Workshop "Apprentissage et simulation en grande dimension",
Airbus Group, June 24-26, 2019

Deep tensor networks

Part II: Approximation

Anthony Nouy

Centrale Nantes, Laboratoire de Mathématiques Jean Leray

Outline

1. Higher-order singular value decomposition and tensor truncation
2. Tree optimization
3. Approximation properties of tree tensor networks

Outline

Hilbertian setting

We consider a tensor u in a Hilbert tensor space $V^1 \otimes \ldots \otimes V^d$ and we assume that u is given as a full tensor or in a certain low-rank format.

We present truncation schemes for finding a low-rank approximation of u with reduced complexity, relying on the standard singular value decomposition of order-two tensors.

We denote by $\|\cdot\|$ the canonical norm on $V^1 \otimes \ldots \otimes V^d$.

For an algebraic tensor in $\mathbb{R}^{I_1} \otimes \ldots \otimes \mathbb{R}^{I_d}$, $\|\cdot\|$ is the Frobenius norm

$$\|u\|^2 = \sum_{i_1 \in I_1} \ldots \sum_{i_d \in I_d} u(i_1, \ldots, i_d)^2$$

Truncated singular value decomposition for order-two tensors

An order-two tensor u in $V^1 \otimes V^2$ admits a singular value decomposition

$$u = \sum_{k \geq 1}^{N} \sigma_k v_k^1 \otimes v_k^2,$$

where the singular values $\sigma(u) = \{\sigma_k\}_{k \geq 1}$ are sorted by decreasing order.

An element of best approximation of u in the set of tensors with rank bounded by r is provided by the truncated singular value decomposition

$$u_r = \sum_{k=1}^{r} \sigma_k v_k^1 \otimes v_k^2,$$

with an error

$$\|u - u_r\|^2 = \min_{\text{rank}(v) \leq r} \|u - v\|^2 = \sum_{k \geq r+1} \sigma_k^2.$$

Truncated singular value decomposition for order-two tensors

An approximation u_r with relative precision ϵ, such that

$$\|u - u_r\| \leq \epsilon \|u\|,$$

can be obtained by choosing a rank r such that

$$\sum_{k \geq r+1} \sigma_k^2 \leq \epsilon^2 \sum_{k \geq 1} \sigma_k^2.$$

The complexity of computing the singular value decomposition of a tensor u is $O(n^3)$ if $\dim(V^1) = \dim(V^2) = n$. If u is given in low-rank format $u = \sum_{k=1}^{R} a_k \otimes b_k$, with a rank $R < n$, the complexity breaks down to $O(R^3 + 2Rn^2)$.

Higher-order singular value decomposition

For a non-empty subset α in $D = \{1, \dots, d\}$, a tensor $u \in V^1 \otimes \dots \otimes V^d$ can be identified with its matricisation

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

an order-two tensor which admits a singular value decomposition

$$\mathcal{M}_\alpha(u) = \sum_{k \geq 1} \sigma_k^\alpha v_k^\alpha \otimes w_k^{\alpha^c}.$$

The set $\sigma^\alpha(u) := \{\sigma_k^\alpha\}_{k \geq 1}$ is called the set of α-singular values of u.

The α-rank of u is the number of non-zero α-singular values

$$\text{rank}_\alpha(u) = \|\sigma^\alpha(u)\|_0.$$

Higher-order singular value decomposition

By sorting the α-singular values by decreasing order, an approximation u_r with α-rank r can be obtained by retaining the r largest singular values, i.e.

$$u_r \quad \text{such that} \quad \mathcal{M}_\alpha(u_r) = \sum_{k=1}^{r} \sigma_k^\alpha v_k^\alpha \otimes w_k^{\alpha^c},$$

which satisfies

$$\|u - u_r\|^2 = \min_{\text{rank}_\alpha(v) \leq r} \|u - v\|^2 = \sum_{k>r} (\sigma_k^\alpha)^2.$$

There are 2^{d-1} different binary partitions $\alpha \cup \alpha^c$ of D, to each of which corresponds a singular value decomposition and a way to truncate a higher-order tensor !

