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

Deep tensor networks

Part I: Tensors, ranks and related tensor formats

Anthony Nouy

Centrale Nantes, Laboratoire de Mathématiques Jean Leray

Outline

Outline

Algebraic tensors

Given d index sets $I_\nu = \{1, \ldots, N_\nu\}$, $1 \leq \nu \leq d$, we introduce the multi-index set

$$I = I_1 \times \ldots \times I_d.$$

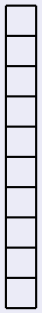
An element v of the vector space $\mathbb{R}^I$ is a tensor of order d and is identified with a multidimensional array

$$(v_i)_{i \in I} = (v_{i_1,\ldots,i_d})_{i_1 \in I_1,\ldots,i_d \in I_d}$$

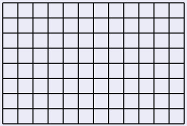
which represents the coefficients of v on the canonical basis of $\mathbb{R}^I$, also denoted

$$v(i) = v(i_1, \ldots, i_d).$$

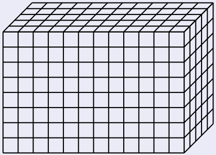
$d = 1$



$d = 2$



$d = 3$



Algebraic tensors

Given d vectors $v^{(\nu)} \in \mathbb{R}^{I_\nu}$, $1 \leq \nu \leq d$, the tensor product of these vectors

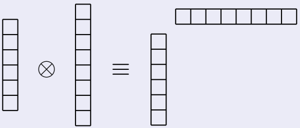
$$v := v^{(1)} \otimes \ldots \otimes v^{(d)}$$

is defined by

$$v(i) = v^{(1)}(i_1) \ldots v^{(d)}(i_d)$$

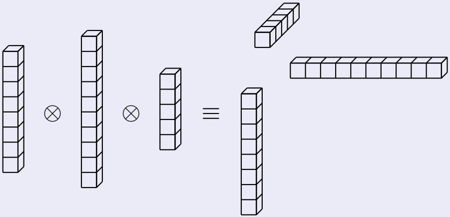
and is called an elementary tensor.

$d = 2$



Using matrix notations, $v \otimes w$ is identified with the matrix vw^T.

$d = 3$



Algebraic tensors

The tensor space $\mathbb{R}^I = \mathbb{R}^{I_1 \times \cdots \times I_d}$, also denoted $\mathbb{R}^{I_1} \otimes \ldots \otimes \mathbb{R}^{I_d}$, is defined by

$$\mathbb{R}^I = \mathbb{R}^{I_1} \otimes \ldots \otimes \mathbb{R}^{I_d} = \mathrm{span}\{v^{(1)} \otimes \ldots \otimes v^{(d)} : v^{(\nu)} \in \mathbb{R}^{I_\nu}, 1 \leq \nu \leq d\}$$

The canonical norm on $\mathbb{R}^I$, also called the Frobenius norm, is given by

$$\|v\| = \sqrt{\sum_{i \in I} v(i)^2}$$

and makes $\mathbb{R}^I$ a Hilbert space. It coincides with the natural norm on $\ell_2(I)$. It is the only norm associated with an inner product and having the property

$$\|v^{(1)} \otimes \ldots \otimes v^{(d)}\| = \|v^{(1)}\|_2 \ldots \|v^{(d)}\|_2.$$

Tensor product of functions

Let $\mathcal{X}_\nu \subset \mathbb{R}$, $1 \leq \nu \leq d$, and $V_\nu \subset \mathbb{R}^{\mathcal{X}_\nu}$ be a space of functions defined on $\mathcal{X}_\nu$.

The tensor product of functions $v^{(\nu)} \in V_\nu$, denoted

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

is a multivariate function defined on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$ and such that

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

for $x = (x_1, \ldots, x_d) \in \mathcal{X}$. For example, for $i \in \mathbb{N}_0^d$, the monomial $x^i = x_1^{i_1} \ldots x_d^{i_d}$ is an elementary tensor.

Tensor product of functions

The algebraic tensor product of spaces V_ν is defined as

$$V_1 \otimes \ldots \otimes V_d = \text{span}\{v^{(1)} \otimes \ldots \otimes v^{(d)} : v^{(\nu)} \in V_\nu, 1 \leq \nu \leq d\}$$

which is the space of multivariate functions v which can be written as a finite linear combination of elementary (separated functions), i.e.

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

Up to a formal definition of the tensor product $\otimes$, the above construction can be extended to arbitrary vector spaces V_ν (not only spaces of functions).

