Principal component analysis and optimal weighted least-squares method for training tree tensor networks

Cécile Haberstich [1,2 †] Anthony Nouy [1] Guillaume Perrin [2]

[1] LMJL (UMR CNRS 6629), Centrale Nantes [2] CEA/DAM/DIF, F-91297, Arpajon, France † **cecile.haberstich@ec-nantes.fr**

High-dimensional approximation

We consider a pair of random variables (X, Y) such that $Y = u(X)$, where $X = (X_1, ..., X_d)$ is a random vector and $u : \mathcal{X} \to \mathbb{R}$ is a function.
In the context of Uncertainty Quantification, Y is the output of a numerical code and X are the input parameters.
When u is costly to evaluate, we replace u by an approximation u^*.

Objectives:

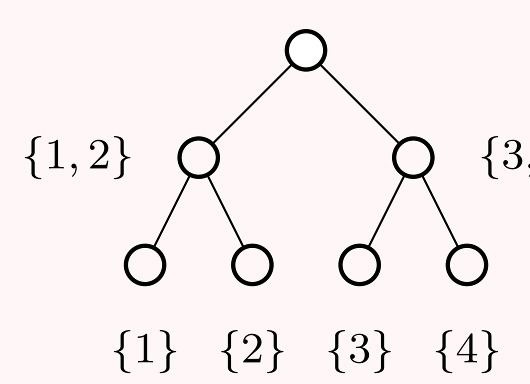
- Construct an approximation u^* using only a few evaluations of u
- Exploit low-rank structures
- Propose a sample-based algorithm with guaranteed stability

Tree-Based Tensor formats

u is in L^2_μ, with μ a product probability measure on $\mathcal{X}$.
L^2_μ is identified with $L^2_{\mu_1} \otimes ... \otimes L^2_{\mu_d}$.
For $\alpha \subset D = \{1, ..., d\}$, u is identified with a bivariate function in $L^2_{\mu_\alpha} \otimes L^2_{\mu_{\alpha^c}}$.
A function u with α-rank r_α:

$$u(x_1, ..., x_d) = \sum_{i=1}^{r_\alpha} v_i^\alpha(x_\alpha) v_i^{\alpha^c}(x_{\alpha^c})$$

If a function u has the above representation for all $\alpha \in T$ where $T \subset 2^{\{1,...d\}}$, and T is a dimension partition tree, u has a representation in tree-based tensor format $\mathcal{T}_r^T$.



A dimension partition tree T over $D = \{1, 2, 3, 4\}$

Higher-order principal component analysis for tree-based formats

The best approximation of u by a function with α-rank r_α is the truncated singular value decomposition:

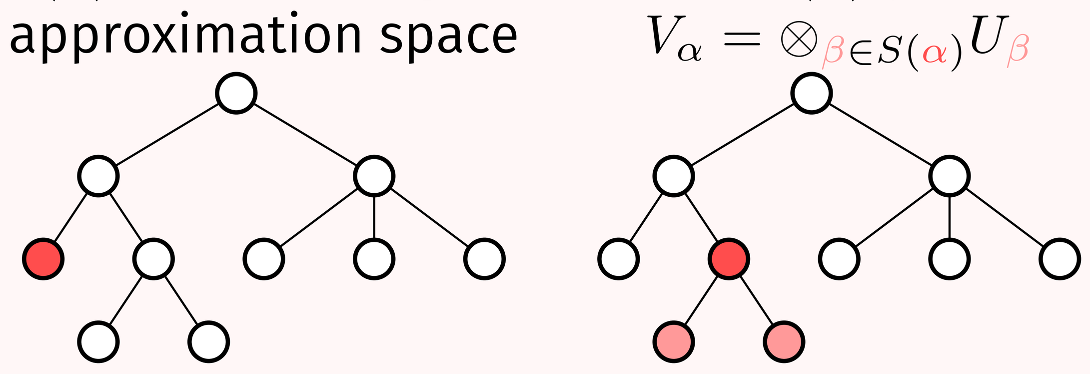
$$u_{r_\alpha}(x_\alpha, x_{\alpha^c}) = \sum_{k=1}^{r_\alpha} \sigma_k v_k^\alpha(x_\alpha) v_k^{\alpha^c}(x_{\alpha^c})$$

$v_1^\alpha, ..., v_{r_\alpha}^\alpha$ are the r_α α-principal components of u and span$\{v_1^\alpha, ..., v_{r_\alpha}^\alpha\} = U_\alpha$ is the α-principal subspace of u.

