CEMRACS,
July 19-23, 2021

Approximation and learning with tensor networks

Anthony Nouy

Centrale Nantes, Laboratoire de Mathématiques Jean Leray

Tensor networks

Tensor networks are prominent tools for the representation or approximation of multivariate functions or multidimensional arrays.

- A long history in quantum physics.
- Tree tensor networks appeared independently in numerical analysis, as an extension of low-rank decompositions to high-order tensors.
- Growing use in statistics, data science and probabilistic modelling.

Computing with tensor networks

- For the approximation of a known tensor u with respect to a certain norm, we aim at finding a tensor network v with low complexity that minimizes

$$\|u - v\|.$$

  Here, the aim is the compression of u or the extraction of information from u (data analysis).

- For the solution of an equation $Au = b$ (e.g. in quantum physics, uncertainty quantification, stochastic calculus), we aim at finding a tensor network v with low complexity that minimizes some distance to u, e.g. some residual norm

$$\|Av - b\|.$$

  The aim is here to obtain an approximation of the solution u with a low computational complexity.

- In tensor completion, knowing some entries $(u(i))_{i\in\Omega}$ of a multidimensional array, we try to find a tensor network that suitably fit the data, e.g. by minimizing

$$\sum_{i\in\Omega} |u(i) - v(i)|^2,$$

The aim is here to recover (or complete) a tensor from partial information, by exploiting low-rank structures of the tensor.

- For inverse problems, we want to identify a tensor u from indirect and partial observations $y = Au$ or $y = Au + \epsilon$, where A is an observation map. We try to find a tensor network that suitably fit the observations by minimizing some distance

$$d(y, Av)$$

between observations and the prediction Av.
Exploiting low-rank structures in u allows to reduce the number of parameters to estimate and possibly makes the problem well-posed.

- Approximating a function u from evaluations $u(x^k)$ at some points x^k, e.g. by minimizing

$$\frac{1}{n}\sum_{k=1}^{n}(u(x^k) - v(x^k))^2.$$

Depending on the context, points can be given or chosen. Here we want to exploit at best the given evaluations or obtain a good approximation using a small number of evaluations.

Computing with tensor networks

- In supervised or unsupervised learning, tensor networks are used as a powerful model class for high-dimensional tasks.
- Supervised learning of the relation between a random variable Y and another random variable X. Introduction of a risk functional

$$\mathcal{R}(v) = \mathbb{E}(\ell(Y, v(X)))$$

that quantifies some expected distance between observations Y and predictions $v(X)$. In practice, using samples $\{(x_k, y_k)\}_{k=1}^{n}$, we optimize an empirical risk

$$\frac{1}{n}\sum_{k=1}^{n} \ell(y^k, v(x^k))$$

- Estimation of the density of a random variable X from samples $\{x_k\}_{k=1}^{n}$. If the density u minimizes some functional

$$\mathcal{R}(v) = \mathbb{E}(\gamma(v, X)),$$

we minimize in practice an empirical risk

$$\frac{1}{n}\sum_{k=1}^{n} \gamma(v, x^k)$$

Outline of the course

- Part I: Tensors, ranks and tensor networks
- Part II: Approximation theory of tree tensor networks
- Part III: Computational aspects

CEMRACS,
July 19-23, 2021

Approximation and learning with tensor networks

Part I: Tensors, ranks and tensor networks

Anthony Nouy

Centrale Nantes, Laboratoire de Mathématiques Jean Leray

Outline

Outline

Algebraic tensors

Given d index sets $I_\nu = \{1, \dots, N_\nu\}$, $1 \le \nu \le d$, 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.

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.

It can be represented by a multidimensional array

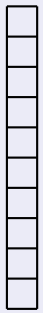
$$(v_i)_{i \in I} = (v_{i_1,\ldots,i_d})_{i_1 \in I_1,\ldots,i_d \in I_d}$$

that contains the coefficients of v in the canonical basis of $\mathbb{R}^I$, also denoted

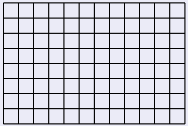
$$v(i) = v(i_1, \ldots, i_d).$$

The order d is the number of dimensions, also known as ways or modes.