For infinite dimensional spaces V_ν, a Hilbert (or Banach) tensor space equipped with a norm $\|\cdot\|$ is obtained by the completion (w.r.t. $\|\cdot\|$) of the algebraic tensor space

$$\overline{V}^{\|\cdot\|} = \overline{V_1 \otimes \ldots \otimes V_d}^{\|\cdot\|}.$$

If the V_ν are Hilbert spaces with inner products $(\cdot,\cdot)_\nu$ and associated norms $\|\cdot\|_\nu$, a canonical inner product on V can be first defined for elementary tensors

$$(v^{(1)} \otimes \ldots \otimes v^{(d)}, w^{(1)} \otimes \ldots \otimes w^{(d)}) = (v^{(1)}, w^{(1)}) \ldots (v^{(d)}, w^{(d)})$$

and then extended by linearity to the whole space V.

The associated norm $\|\cdot\|$ is called the canonical norm.

Infinite dimensional tensor spaces

Example (L^p spaces)

Let $1 \leq p < \infty$. If $V_\nu = L^p_{\mu_\nu}(\mathcal{X}_\nu)$, then

$$L^p_{\mu_1}(\mathcal{X}_1) \otimes \ldots \otimes L^p_{\mu_d}(\mathcal{X}_d) \subset L^p_\mu(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$$

with $\mu = \mu_1 \otimes \ldots \otimes \mu_d$, and

$$\overline{L^p_{\mu_1}(\mathcal{X}_1) \otimes \ldots \otimes L^p_{\mu_d}(\mathcal{X}_d)}^{\|\cdot\|} = L^p_\mu(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$$

where $\|\cdot\|$ is the natural norm on $L^p_\mu(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$.

Example (Bochner spaces)

Let $\mathcal{X}$ be equipped with a finite measure μ, and let W be a Hilbert (or Banach) space. For $1 \leq p < \infty$, the Bochner space $L^p_\mu(\mathcal{X}; W)$ is the set of Bochner-measurable functions $u : \mathcal{X} \to W$ with bounded norm $\|u\|_p = (\int_{\mathcal{X}} \|u(x)\|^p_W \mu(dx))^{1/p}$, and

$$L^p_\mu(\mathcal{X}; W) = \overline{W \otimes L^p_\mu(\mathcal{X})}^{\|\cdot\|_p}.$$

Example (Sobolev spaces)

The Sobolev space $H^k(\mathcal{X})$ of functions defined on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$, equipped with the norm

$$\|u\|_{H^k}^2 = \sum_{|\alpha|_1 \leq k} \|D^\alpha u\|_{L^2}^2,$$

is a Hilbert tensor space

$$H^k(\mathcal{X}) = \overline{H^k(\mathcal{X}_1) \otimes \ldots \otimes H^k(\mathcal{X}_d)}^{\|\cdot\|_{H^k}}.$$

The Sobolev space $H^k_{mix}(\mathcal{X})$ equipped with the norm

$$\|u\|_{H^k_{mix}}^2 = \sum_{|\alpha|_\infty \leq k} \|D^\alpha u\|_{L^2}^2,$$

is a different tensor Hilbert space

$$H^k_{mix}(\mathcal{X}) = \overline{H^k(\mathcal{X}_1) \otimes \ldots \otimes H^k(\mathcal{X}_d)}^{\|\cdot\|_{H^k_{mix}}}.$$

$\|u\|_{H^k_{mix}}^2$ is the canonical tensor norm on $H^k(\mathcal{X}_1) \otimes \ldots \otimes H^k(\mathcal{X}_d)$.

Tensor product basis

If $\{\psi_i^{(\nu)}\}_{i \in I_\nu}$ is a basis of V_ν, then a basis of $V = V_1 \otimes \ldots \otimes V_d$ is given by

$$\left\{ \psi_i = \psi_{i_1}^{(1)} \otimes \ldots \otimes \psi_{i_d}^{(d)} : i \in I = I_1 \times \ldots \times I_d \right\}.$$

A tensor $v \in V$ admits a decomposition

$$v = \sum_{i \in I} a_i \psi_i = \sum_{i_1 \in I_1} \ldots \sum_{i_d \in I_d} a_{i_1, \ldots, i_d} \psi_{i_1}^{(1)} \otimes \ldots \otimes \psi_{i_d}^{(d)},$$

and v can be identified with the set of its coefficients

$$a \in \mathbb{R}^I.$$

Hilbert tensor spaces

If the $\{\psi_i^{(\nu)}\}_{i \in I_\nu}$ are orthonormal bases of spaces V_ν, then $\{\psi_i\}_{i\in I}$ is an orthonormal basis of $\overline{V}^{\|\cdot\|}$. A tensor

$$v = \sum_{i \in I} a_i \psi_i$$

is such that

$$\|v\|^2 = \sum_{i \in I} a_i^2 := \|a\|^2.$$

Therefore, the map

$$\Psi : a \mapsto \sum_{i \in I} a_i \psi_i$$

defines a linear isometry from $\mathbb{R}^I$ to V for finite dimensional spaces, and between $\ell_2(I)$ and $\overline{V}^{\|\cdot\|}$ for infinite dimensional spaces.