Truncation scheme for the approximation in Tucker format

For each $\nu \in \{1, \ldots, d\}$, we consider the singular value decomposition of the matricisation $\mathcal{M}_\nu(u)$ of a tensor u

$$\mathcal{M}_\nu(u) = \sum_{k \geq 1} \sigma_k^\nu v_k^\nu \otimes w_k^\nu.$$

Let $U_{r_\nu}^\nu = \text{span}\{v_k^\nu\}_{k=1}^{r_\nu}$ be the subspace of V^ν generated by the r_ν dominant left singular vectors of $\mathcal{M}_\nu(u)$, and by $P_{U_{r_\nu}}$ the orthogonal projection from V^ν to $U_{r_\nu}^\nu$.

The tensor

$$u_r = (P_{U_{r_1}^1} \otimes \ldots \otimes P_{U_{r_d}^d})u$$

is the orthogonal projection of u onto the reduced tensor space

$$U_{r_1}^1 \otimes \ldots \otimes U_{r_d}^d$$

and therefore,

$$u_r \in \mathcal{T}_r = \{v : \text{rank}_{\{\nu\}}(v) \leq r_\nu, 1 \leq \nu \leq d\}.$$

Truncation scheme for the approximation in Tucker format

The operator

$$\mathcal{P}^{\nu}_{r_\nu} = \mathcal{M}_\nu^{-1} P_{U^{\nu}_{r_\nu}} \mathcal{M}_\nu = I \otimes \ldots \otimes P_{U^{\nu}_{r_\nu}} \otimes \ldots \otimes I$$

is the orthogonal projection from V onto

$$V^1 \otimes \ldots \otimes U^{\nu}_{r_\nu} \otimes \ldots \otimes V^d,$$

which is such that

$$\|u - \mathcal{P}^{\nu}_{r_\nu} u\| = \min_{\text{rank}_\nu(v) \leq r_\nu} \|u - v\| = \sum_{k \geq r_\nu + 1} (\sigma^{\nu}_k)^2.$$

The approximation u_r can then be written

$$u_r = \mathcal{P}^{1}_{r_1} \ldots \mathcal{P}^{d}_{r_d} u,$$

and satisfies

$$\|u - u_r\|^2 \leq \sum_{\nu=1}^{d} \|u - \mathcal{P}^{\nu}_{r_\nu} u\|^2 = \sum_{\nu=1}^{d} \min_{\text{rank}_\nu(v) \leq r_\nu} \|u - v\|^2,$$

from which we deduce the quasi-optimality property

$$\|u - u_r\| \leq \sqrt{d} \min_{v \in \mathcal{T}_r} \|u - v\|.$$

Truncation scheme for the approximation in Tucker format

Also, from

$$\|u - u_r\|^2 \leq \sum_{\nu=1}^{d} \|u - \mathcal{P}^{\nu}_{r_\nu} u\|^2 = \sum_{\nu=1}^{d} \sum_{k_\nu > r_\nu} (\sigma^{\nu}_{k_\nu})^2,$$

we deduce that if we select the ranks $(r_1, \dots, r_d)$ such that for each ν

$$\sum_{k_\nu > r_\nu} (\sigma^{\nu}_{k_\nu})^2 \leq \frac{\epsilon^2}{d} \sum_{k_\nu \geq 1} (\sigma^{\nu}_{k_\nu})^2 = \frac{\epsilon^2}{d} \|u\|^2,$$

then the truncated singular value decomposition $\mathcal{P}^{\nu}_{r_\nu} u$ has a relative precision $\epsilon/\sqrt{d}$ and we finally obtain an approximation u_r with relative precision ϵ,

$$\|u - u_r\| \leq \epsilon \|u\|.$$

Note that the definition of u_r is independent on the order of the projections $\mathcal{P}^{\nu}_{r_\nu}$.

Truncation scheme for tree-based tensor formats

For tree-based tensor formats

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

where T is a dimension partition tree over $D = \{1, \dots, d\}$, a higher order singular value decomposition (also called hierarchical singular value decomposition) can also be defined from singular value decompositions of matricisations $\mathcal{M}_\alpha(u)$ of a tensor u.