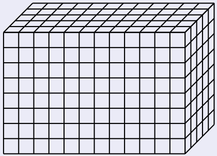
$d = 1$



$d = 2$



$d = 3$



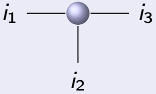
Tensor diagram notations

A tensor is represented by a solid shape and tensor indices are notated by lines emanating from this shape.

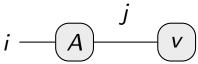
$d = 1$

$d = 2$

$d = 3$



Connecting two index lines means contraction (or summation) over the corresponding indices.



$$= \sum_j A(i,j)v(j)$$

Algebraic tensors

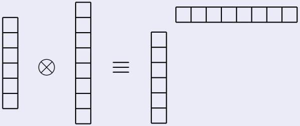
Given d vectors $v^{(\nu)} \in \mathbb{R}^{I_\nu}$, $1 \leq \nu \leq d$, the tensor product of these vectors

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

is called an elementary tensor and is such that

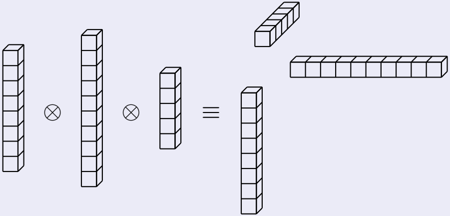
$$v(i) = v^{(1)}(i_1) \ldots v^{(d)}(i_d)$$

$d = 2$



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

$d = 3$



Algebraic tensors

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

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

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

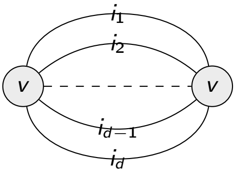
$$\|v\| = \sqrt{\sum_{i \in I} v(i)^2}$$

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 crossnorm property

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

In tensor diagram notations

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



Tensor product of functions

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

$\mathcal{X}_\nu$ can be (a subset of) $\mathbb{R}$, $\mathbb{C}$, $\mathbb{N}$, $\mathbb{Z}$, or a set of vectors, sequences, graphs, images...

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

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

Tensor product of functions

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

$\mathcal{X}_\nu$ can be (a subset of) $\mathbb{R}$, $\mathbb{C}$, $\mathbb{N}$, $\mathbb{Z}$, or a set of vectors, sequences, graphs, images...

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

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

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

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

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

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

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.

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

Example

A polynomial $\sum_i a_i x^i$ with $x^i = x_1^{i_1} \ldots x_d^{i_d}$.

Tensor product of functions

The algebraic tensor product of spaces V_ν is defined as

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

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

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

Example

A polynomial $\sum_i a_i x^i$ with $x^i = x_1^{i_1} \ldots x_d^{i_d}$.

Up to a formal definition of the tensor product $\otimes$, the above construction can be extended to more general vector spaces (not only spaces of functions), including spaces of matrices or operators.

Infinite dimensional tensor spaces

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.

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 $L^p_{\mu}(\mathcal{X}_1 \times \ldots \times \mathcal{X}_d)$.

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

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

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

Tensor product basis

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

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

A tensor $v \in V$ admits a decomposition

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

and v can be identified with the set of its coefficients

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

Hilbert tensor spaces

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

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

is such that

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

Therefore, the map

$$a \mapsto \sum_{i\in I} a_i \phi_i$$

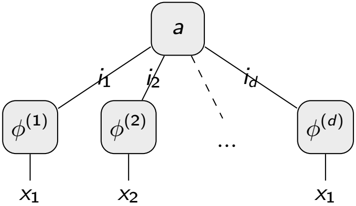
defines a linear isometry from $\ell_2(I)$ to V for finite dimensional spaces, and between $\ell_2(I)$ and $\overline{V}^{\|\cdot\|}$ for infinite dimensional spaces.

If V is a space of functions defined on $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d$, we introduce the feature map $\phi^{(\nu)}(x_\nu) = (\phi^{(\nu)}_{i_\nu}(x_\nu))_{i_\nu \in I_\nu} \in \mathbb{R}^{I_\nu}$ and the tensor product feature map $\Phi : \mathcal{X} \to \mathbb{R}^I$ such that

$$\Phi(x) = \phi^{(1)}(x_1) \otimes \ldots \otimes \phi^{(d)}(x_d) \in \mathbb{R}^I$$