Curse of dimensionality

A tensor $a \in \mathbb{R}^I = \mathbb{R}^{I_1 \times \ldots \times I_d}$ or a corresponding tensor $v = \sum_{i \in I} a_i \psi_i$, when $\# I_\nu = O(n)$ for each ν, has a storage complexity

$$\# I = \# I_1 \ldots \# I_d = O(n^d)$$

which grows exponentially with the dimension.

Manipulating tensors requires exploiting special properties of tensors such as sparsity or low rankness.

Outline

Rank of order-two tensors

The rank of an order-two tensor $u \in V \otimes W$, denoted $\mathrm{rank}(u)$, is the minimal integer r such that

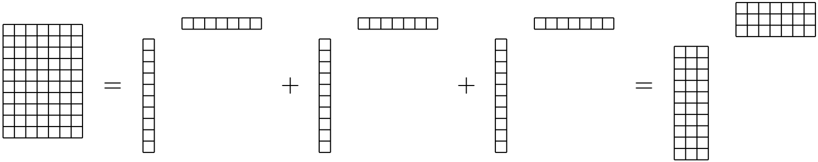
$$u = \sum_{k=1}^{r} v_k \otimes w_k$$

for some $v_k \in V$ and $w_k \in W$.

A tensor $u \in \mathbb{R}^n \otimes \mathbb{R}^m$ is identified with a matrix $u \in \mathbb{R}^{n \times m}$. The rank of u coincides with the matrix rank, which is the minimal integer r such that

$$u = \sum_{k=1}^{r} v_k w_k^T = VW^T,$$

where $V = (v_1, \ldots, v_r) \in \mathbb{R}^{n \times r}$ and $W = (w_1, \ldots, w_r) \in \mathbb{R}^{m \times r}$.



Singular value decomposition of order-two tensors

When V and W are Hilbert spaces (possibly infinite-dimensional), an algebraic tensor $u \in V \otimes W$ admits a singular value decomposition

$$u = \sum_{k \geq 1} \sigma_k v_k \otimes w_k,$$

where v_k and w_k are orthonormal vectors (singular vectors) and $\sigma_k \in \mathbb{R}^+$ are the singular values.

The rank of u is finite and coincides with the number of non-zero singular values,

$$\text{rank}(u) = \#\{k : \sigma_k \neq 0\}.$$

Example (Singular value decomposition of matrices)

For $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, u is identified with a matrix in $\mathbb{R}^{n \times m}$ and

$$u = \sum_{k=1}^{\text{rank}(u)} \sigma_k v_k w_k^T = \mathbf{V}\mathbf{S}\mathbf{W}^T.$$

with orthogonal matrices $\mathbf{V}$ and $\mathbf{W}$, and a diagonal matrix $\mathbf{S}$.

Singular value decomposition of order-two tensors

An algebraic tensor $u \in V \otimes W$ can be identified with a linear operator from W to V with rank equal to $\text{rank}(u)$.

For infinite dimensional Hilbert spaces, the closure $\overline{V \otimes W}^{\|\cdot\|_\vee}$ of $V \otimes W$ with respect to the injective norm (corresponding to the operator norm) coincides with the space of compact operators.

A tensor $u \in \overline{V \otimes W}^{\|\cdot\|_\vee}$ still admits a singular value decomposition

$$u = \sum_{k \geq 1} \sigma_k v_k \otimes w_k.$$

and the rank (number of non-zero singular values) is possibly infinite.

Singular value decomposition of order-two tensors

Example (Proper Orthogonal Decomposition)

For $\Omega \times I$ a space-time domain and V a Hilbert space of functions defined on Ω, a function $u \in L^2(I; V)$ admits a singular value decomposition

$$u(t) = \sum_{k=1}^{\infty} \sigma_k v_k w_k(t)$$

which is known as the Proper Orthogonal Decomposition (POD).

Example (Karhunen-Loeve decomposition)

For a probability space (Ω, μ), an element $u \in L^2_\mu(\Omega; V)$ is a second-order V-valued random variable. If u is zero-mean, the singular value decomposition of u is known as the Karhunen-Loeve decomposition

$$u(\omega) = \sum_{k=1}^{\infty} \sigma_k v_k w_k(\omega)$$

where $w_k : \Omega \to \mathbb{R}$ are uncorrelated (orthogonal) random variables.