Truncation scheme for tree-based tensor formats

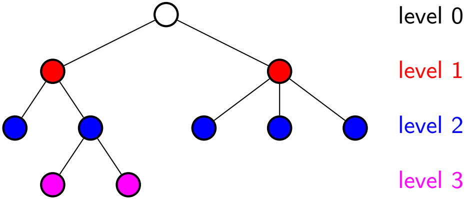
Letting $U^\alpha_{r_\alpha}$ be the subspace generated by the r_α dominant left singular vectors of $\mathcal{M}_\alpha(u)$, and letting $P_{U^\alpha_{r_\alpha}}$ be the orthogonal projector from V^α to $U^\alpha_{r_\alpha}$, we define the orthogonal projection

$$\mathcal{P}^\alpha_{r_\alpha} = \mathcal{M}_\alpha^{-1} P_{U^\alpha_{r_\alpha}} \mathcal{M}_\alpha.$$

Then, an approximation with tree-based rank $r = (r_\alpha)_{\alpha \in T}$ can be defined by

$$u_r = \mathcal{P}_r^{(L)} \mathcal{P}_r^{(L-1)} \dots \mathcal{P}_r^{(1)} u \quad \text{with} \quad \mathcal{P}^{(\ell)} = \prod_{\substack{\alpha \in T \\ \text{level}(\alpha)=\ell}} \mathcal{P}^\alpha_{r_\alpha}$$

where we apply to u a sequence of projections ordered by increasing level in the tree (from the root to the leaves). Here $L = \max_{\alpha \in T} \text{level}(\alpha)$.



Truncation scheme for tree-based tensor formats

The obtained approximation u_r is such that

$$\|u - u_r\|^2 \leq \sum_{\alpha \in T \setminus D} \min_{\text{rank}_\alpha(v) \leq r_\alpha} \|u - v\|^2 = \sum_{\alpha \in T \setminus D} \sum_{k_\alpha > r_\alpha} (\sigma^\alpha_{k_\alpha})^2,$$

from which we deduce that u_r is a quasi-optimal approximation of u in $\mathcal{T}^T_r$ such that

$$\|u - u_r\| \leq C(T) \min_{v \in \mathcal{T}^T_r} \|u - v\|,$$

where $C(T) = \sqrt{\#T - 1}$ is the square root of the number of projections applied to the tensor. The number of nodes of a dimension partition tree T being bounded by $2d - 1$,

$$C(T) \leq \sqrt{2d - 2}.$$

Also, if we select the ranks $(r_\alpha)_{\alpha \in T \setminus D}$ such that for all α

$$\sum_{k_\alpha > r_\alpha} (\sigma^\alpha_{k_\alpha})^2 \leq \frac{\epsilon^2}{C(T)^2} \sum_{k_\alpha \geq 1} (\sigma^\alpha_{k_\alpha})^2 = \frac{\epsilon^2}{C(T)^2} \|u\|^2,$$

we finally obtain an approximation u_r with relative precision ϵ,

$$\|u - u_r\| \leq \epsilon \|u\|.$$

Example: a home made function

We consider the function $u : [0,1]^6 \to \mathbb{R}$ defined by

$$f(x) = \frac{1}{3 + x_1 + x_3} + x_2 + \cos(x_4 + x_5).$$

The function is evaluated on a product grid $\Gamma \times \ldots \times \Gamma$ with N^6 points, where

$$\Gamma = \{t_k = \frac{k-1}{N-1} : 1 \leq k \leq N-1\} \subset [0,1]$$

.
We consider the tensor of order 6 in $(\mathbb{R}^N)^{\otimes 6} = \mathbb{R}^{N^6}$

$$u(i_1, \ldots, i_d) = f(t_{i_1}, \ldots, t_{i_d}).$$

See live results...

Example: conditional probability table

Let $X_1, \ldots, X_d$ be d i.i.d. random variables taking values 0 or 1 and let Y be a random variable taking the value 1 if $\sum_{i=1}^d X_i = k$ and 0 otherwise.

Let consider the function $f : \{0,1\}^{d+1} \to \{0,1\}$ defined by

$$f(y, x_1, \ldots, y_d) = \mathbb{P}(Y = y | X_1 = x_1, \ldots, X_d = x_d) = \mathbf{1}_{g(x_1, \ldots, x_d) = y}$$

with

$$g(x_1, \ldots, x_d) = \mathbf{1}_{\sum_{i=1}^d x_i = k}.$$

The function f is identified with a tensor in $(\mathbb{R}^2)^{\otimes(d+1)}$.

Example: conditional probability table

We compress the function f using tree-based tensor formats with a linear tree (tensor train format).

The ranks are chosen so that the compression relative error (in Frobenius norm) is below 10^{-10}.

The following table gives the storage complexity of the compressed tensor. The storage complexity of the initial tensor, equal to 2^{d+1}, is provided inside parentheses.