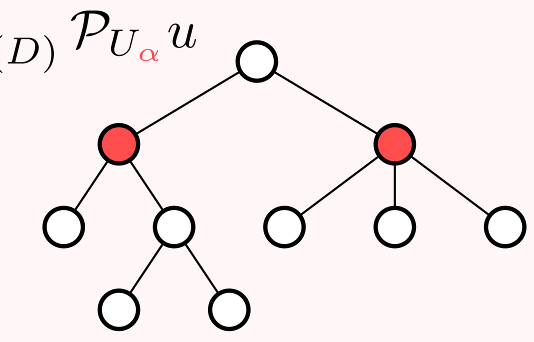
Extension to tree-based formats:
From the leaves of the tree to the root, determine the subspaces U_α of principal components of $u_\alpha = \mathcal{P}_{V_\alpha} u$ for all α, with $\mathcal{P}_{V_\alpha}$ the orthogonal projection onto $U_\alpha \otimes L^2_{\mu_{\alpha^c}}$:

if $S(\alpha) = \emptyset$, V_α is a given approximation space

if $S(\alpha) \neq \emptyset$ $V_\alpha = \otimes_{\beta \in S(\alpha)} U_\beta$



Finally project u onto the tensor space $\otimes_{\alpha \in S(D)} U_\alpha$, to obtain the approximation $u^* = \prod_{\alpha \in S(D)} \mathcal{P}_{U_\alpha} u$



Control of the error:

- algorithm with prescribed rank:
$$\|u - u^*\|^2 \leq \#T \min_{v \in \mathcal{T}_r^T} \|v - u\|^2 + \sum_{\alpha \in \mathcal{L}(T)} \|u - \mathcal{P}_{V_\alpha} u\|^2$$
- algorithm with prescribed tolerance:
if $\|u - \mathcal{P}_{V_\alpha} u\| \leq \epsilon / \sqrt{d} \|u\|$
and if r_α is chosen such that:
$\|u_\alpha - \mathcal{P}_{U_\alpha} u_\alpha\| \leq \epsilon / \sqrt{\#T} \|u_\alpha\|$
then $\|u - u^*\| \leq \epsilon \|u\|$

Optimal Weighted Least-Squares

PCA involves projections $P_V u$ with $V = V_\alpha$ or $V = U_\alpha$ subspaces of $L^2_{\mu_\alpha}$.
For a function $f \in L^2_\rho$, replace the ideal orthogonal projection:

$$P_V f = \arg\min_{v \in W} \|v - f\|_{L^2_\rho}$$

by a weighted least-squares projection:

$$P_V^w f = \arg\min_{v \in V} \frac{1}{n} \sum_{i=1}^n w^i |v(z^i) - f(z^i)|^2$$

with $(z^i)_{i=1}^n$ i.i.d samples from the measure $d\rho_w$.
$\|P_V^w f - f\|_{L^2_\rho}$ should be close to the error of best approximation and n as close as possible to m, the dimension of V.

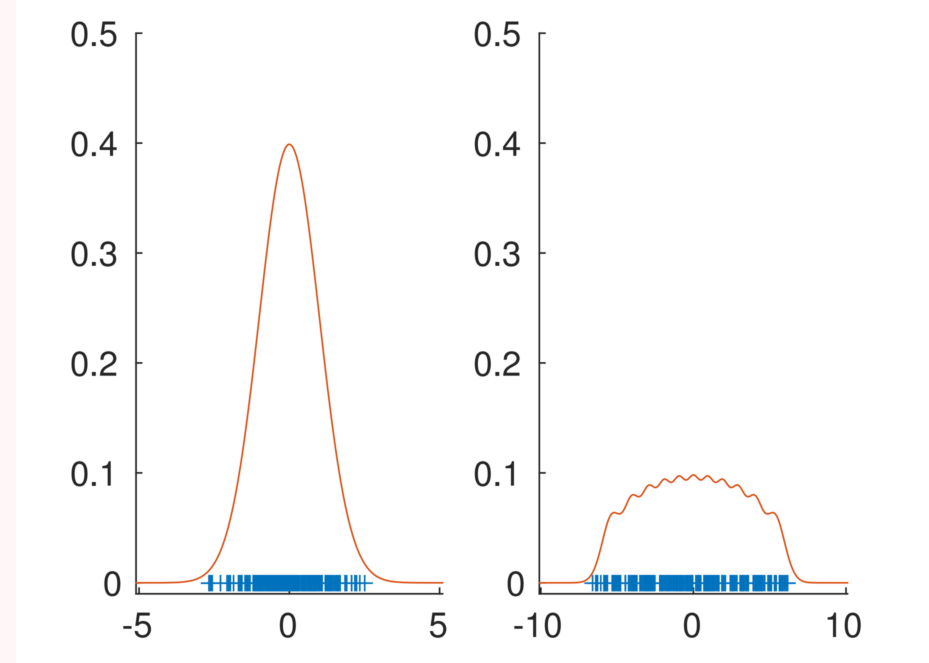
Optimal choice [1]:

1. Sample $(z^i)_{i=1}^n$ from $d\rho_w = w^{-1} d\rho$ with $w^{-1}(z) = \frac{1}{m} \sum_{i=1}^m \varphi_i(z)$ where $(\varphi_i)_{i=1}^m$ is an orthonormal basis of V
2. Choose weigths $(w^i)_{i=1}^n = (w(z^i))_{i=1}^n$

The condition $n \geq c(r) m \log(m)$ ensures the stability property:

$$Pr\{\|G - I\| \geq \frac{1}{2}\} \leq m^{-r}$$

with G the empirical grammian matrix of basis $(\varphi_i)_{i=1}^m$



Left: Pdf and samples of the gaussian measure ρ,
Right: Pdf and samples of the weighted measure ρ_w with $m = 11$

The sampling according to optimal measures exploits the tensor product structure of the functions involved.

Statistical estimation of principal components U_α

In practice we determine U_α by solving:

$$\min_{\dim(U_\alpha) = r_\alpha} \frac{1}{n_\alpha} \sum_{k=1}^{n_\alpha} \|u_\alpha(\cdot, x_{\alpha^c}^k) - P_{U_\alpha} u_\alpha(\cdot, x_{\alpha^c}^k)\|_{L^2_{\mu_\alpha}}$$

$(x_{\alpha^c}^k)_{k=1}^{n_\alpha}$ are i.i.d. samples of the group of variables x_{α^c}
P_{U_α} is the orthogonal projection onto U_α.

Numerical experiments $f(x) = \sin(\sum_{i=1}^{10} x_i)$

We consider $\mathcal{X} = \mathbb{R}^d$, with the standard Gaussian measure, and a fixed α-rank equal to 2.

polynomial degree of the approximation basis	Approximation error in $\log_{10}$			Number of samples		
	WLS	SLS	Interpolation	WLS	SLS	Interpolation
5	$[-1.1, -0.9]$	$[-1, -0.1]$	$[-0.6, -0.4]$	1292	1292	188
10	$[-3.6, -3.5]$	$[-3.4, -2.3]$	$[-4.3, -3.9]$	2212	2212	288
20	$[-9.8, -9.5]$	$[-9.7, -7.5]$	$[-13.2, -10.8]$	4232	4232	488
25	$[-13.3, -12.9]$	$[-8.8, -4.1]$	$[-13.1, -8.4]$	5292	5292	588
40	$[-14.5, -14]$	$[-7.4, -2.5]$	error	8652	8652	error

Conclusions

The optimal-weighted projection in the higher-order PCA algorithm increases the stability compared to a standard least-squares projection or interpolation. This stability is guaranteed with conditions on the number of samples.

Ongoing works:

- For a fixed rank, provide conditions on the number of samples n_α that garantee a quasi-optimality result in expectation or probability for the empirical PCA
- Provide fully adaptive algorithms (in samples, ranks) that achieve (in expectation or high probability) a prescribed precision
- Exploit sparsity of tensors

References

[1] A. Cohen and G. Migliorati. Optimal weighted least-squares methods. *SMAI Journal of Computational Mathematics*, 3:181–203, 2017.

[2] A. Nouy. Higher-order principal component analysis for the approximation of tensors in tree-based low rank formats. arXiv:1705.00880v1, 2017.