Low-rank format for order-two tensors

The set of tensors in $V \otimes W$ with rank bounded by r, denoted

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

is not a linear space nor a convex set. However, it has many favorable properties for a numerical use.

- The application $v \mapsto \text{rank}(v)$ is lower semi-continuous, and therefore the set $\mathcal{R}_r$ is closed, which makes best approximation problems in $\mathcal{R}_r$ well posed.
- $\mathcal{R}_r$ is the union of smooth manifolds of tensors with fixed rank.

Canonical rank of higher-order tensors

For tensors $u \in V_1 \otimes \ldots \otimes V_d$ with $d \geq 3$, there are different notions of rank.

The canonical rank, which is the natural extension of the notion of rank for order-two tensors, is the minimal integer r such that

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

for some vectors $v_k^{(\nu)} \in V_\nu$.

Canonical format

The subset of tensors in $V = V_1 \otimes \ldots \otimes V_d$ with canonical rank bounded by r is denoted

$$\mathcal{R}_r = \{v \in V : \operatorname{rank}(v) \leq r\}.$$

A tensor in $\mathcal{R}_r$ has a representation

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

The storage complexity of tensors in $\mathcal{R}_r$ is

$$\text{storage}(\mathcal{R}_r) = r \sum_{\nu=1}^{d} \dim(V_\nu) = O(rdn)$$

for $\dim(V_\nu) = O(n)$.

$\mathcal{R}_r$ is a universal approximation tool since

$$\bigcup_{r \geq 1} \mathcal{R}_r \text{ is dense in } V$$

so that for any $u \in V$, we can find a sequence $\{u_r\}_{r \geq 1}$ with $u_r \in \mathcal{R}_r$ concerving to u.

Canonical format

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.
- The set $\mathcal{R}_r$ is not an algebraic variety.
- No notion of singular value decomposition.
- The application $v \mapsto \text{rank}(v)$ is not lower semi-continuous and therefore, $\mathcal{R}_r$ is not closed.

Example

Consider the order-3 tensor

$$v = a \otimes a \otimes b + a \otimes b \otimes a + b \otimes a \otimes a$$

where a and b are linearly independent vectors in $\mathbb{R}^m$. The rank of v is 3. The sequence of rank-2 tensors

$$v_n = n(a + \frac{1}{n}b) \otimes (a + \frac{1}{n}b) \otimes (a + \frac{1}{n}b) - na \otimes a \otimes a$$

converges to v as $n \to \infty$.

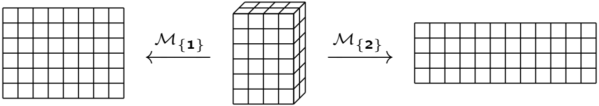
- The consequence is that for most problems involving approximation in canonical format $\mathcal{R}_r$, there is no robust method when $d > 2$.

α-rank

For a non-empty subset α of $D = \{1, \dots, d\}$, a tensor $u \in V = V_1 \otimes \dots \otimes V_d$ can be identified with an order-two tensor

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

where $V_\alpha = \bigotimes_{\nu \in \alpha} V_\nu$, and $\alpha^c = D \setminus \alpha$. The operator $\mathcal{M}_\alpha = V \to V_\alpha \otimes V_{\alpha^c}$ is called the matricisation operator.



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

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

which is the minimal integer r_α such that

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

for some $v_k^\alpha \in V_\alpha$ and $w_k^{\alpha^c} \in V_{\alpha^c}$. We note that $\mathrm{rank}_\alpha(u) = \mathrm{rank}_{\alpha^c}(u)$.

α-rank

A multivariate function $u(x_1, \dots, x_d)$ with $\text{rank}_\alpha(u) \le r_\alpha$ is 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

Example

- $u(x) = u^1(x_1) \dots u^d(x_d)$ can be written $u(x) = u^\alpha(x_\alpha) u^{\alpha^c}(x_{\alpha^c})$, with $u^\alpha(x_\alpha) = \prod_{\nu \in \alpha} u^\nu(x_\nu)$. Therefore, for any α, $\text{rank}_\alpha(u) = 1$.

- $u(x) = \prod_{\alpha \in T} u^\alpha(x_\alpha)$ with T a collection of disjoint subsets, is such that $\text{rank}_\alpha(u) = 1$ for all $\alpha \in T$, and $\text{rank}_\gamma(u) \leq \prod_{\alpha \in T, \alpha \cap \gamma \neq \emptyset} \text{rank}_{\gamma \cap \alpha}(u^\alpha)$ for all γ.