	$d = 15$ (65536)	$d = 20$ (2097152)	$d = 25$ (67108864)
$k = 1$	308	418	528
$k = 3$	620	890	1176
$k = 5$	900	1410	1984
$k = 10$	900	2082	3818
$k = 15$	180	1410	3992

Outline

Tree optimization

Consider a tensor v in a finite dimensional tensor space $\mathbb{R}^{N_1} \otimes \ldots \otimes \mathbb{R}^{N_d}$.

For a given dimension tree T, v admits an exact representation in the format $\mathcal{T}_r^T$ with $r = \mathrm{rank}_T(u)$ and a storage complexity

$$C(T, r) = \sum_{\alpha \in T \setminus \mathcal{L}(T)} r_\alpha \prod_{\beta \in S(\alpha)} r_\beta + \sum_{\alpha \in \mathcal{L}(T)} N_\alpha r_\alpha .$$

Then, we would like to find a tree T solution of

$$\min_T C(T, \mathrm{rank}_T(v)),$$

which is a combinatorial problem.

Stochastic algorithm for tree adaptation

Starting from an initial tree T and a representation of v in the corresponding format, we repeatedly

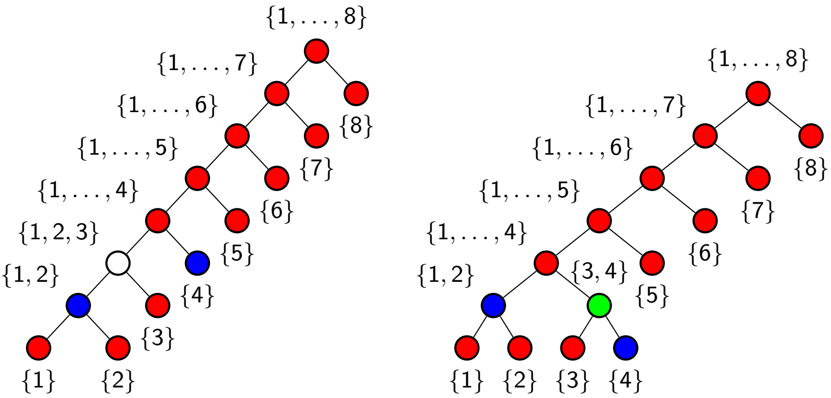
- Draw a new tree $\widetilde{T}$ from a suitable distribution over the set of trees,
- Check if

$$C(\widetilde{T}, \mathrm{rank}_{\widetilde{T}}(v)) < C(T, \mathrm{rank}_T(v)),$$

  in which case we set $T \leftarrow \widetilde{T}$ and change the representation of v.

Stochastic algorithm for tree adaptation

The tree $\widetilde{T}$ is obtained random permutations of nodes.



(a) Before permutation

(b) After permutation

Figure: Permutation of the blue nodes $\{1,2\}$ and $\{4\}$. White node is removed. Green node is added.

$\{1,\ldots,8\}$

Stochastic algorithm for tree adaptation

The tree $\widetilde{T}$ is defined as $\widetilde{T} = \sigma_m \circ \ldots \circ \sigma_1(T)$, where the map σ_{i+1} permutes two nodes of the tree $T_i = \sigma_i \circ \ldots \circ \sigma_1(T)$, with $T_0 := T$.

We first draw the number of permutations m.

Then given T_i, σ_{i+1} permutes nodes α and β in T_i, with

- α drawn with a probability increasing with the rank of its parent (since the parent will be removed).
- β drawn with a probability depending on the computational complexity for permuting α and β.

For changing the representation from T_i to T_{i+1}, use of higher-order singular value decomposition (change of representation with arbitrary accuracy).

Outline

1. Higher-order singular value decomposition and tensor truncation
2. Tree optimization
3. Approximation properties of tree tensor networks

Approximation

The goal is to approximate a function

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

by an element of a subset X_n described by a moderate number of parameters.

For a function u from a normed vector space X, the error of best approximation of u by elements of X_n is defined by

$$e_n(u) = \inf_{v \in X_n} \|u - v\|_X$$

and quantifies the best we can expect from X_n.

A sequence of subsets (X_n) is called an approximation tool.

Approximation

For a given approximation tool, fundamental problems are to

- determine if $e_n(u)$ tends to 0 for a certain class of functions (universality result)
- determine how fast $e_n(u)$ tends to 0 for a certain class of functions, e.g.

$$e_n(u) \leq M n^{-r}$$

- characterize approximation classes, i.e. sets of functions for which the approximation tool gives a certain convergence rate, e.g.

$$\mathcal{A}_r = \{u : \sup_n n^r e_n(u) < +\infty\}$$

- provide algorithms which produce approximations $u_n \in X_n$ such that

$$\|u - u_n\|_X \leq C e_n(u)$$

with C independent of n (instance optimality) or $C(n)e_n(u) \to 0$ as $n \to \infty$

What can we expect from an ideal approximation tool ?

