significantly decreased.

Hierarchical boosted least-squares

- Considering the former s-BLS method at each node of the tree, it is possible to recover a bound in expectation for the TBTA $\widehat{y}$ of y :

$$\mathbb{E}(\|y-\widehat{y}\|^2) \leq C_1 \varepsilon_{svd}^2 + C_2 \varepsilon_{dis}^2.$$

- We observed empirically that this bound was often very **loose**, in the sense that to get a desired precision ε^2, choosing ε_{svd}^2 and ε_{dis}^2 such that $\varepsilon^2 = C_1 \varepsilon_{svd}^2 + C_2 \varepsilon_{dis}^2$ is very likely to lead to values of $\mathbb{E}(\|y-\widehat{y}\|^2)$ much lower than ε^2.

⇒ **Using cross validation** techniques, it is however possible to adapt $n_{\alpha^c}, \dim(V_i)$ to impose at each leaf of the tree a chosen discretisation error ε_{dis}^2, and at each node of the tree a chosen SVD error ε_{svd}^2.

⇒ Given a desired precision ε^2, constants C_1 and C_2 can then be replaced by **heuristic values** $C_1^\star$ and $C_2^\star$ to adapt these errors so that :

$$\varepsilon_{svd}^2 \leq \frac{\varepsilon^2}{2C_1^\star}, \quad \varepsilon_{dis}^2 \leq \frac{\varepsilon^2}{2C_2^\star}.$$

Illustration of the conservative character of C_1 and C_2

- Borehole function (water flow)

$$y(x_1,\cdots,x_8) = \frac{2\pi x_3(x_4 - x_6)}{(x_2 - \log(x_1))(1 + \frac{2x_7x_3}{(x_2-\log(x_1))x_1^2x_8} + \frac{x_3}{x_5})}$$

- **Desired precision** $\varepsilon = 10^{-2}$.
- ε_{svd}^2 and ε_{dis}^2 chosen such that $\varepsilon_{svd}^2 \leq \frac{\varepsilon^2}{2C_1^\star}$, $\varepsilon_{dis}^2 \leq \frac{\varepsilon^2}{2C_2^\star}$.

	(randomly chosen) Balanced tree		
$C_1^\star$ and $C_2^\star$	$\log_{10}(\sqrt{\mathbb{E}(\\|y-\widehat{y}\\|^2)})$	m^{tot}	n^{tot}
$= C_1$, $= C_2$	-9.4	[1349 ; 2459]	[1597 ; 2742]
in $\mathcal{O}(d\widehat{C}^{l(d)})$	-3.7	[141 ; 177]	[342 ; 379]
in $\mathcal{O}(1)$	-2.0	[34 ; 51]	[168 ; 188]

Table – Different heuristics for the control of the precision, and associated confidence intervals of levels 10% and 90% for the total storage complexity m^{tot} and the total number of evaluations n^{tot}.

Empirical control of the approximation error

- $y(x) = \frac{1}{(10+2x_1+x_3+2x_4-x_5)^2}$, $\mathcal{X} = [-1,1]^6$, $\mu \sim U([-1,1]^6)$.
- Polynomial approximation spaces $V_i = \mathbb{P}_p(\mathcal{X}_i)$, with p chosen adaptively to reach a negligible discretization error using adaptive s-BLS.
- T is a (randomly chosen) balanced binary tree.
- Adaptive strategy for choosing $n_\alpha, n_{\alpha^c} + C_1^\star$ and $C_2^\star$ chosen in $\mathcal{O}(\widehat{C}^{l(d)})$.

$\log_{10}(\varepsilon)$	$\log_{10}(\sqrt{\mathbb{E}(\|y-\widehat{y}\|^2)})$	m^{tot}	n^{tot}
-2	-3	[193 ; 290]	[328 ; 403]
-3	-4.1	[309 ; 430]	[455 ; 579]
-4	-4.4	[385 ; 531]	[534 ; 697]
-5	-5.3	[588 ; 805]	[751 ; 985]
-6	-6.1	[827 ; 1268]	[1028 ; 1503]
-7	-7.0	[1203 ; 1861]	[1463 ; 2230]

→ The observed error matches with the desired precision.

→ We find values of n^{tot} (code evaluations) close to m^{tot} (complexity).

Outline

Conclusions and prospects

We proposed an algorithm that constructs, at a reasonable computational cost (close to the model complexity), a stable and controlled approximation (in expectation) of a function y in tree-based tensor format. It relies on :

- → the **BLS projection** [Haberstich et al., 2022b],
- → **adapt. strategies** for controlling the discretization error [Haberstich et al., 2022a],
- → **adapt. strategies** for controlling the construction of the α-principal subspaces U_α [Haberstich et al., 2021].

However...

- → Theoretical bounds C_1, C_2 are high compared to what we observe in numerical experiments (what hypotheses to add to better match theory and practice ?).
- → The **offline cost** remains important compared to an interpolation method for example (generation of n_α times a n_{α^c}-samples + greedy strategy).

References I

Cohen, A. and Migliorati., G. (2017).
Optimal weighted least-squares methods.
SMAI Journal of Computational Mathematics, 86(3) :181–203.

Haberstich, C., Nouy, A., and Perrin, G. (2021).
Active learning for tree tensor networks using boosted least squares.
arXiv :2104.13436v1.

Haberstich, C., Nouy, A., and Perrin, G. (2022a).
Adaptive boosted optimal weighted least-squares methods.
In preparation.

Haberstich, C., Nouy, A., and Perrin, G. (2022b).
Boosted optimal weighted least-squares methods.
Math. Comp.

Thank you for your attention.