- $u(x) = u^1(x_1) + \dots + u^d(x_d)$ can be written $u(x) = u^\alpha(x_\alpha) + u^{\alpha^c}(x_{\alpha^c})$, with $u^\alpha(x_\alpha) = \sum_{\nu \in \alpha} u^\nu(x_\nu)$. Therefore, $\text{rank}_\alpha(u) \leq 2$.

- $u(x) = \sum_{k=1}^r u_k^1(x_1) \dots u_k^d(x_d)$ can be written $\sum_{k=1}^r u_k^\alpha(x_\alpha) u_k^{\alpha^c}(x_{\alpha^c})$ with $u_k^\alpha(x_\alpha) = \prod_{\nu \in \alpha} u_k^\nu(x_\nu)$. Therefore, for any α, $\text{rank}_\alpha(u) \leq r$, with equality if the functions $\{u_k^\alpha(x_\alpha)\}$ and the functions $\{u_k^{\alpha^c}(x_{\alpha^c})\}$ are linearity independent.

  We deduce the following relation between α-ranks and canonical rank:

$$\text{rank}_\alpha(u) \leq \text{rank}(u), \quad \text{for any } \alpha.$$

α-rank and minimal subspace

For a subset α of $D = \{1, \dots, d\}$, the minimal subspace

$$U_\alpha^{min}(u)$$

of a tensor $u \in V_1 \otimes \dots \otimes V_d$ is defined as the smallest subspace

$$U_\alpha \subset V_\alpha = \bigotimes_{\nu \in \alpha} V_\nu$$

such that

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

The α-rank of u is the dimension of the minimal subspace $U_\alpha^{min}(u)$,

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

If u admits the representation

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

then $U_\alpha^{min}(u) = span\{v_k^\alpha : 1 \leq k \leq \text{rank}_\alpha(u)\}$.

α-ranks and related tensor formats

The subset of tensors

$$\mathcal{T}_{r_\alpha}^{\{\alpha\}} = \{v : \text{rank}_\alpha(v) \leq r_\alpha\}$$

is also characterized by

$$\mathcal{T}_{r_\alpha}^{\{\alpha\}} = \{v : \dim(U_\alpha^{min}(v)) \leq r_\alpha\}$$

and has the nice properties of low-rank formats for order-two tensors.

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

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

Outline

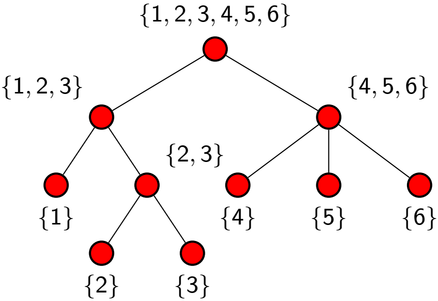
Tree-based tensor format

Tree-based tensor formats are subsets of tensors

$$\mathcal{T}_r^T = \{v \in V : \text{rank}_T(v) \leq r\}$$

where T is a dimension partition tree T over $D = \{1, \dots, d\}$, with root D and leaves

$$\mathcal{L}(T) = \{\{\nu\} : 1 \leq \nu \leq d\}.$$



The tree-based rank of a tensor v is the tuple $\text{rank}_T(v) = (\text{rank}_\alpha(v))_{\alpha \in T}$.

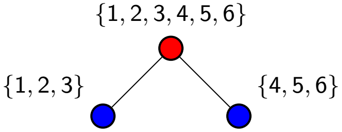
By convention, $\text{rank}_D(v) = 1$.

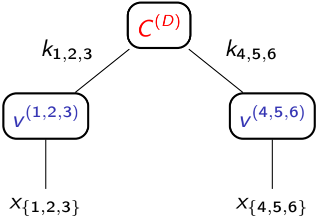
Tree-based tensor format

Elements of $\mathcal{T}_r^T$ admit an explicit representation. Let $v \in \mathcal{T}_r^T$ with T-rank $r = (r_\alpha)_{\alpha \in T}$. At the first level, v admits the representation

$$v(x) = \sum_{k_{\beta_1}=1}^{r_{\beta_1}} \dots \sum_{k_{\beta_s}=1}^{r_{\beta_s}} C^{(D)}_{k_{\beta_1},\dots,k_{\beta_s}} v^{(\beta_1)}_{k_{\beta_1}}(x_{\beta_1}) \dots v^{(\beta_s)}_{k_{\beta_s}}(x_{\beta_s})$$

where $\{\beta_1, \dots, \beta_s\} = S(D)$ are the children of the root node D, and $\{v^{(\beta)}_{k_\beta}\}_{1 \le k_\beta \le r_\beta}$ form a basis of the minimal subspace $U^{min}_\beta(v)$.