For a set of functions K in a normed vector space X, the Kolmogorov n-width of K is

$$d_n(K)_X = \inf_{\dim(X_n)=n} \sup_{u \in K} \inf_{v \in X_n} \|u - v\|_X$$

where the infimum is taken over all linear subspaces of dimension n.

$d_n(K)_X$ measures how well the set of functions K can be approximated by a n-dimensional space.

It measures the ideal performance that we can expect from linear approximation methods.

The curse of dimensionality

- For $X = L^2(\mathcal{X})$ with $\mathcal{X} = (0,1)^d$ or $\mathcal{X} = \mathbb{T}^d$, and K the unit ball of the Sobolev space $H^k(\mathcal{X})$,

$$d_n(K)_X \sim n^{-k/d}$$

  this optimal rate being achieved with splines or trigonometric polynomials. The complexity for achieving an accuracy ϵ,

$$n(\epsilon) \lesssim \epsilon^{-d/k},$$

  may grow exponentially with d, which is the curse of dimensionality.
- For $X = L^2(\mathcal{X})$ with $\mathcal{X} = \mathbb{T}^d$, and K the unit ball of the mixed Sobolev space $H^k_{mix}(\mathcal{X})$,

$$d_n(K)_X \sim n^{-k} \log(n)^{k(d-1)},$$

  this rate being achieved by sparse tensors (hyperbolic cross approximation), and

$$n(\epsilon) \lesssim k^{-(d-1)} \epsilon^{-1/k} |\log(\epsilon)|^{d-1}$$

  The curse of dimensionality is still present.
- For $X = L^\infty(\mathcal{X})$ with $\mathcal{X} = (0,1)^d$ and $K = \{v \in C^\infty(\mathcal{X}) : \sup_\alpha \|D^\alpha v\|_{L^\infty} < \infty\}$,

$$\min\{n : d_n(K)_X \leq 1/2\} \geq c2^{d/2}$$

  No blessing of smoothness !

What about tree-based tensor formats ?

Tree-based tensor formats are nonlinear approximation tools, so the previous results do not apply.

The question is: do they perform better for smoothness classes ?

Singular value decomposition of multivariate functions

We consider a function u in $L^2_\mu(\mathcal{X})$, where $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ is equipped with a product measure $\mu = \mu_1 \otimes \ldots \otimes \mu_d$.

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

u can be identified with a bivariate function $u(x_\alpha, x_{\alpha^c})$ in $L^2_{\mu_\alpha \otimes \mu_{\alpha^c}}(\mathcal{X}_\alpha \times \mathcal{X}_{\alpha^c})$.

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

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

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

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

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

α-principal subspaces and associated projections

The subspace of principal components

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

is such that

$$u_{r_\alpha}(\cdot, x_{\alpha^c}) = \mathcal{P}_{U_\alpha} u(\cdot, x_{\alpha^c})$$

where $\mathcal{P}_{U_\alpha}$ is the orthogonal projection onto U_α.

It is solution of

$$\min_{\dim(U_\alpha)=r_\alpha} \|u - \mathcal{P}_{U_\alpha} u\|^2$$

that is

$$\min_{\dim(U_\alpha)=r_\alpha} \int \|u(\cdot, x_{\alpha^c}) - \mathcal{P}_{U_\alpha} u(\cdot, x_{\alpha^c})\|^2_{L^2_{\mu_\alpha}} \, d\mu_{\alpha^c}(x_{\alpha^c}).$$

Consider the set of functions

$$K_\alpha(u) = \{u(\cdot, x_{\alpha^c}) : x_{\alpha^c} \in \mathcal{X}_{\alpha^c}\} \subset L^2_{\mu_\alpha}(\mathcal{X}_\alpha)$$

and let ν_{α^c} be the push-forward measure of μ_{α^c} over $K_\alpha(u)$ through the map $x_{\alpha^c} \mapsto u(\cdot, x_{\alpha^c})$.