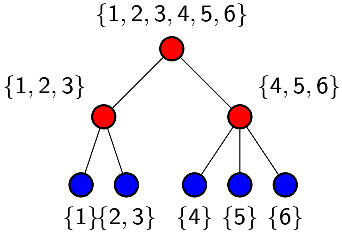


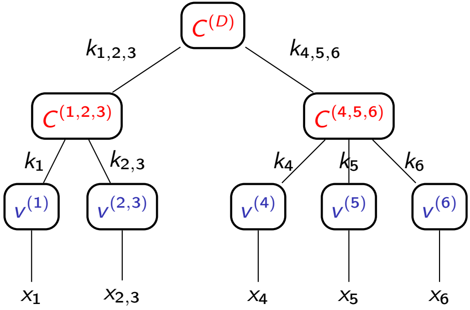


Tree-based tensor format

Then, for an interior node α of the tree, with children $S(\alpha) = \{\beta_1, \ldots, \beta_s\}$, the functions (or tensors) $v_{k_\alpha}^{(\alpha)}$ admit the representation

$$v_{k_\alpha}^{(\alpha)}(x_\alpha) = \sum_{k_{\beta_1}=1}^{r_{\beta_1}} \ldots \sum_{k_{\beta_s}=1}^{r_{\beta_s}} C_{k_\alpha, k_{\beta_1}, \ldots, k_{\beta_s}}^{(\alpha)} v_{k_{\beta_1}}^{(\beta_1)}(x_{\beta_1}) \ldots v_{k_{\beta_s}}^{(\beta_s)}(x_{\beta_s}).$$

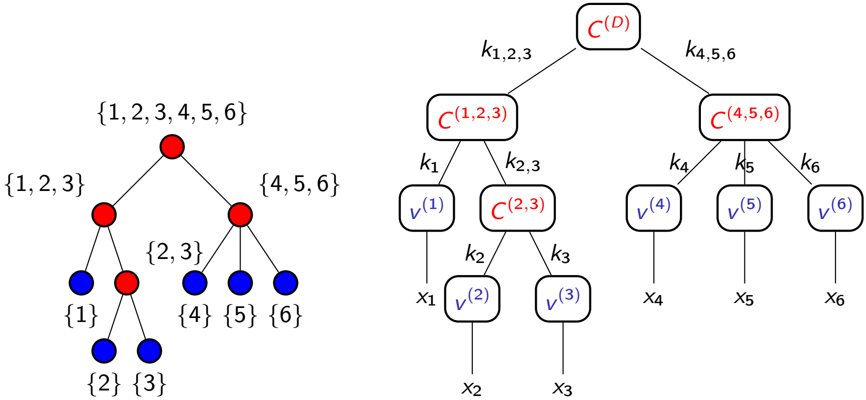




Tree-based tensor format

Finally, the tensor v admits the representation

$$v(x) = \sum_{\substack{1 \le k_\beta \le r_\beta \\ \beta \in T}} \prod_{\alpha \in T \setminus \mathcal{L}(T)} C^{(\alpha)}_{(k_\beta)_{\beta \in S(\alpha)}, k_\alpha} \prod_{\nu \in \mathcal{L}(T)} v^{(\nu)}_{k_\nu}(x_\nu)$$



Tree-based tensor format

Given bases $\{\phi^{\alpha}_{i_\alpha}(x_\alpha)\}_{i_\alpha \in I^\alpha}$ of functions for the spaces V_α for $\alpha \in \mathcal{L}(T)$,

$$v(x) = \sum_{\substack{1 \le k_\beta \le r_\beta \\ \beta \in T}} \prod_{\alpha \in T \setminus \mathcal{L}(T)} C^{(\alpha)}_{(k_\beta)_{\beta \in S(\alpha)}, k_\alpha} \prod_{\alpha \in \mathcal{L}(T)} \underbrace{\Big(\sum_{i_\alpha \in I^\alpha} C^{(\alpha)}_{i_\alpha, k_\alpha} \phi^{(\alpha)}_{i_\alpha}(x_\alpha) \Big)}_{v^{(\alpha)}_{k_\alpha}(x_\alpha)}$$

or equivalently

$$v(x) = \sum_{i_1 \in I^1} \dots \sum_{i_d \in I^d} a_{i_1, \dots, i_d} \phi_{i_1}(x_1) \dots \phi_{i_d}(x_d)$$

where $a \in \mathbb{R}^{I^1 \times \dots \times I^d}$ is such that

$$a_{i_1, \dots, i_d} = \sum_{\substack{1 \le k_\beta \le r_\beta \\ \beta \in T}} \prod_{\alpha \in T \setminus \mathcal{L}(T)} C^{(\alpha)}_{(k_\beta)_{\beta \in S(\alpha)}, k_\alpha} \prod_{\alpha \in \mathcal{L}(T)} C^{(\alpha)}_{i_\alpha, k_\alpha}$$