The best approximation error $e^\alpha_{r_\alpha}(u)$ is such that

$$e^\alpha_{r_\alpha}(u)^2 = \min_{\dim(U_\alpha)=r_\alpha} \int_{K_\alpha(u)} \|v - \mathcal{P}_{U_\alpha} v\|^2_{L^2_{\mu_\alpha}} d\nu_{\alpha^c}(v)$$

and defines a linear width of the set $K_\alpha(u)$ which measures how well it can be approximated by a r_α dimensional space U_α. It quantifies the ideal performance of a linear approximation method in $L^2_{\mu_\alpha}(\mathcal{X}_\alpha)$ in a mean-square sense.

Assuming μ is finite,

$$e^{\alpha}_{r_\alpha}(u) \lesssim \min_{\dim(U_\alpha)=r_\alpha} \sup_{v\in K_\alpha(u)} \|v - \mathcal{P}_{U_\alpha} v\|_{L^2_{\mu_\alpha}} = d_{r_\alpha}(K_\alpha(u))_{L^2_{\mu_\alpha}},$$

this upper bound being the Kolmogorov r_α-width of $K_\alpha(u)$ in $L^2_{\mu_\alpha}(\mathcal{X}_\alpha)$.

Furthermore, since

$$e^{\alpha}_{r_\alpha}(u) = e^{\alpha^c}_{r_\alpha}(u),$$

we have

$$e^{\alpha}_{r_\alpha}(u) \leq \min\left\{d_{r_\alpha}(K_\alpha(u))_{L^2_{\mu_\alpha}}, d_{r_\alpha}(K_{\alpha^c}(u))_{L^2_{\mu_{\alpha^c}}}\right\}$$

Best approximation in tree-based tensor format

We would like a bound on the best approximation error using tree-based tensor format

$$e_r^T(u) = \inf_{v \in \mathcal{T}_r^T} \|u - v\|_{L^2_\mu}.$$

Given a dimension tree T, for each $\alpha \in T$, we let U_α be a r_α-dimensional principal subspace of $L^2_{\mu_\alpha}$ and define

$$u_r = \mathcal{P}^{(L)} \mathcal{P}^{(L-1)} \dots \mathcal{P}^{(1)} u \quad \text{with} \quad \mathcal{P}^{(\ell)} = \prod_{\substack{\alpha \in T \\ \text{level}(\alpha) = \ell}} \mathcal{P}_{U_\alpha}$$

where we apply to u a sequence of orthogonal projections $\mathcal{P}_{U_\alpha}$ onto $U_\alpha \otimes L^2_{\mu_{\alpha^c}}$, ordered by increasing level in the tree (from the root to the leaves). Here $L = \max_{\alpha \in T} \text{level}(\alpha)$.

Best approximation in tree-based tensor format

We have that

$$\|u - u_r\|^2 \leq \sum_{\alpha \in T \setminus \{D\}} \|u - \mathcal{P}_{U_\alpha} u\|^2.$$

By taking the best possible subspaces U_α, we obtain

$$e_r^T(u) \leq \|u - u_r\|^2 \leq \sum_{\alpha \in T \setminus \{D\}} e_{r_\alpha}^\alpha(u)^2$$

with

$$e_{r_\alpha}^\alpha(u) \leq \min \left\{ d_{r_\alpha}(K_\alpha(u))_{L^2_{\mu_\alpha}}, d_{r_\alpha}(K_{\alpha^c}(u))_{L^2_{\mu_{\alpha^c}}} \right\}$$

where

$$K_\beta(u) = \{u(\cdot, x_{\beta^c}) : x_{\beta^c} \in \mathcal{X}_{\beta^c}\}$$

Approximation of smoothness classes in tree-based tensor format

Consider the approximation of a function $u \in H^k_{mix}(\mathcal{X})$ with $\mathcal{X} = (0,1)^d$. Then for any β,

$$K_\beta(u) \subset H^k_{mix}(\mathcal{X}_\beta)$$

and

$$e^\alpha_{r_\alpha}(u) \lesssim r_\alpha^{-k} \log(r_\alpha)^{k(d_\alpha - 1)}, \quad d_\alpha = \min\{\#\alpha, d - \#\alpha\}$$

If the ranks are chosen such that

$$r_\alpha \sim \epsilon^{-1/k} \log(\epsilon^{-1})^{d_\alpha - 1} (\# T)^{1/(2k)},$$

it guarantees $e^\alpha_{r_\alpha}(u) \leq \epsilon/\sqrt{\# T}$, and therefore

$$e^T_r(u) \leq \epsilon$$

for a complexity (for binary trees)

$$c(\epsilon) \lesssim \epsilon^{-3/k} \log(\epsilon^{-1})^d d^{1+3/(2k)} \quad \text{up to powers of } \log(\epsilon^{-1})$$

It performs almost as well as hyperbolic cross approximation (sparse tensors), but not better !

The result only depends on the arity of the tree. However, for a particular function, the tree may have a strong influence.

Approximation of smoothness classes in tree-based tensor format

Consider the approximation of a function $u \in H^k(\mathcal{X})$ with $\mathcal{X} = (0,1)^d$.

We can prove that

$$e^{\alpha}_{r_\alpha}(u) \lesssim r_\alpha^{-k/d_\alpha}, \quad d_\alpha = \min\{\#\alpha, d - \#\alpha\}$$

If the ranks are chosen such that

$$r_\alpha \sim \epsilon^{-d_\alpha/k} (\#T)^{d_\alpha/(2k)},$$

it guarantees $e^{\alpha}_{r_\alpha}(u) \leq \epsilon/\sqrt{\#T}$, and therefore

$$e^T_r(u) \leq \epsilon$$

for a complexity

$$c(\epsilon) \lesssim d^{1+\frac{3d}{4k}} \epsilon^{-\frac{3d}{2k}}$$

It performs almost as well as splines, but not better.

About nonlinear approximation results

There exists notions of nonlinear widths $\delta_n(K)_X$ that measure the ideal performance of nonlinear approximation tools for standard smoothness classes.

For standard smoothness classes, the performance of tree tensor networks is almost the best over all nonlinear approximation tools (covered by these notions of widths).

A first conclusion is that no (reasonable) approximation tool is able to overcome the curse of dimensionality for all functions from standard smoothness classes.

But of course, a certain approximation tool may behave well for a particular function with low-dimensional features that the approximation tool is able to capture.

Approximation of composition of functions

Consider a ridge function u defined on $\mathcal{X} = (0,1)^d$

$$u(x) = f(w^T x) = f(w_1 x_1 + \ldots + w_d x_d),$$

with $\|w\|_1 \leq 1$ and $f \in L^\infty(0,1)$.

Assuming there exists an approximation using exponential sums

$$f_r(t) = \sum_{k=1}^{r} a_k e^{b_k t} \quad \text{such that} \quad \|f - f_r\|_{L^\infty(0,1)} \leq c r^{-\gamma},$$

the function u admits a representation in canonical tensor format

$$u_r(x) = \sum_{k=1}^{r} a_k u_1^k(x_1) \ldots u_d^k(x_d), \quad u_\nu^k(x_\nu) = e^{b_k w_\nu x_\nu},$$

such that

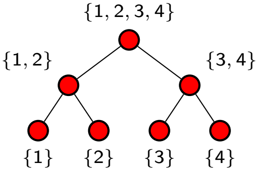
$$\|u - u_r\|_{L^\infty} \leq c r^{-\gamma} \quad \text{(No curse of dimensionality !)}$$

The same type of results holds for $u(x) = f \circ g(x)$ with $g(x) = g_1(x_1) + \ldots g_d(x_d)$ such that $\|g\|_{L^\infty} \leq 1$.

Approximation of composition of functions

Consider a function u defined on $\mathcal{X} = (0,1)^d$ which is obtained by compositions of a collection of functions $\{f_\alpha\}_{\alpha \in T}$, with T a dimension tree.

$$f_{1,2,3,4}\left(f_{1,2}\left(f_1(x_1), f_2(x_2)\right), f_{3,4}\left(f_3(x_3), f_4(x_4)\right)\right)$$



Assume that the functions $f_\alpha \in W^{k,\infty}$ with $\|f_\alpha\|_{L^\infty} \leq 1$ and $\|f_\alpha\|_{W^{k,\infty}} \leq B$.

Approximation of composition of functions

For all $\alpha \in T$,

$$e^{\alpha}_{r_\alpha}(u) \lesssim d_{r_\alpha}(K_\alpha(u))_{L^\infty_{\mu_\alpha}} \lesssim B^{k\ell_\alpha} \ell_\alpha^{k-1} r_\alpha^{-k}$$

where ℓ_α is the level of the node α.

Then

$$e^T_r(u) \leq \epsilon$$

is guaranteed with ranks

$$r_\alpha \sim \epsilon^{-1/k} \ell_\alpha^{1-1/k} B^{\ell_\alpha} (\# T)^{1/(2k)}.$$

This gives the following simplified bound on the complexity $c(\epsilon)$ to achieve accuracy ϵ

$$c(\epsilon) \lesssim \epsilon^{-(s+1)/k} (L+1)^{s+1} B^{(s+1)(L+1)} d^{1+(s+1)/2k}$$

with L the depth of the tree, and s its arity.