Tree-based tensor format as a tensor network

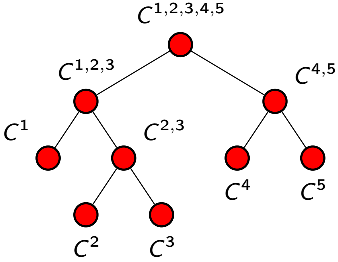
The parameters $\{C^{(\alpha)}\}_{\alpha \in T}$ form a tree network of low-order tensors such that

$$C^{(\alpha)} \in \mathbb{R}^{\# I^\alpha \times r_\alpha}$$

for a leaf node α, and

$$C^{(\alpha)} \in \mathbb{R}^{r_{\beta_1} \times \dots r_{\beta_s} \times r_\alpha}$$

for an interior node α with children $S(\alpha) = \{\beta_1, \dots, \beta_s\}$.



Tree-based tensor format

The storage complexity for the representation of a tensor u in $\mathcal{T}_r^T$ is

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

If $r_\alpha = O(R)$ and $\# I^\alpha = O(N)$,

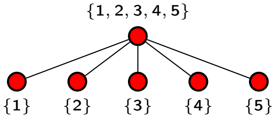
$$C(T,r) = O(dNR + (\#T - d - 1)R^{s+1} + R^s),$$

where $s = \max_{\alpha \in T \setminus \mathcal{L}(T)} \# S(\alpha)$ is the arity of the tree.

Tree-based tensor format

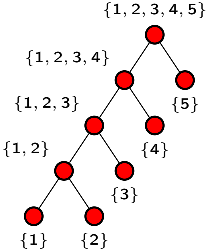
- For a trivial tree with one level (Tucker format), $s = d$, $\#T = d + 1$, and

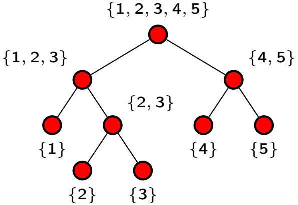
$$C(T, r) = O(dNR + R^d)$$



- For any binary tree such as a linear binary tree (Tensor Train Tucker format) or a balanced binary tree (Hierarchical Tucker format), $s = 2$, $\#T = 2d - 1$, and

$$C(T, r) = O(dNR + (d-2)R^3 + R^2)$$





- For an arbitrary tree with arity $s = O(1)$, since $\#T = O(d)$,

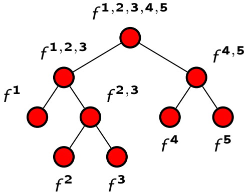
$$C(T, r) = O(dNR + dR^{s+1})$$

Tree-based tensor format as a deep network

By identifying a tensor $C^{(\alpha)} \in \mathbb{R}^{n_1 \times \ldots \times n_s \times r_\alpha}$ with a $\mathbb{R}^{r_\alpha}$-valued multilinear function

$$f^{(\alpha)} : \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_s} \to \mathbb{R}^{r_\alpha},$$

a function v in $\mathcal{T}_r^T$ admits a representation as a tree-structured composition of multilinear functions $\{f^{(\alpha)}\}_{\alpha \in T}$.



$$v(x) = f^D(f^{1,2,3}(f^1(\Phi^1(x_1)), f^{2,3}(f^2(\Phi^2(x_2)), f^3(\Phi^3(x_3))), f^{4,5}(f^4(\Phi^4(x_4)), f^5(\Phi^5(x_5))))$$

where $\Phi^\nu(x_\nu) = (\phi^\nu_{i_\nu}(x_\nu))_{i_\nu \in I^\nu} \in \mathbb{R}^{\# I^\nu}$.

It corresponds to a deep network with a sparse architecture (given by T), a depth bounded by $d-1$, and width at level ℓ related to the α-ranks of the nodes α of level ℓ.

Universality result

For any fixed T (a tree or not), $\mathcal{T}_r^T$ is a universal approximation tool since

$$\bigcup_{r \geq 0} \mathcal{T}_r^T \text{ is dense in } V$$

or equal to V if $\dim(V) < \infty$.

Therefore, for any $u \in V$, we can find a sequence $\{u_r\}_{r \geq 1}$ with $u_r \in \mathcal{T}_r^T$ which converges to u.

Canonical versus tree-based format

Consider a finite dimensional tensor space $V = V^1 \otimes \ldots \otimes V^d$ with $\dim(V_\nu) = \mathbb{R}^n$, which is identified with $\mathbb{R}^{n\times\ldots\times n}$. Denote by $\mathcal{T}_r^T = \{v : \text{rank}_\alpha(v) \leq r, \alpha \in T\}$.