Approximation of composition of functions

For binary trees, the complexity to achieve precision ϵ is

$$c(\epsilon) \lesssim \epsilon^{-3/k}(L+1)^3 B^{3L} d^{1+3/2k}$$

with $L = \log_2(d)$ for a balanced tree and $L+1 = d$ for a linear tree.

- We observe a bad influence of the depth through the exponent of the norm B of functions f_α (roughness).
- For $B \leq 1$ (and even for 1-Lipschitz functions), the complexity only scales polynomially in d

$$c(\epsilon) \lesssim \epsilon^{-3/k}(L+1)^3 d^{1+3/2k} \quad \text{(no curse of dimensionality)}$$

- The choice of tree is here crucial.

Approximation of functions through tensorization

For a function $u(x)$ defined for $x \in [0,1)$, we introduce the corresponding multivariate function v defined on $\{0,...,b-1\}^d \times [0,1)$ such that

$$u(x) = v(i_0, \ldots, i_{d-1}, y)$$

where

$$x = b^{-d} y + b^{-d} \sum_{k=0}^{d-1} i_k b^k.$$

- This allows the identification (through a linear isometry)

$$L^2(0,1) = \mathbb{R}^b \ldots \ldots \mathbb{R}^b \otimes L^2(0,1).$$

- In practice, introduction of an approximation space $S_p \subset L^2(0,1)$ (e.g. polynomial space) and approximations in

$$V_{b,d,p} = \mathbb{R}^b \ldots \ldots \mathbb{R}^b \otimes S_p,$$

  using tree-based formats.
- For example, $V_{2,d,0}$ corresponds to the space of piecewise constant functions on a uniform mesh with 2^d elements.

Approximation of functions through tensorization

Exploiting low-rank structures of the tensorized function allows to achieve better performance than splines on adapted meshes for functions with singularities or multiscale functions [Kazeev and Schwab 2015 , Kazeev et al. 2017].

- For $u(x) = x^{\alpha}$, $0 < \alpha \leq 1$,
  - a piecewise constant approximation on a uniform mesh with n elements gives a convergence in $O(n^{-\alpha})$ in L^{∞},
  - a piecewise constant approximation on an optimal mesh with n elements gives a convergence in $O(n^{-1})$ in L^{∞},
  - a piecewise constant approximation on a uniform mesh with 2^d elements exploiting low-rank structures gives an exponential convergence in $O(\beta^{-n})$, where n is the complexity of the representation.
- For $u(x) = e^{zx}$, $z \in \mathbb{C}$,

$$v(i_0, \ldots, i_{d-1}, y) = u_1(i_0), \ldots u_d(i_{d-1}) u_{d+1}(y), \quad \text{with } u_k(t) = e^{ztb^{k-d}},$$

is a rank-one function whatever z.

Approximation of functions through tensorization

A promising route for high-resolution simulations in low-dimension.

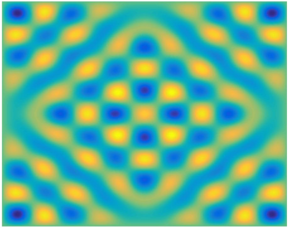


Figure: Scattering problem: tensorization with base $b = 2$, piecewise constant approximation, storage complexity at precision 10^{-3} (resp. 10^{-5}) goes from 260100 to 3532 (resp. 6170) by exploiting low-rank structures.

References I

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.

W. Hackbusch.
Tensor spaces and numerical tensor calculus, volume 42 of *Springer series in computational mathematics*.
Springer, Heidelberg, 2012.

R. Schneider and A. Uschmajew.
Approximation rates for the hierarchical tensor format in periodic sobolev spaces.
Journal of Complexity, 30(2):56 – 71, 2014.
Dagstuhl 2012.

Vladimir Kazeev and Christoph Schwab.
Approximation of singularities by quantized-tensor fem.
PAMM, 15(1):743–746, 2015.

Vladimir Kazeev, Ivan Oseledets, Maxim Rakhuba, and Christoph Schwab.
Qtt-finite-element approximation for multiscale problems i: model problems in one dimension.
Advances in Computational Mathematics, 43(2):411–442, Apr 2017.