- From canonical format to tree-based format.
  For any v in V and any $\alpha \subset D$, the α-rank is bounded by the canonical rank:

  $$\text{rank}_\alpha(v) \leq \text{rank}(v).$$

  Therefore, for any tree T,

  $$\mathcal{R}_r \subset \mathcal{T}_r^T,$$

  so that an element in $\mathcal{R}_r$ with storage complexity $O(dnr)$ admits a representation in $\mathcal{T}_r^T$ with a storage complexity $O(dnr + dr^{s+1})$ where s is the arity of the tree T.

- From tree-based format to canonical format. For a balanced or linear binary tree, the subset

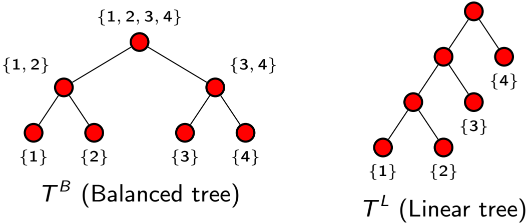
  $$S = \{v \in \mathcal{T}_r^T : \text{rank}(v) < q^{d/2}\}, \quad q = \min\{n, r\},$$

  is of Lebesgue measure 0.

  Then a typical element $v \in \mathcal{T}_r^T$ with storage complexity of order $dnr + dr^3$ admits a representation in canonical format with a storage complexity of order $dnq^{d/2}$.

Influence of the tree

- For some functions, the choice of tree is not crucial. For example, an additive function
$$u_1(x_1) + \ldots + u_d(x_d)$$
has α-ranks equal to 2 whatever $\alpha \subset D$.
- But usually, different trees lead to different complexities of representations.



- If $\text{rank}_{T^L}(u) \leq r$ then $\text{rank}_{T^B}(u) \leq r^2$
- If $\text{rank}_{T^B}(u) \leq r$ then $\text{rank}_{T^L}(u) \leq r^{\log_2(d)/2}$

Influence of the tree

Given a tree T and a permutation σ of $D = \{1, \ldots, d\}$, we define a tree T_σ

$$T_\sigma = \{\sigma(\alpha) : \alpha \in T\}$$

having the same structure as T but different nodes.

T $\qquad$ T_σ with $\sigma = (3, 1, 2, 4)$

If $\text{rank}_T(u) \leq r$ then $\text{rank}_{T_\sigma}(u)$ typically depends on d.

Influence of the tree

- Consider the Henon-Heiles potential

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

  Using a linear tree $T = \{\{1\},\{2\},\dots,\{d\},\{1,2\},\{1,2,3\},\dots,\{1,\dots,d-1\},D\}$,

$$\mathrm{rank}_T(u) \leq 4, \quad storage(u) = O(d)$$

  but for the permutation

$$\sigma = (1,3,\dots,d-1,2,4,\dots,d) \qquad (\star)$$

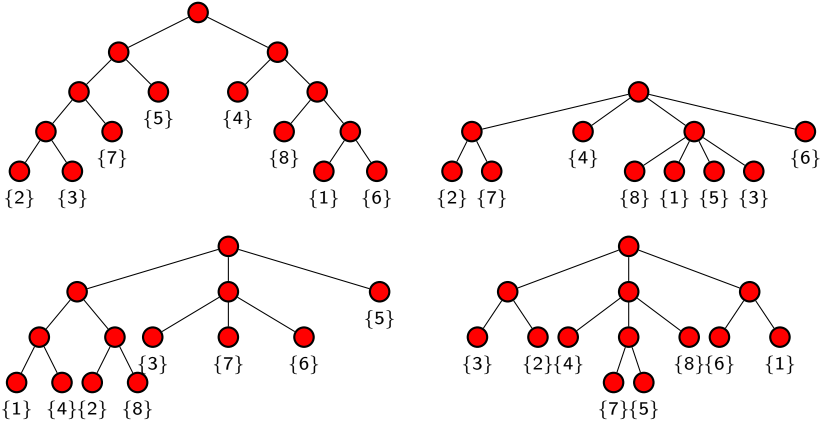
  and the corresponding linear tree T_σ,

$$\mathrm{rank}_{T_\sigma}(u) \leq 2d+1, \quad storage(u) = O(d^3).$$

- For a typical tensor in $\mathcal{T}_r^T$ with T a binary tree, its representation in tree based format with tree T_σ, with σ as in $(\star)$, has a complexity scaling exponentially with d.

How to choose a good tree ?

A combinatorial problem...



References I

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

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.