and a tensor v in V admits the representation

$$v(x) = (a, \Phi(x)) =$$



Outline

Rank of order-two tensors

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

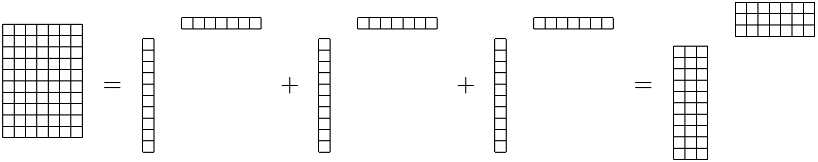
$$u = \sum_{k=1}^{r} v_k \otimes w_k$$

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

$$u = \sum_{k=1}^{\text{rank}(u)} \sigma_k v_k w_k^T = \text{VSW}^T$$

with orthogonal matrices V and W, and a diagonal matrix 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 or spectral 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 (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

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

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

- The application $v \mapsto \mathrm{rank}(v)$ 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 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 $v_k^{(\nu)} \in V_\nu$.

Example

- A monomial $x^i = x_1^{i_1} \ldots x_d^{i_d}$ has rank 1.
- A polynomial $\sum_{i \in \Lambda} a_i x^i$ has rank $\#\Lambda$.
- A Gaussian function $\exp(-\alpha \|x - a\|_2^2) = \prod_{i=1}^{d} \exp(-\alpha(x_i - a_i)^2)$ has rank 1.
- The function $\frac{1}{\|x\|_2}$ has infinite rank.

Canonical format

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

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

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

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.

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.

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.

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.

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 $n \to \infty$.

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 $n \to \infty$.

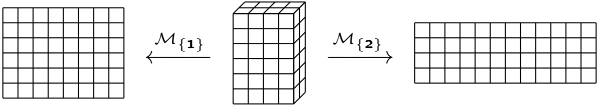
- The consequence is that for most problems involving approximation in canonical format $\mathcal{R}_r$, there is no robust method when $d > 2$.

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 (or unfolding) 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 $\mathrm{rank}_\alpha(u) = \mathrm{rank}_{\alpha^c}(u)$.

A multivariate function $u(x_1, \ldots, x_d)$ with $\mathrm{rank}_\alpha(u) \leq r_\alpha$ 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

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

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, $\mathrm{rank}_\alpha(u) = 1$.

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, $\text{rank}_\alpha(u) = 1$.

- $u(x) = \sum_{k=1}^r u_k^1(x_1)\dots u_k^d(x_d)$ 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, $\text{rank}_\alpha(u) \le r$, 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:

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

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, $\text{rank}_\alpha(u) = 1$.

- $u(x) = \sum_{k=1}^r u_k^1(x_1) \dots u_k^d(x_d)$ 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:

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

- $u(x) = u^1(x_1) + \dots + u^d(x_d)$ 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, $\text{rank}_\alpha(u) \leq 2$.

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, $\text{rank}_\alpha(u) = 1$.

- $u(x) = \sum_{k=1}^r u_k^1(x_1)\dots u_k^d(x_d)$ 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, $\text{rank}_\alpha(u) \le r$, 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:

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

- $u(x) = u^1(x_1) + \dots + u^d(x_d)$ 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, $\text{rank}_\alpha(u) \le 2$.

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

α-ranks and minimal subspaces

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

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

If u admits the representation

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

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

α-ranks and minimal subspaces

For any partition $\{\alpha_1, \ldots, \alpha_m\}$ of D, an algebraic tensor u is such that

$$u \in U_{\alpha_1}^{min}(u) \otimes \ldots \otimes U_{\alpha_m}^{min}(u)$$

Moreover, for any $\alpha \subset D$ and any partition $\{\beta_1, \ldots, \beta_s\}$ of α, it holds

$$U_{\alpha}^{min}(u) \subset U_{\beta_1}^{min}(u) \otimes \ldots \otimes U_{\beta_s}^{min}(u)$$

that implies

$$\mathrm{rank}_{\alpha}(u) \leq \prod_{k=1}^{s} \mathrm{rank}_{\beta_k}(u)$$

Also, for any $p \in \{1, ..., s\}$

$$\mathrm{rank}_{\beta_p}(u) \leq \mathrm{rank}_{\alpha}(u) \prod_{\substack{k=1 \\ k \neq p}}^{s} \mathrm{rank}_{\beta_k}(u)$$

Example

The function

$$u(x_1, x_2, x_3) = \cos(x_1 + x_2) + x_1(x_2 + 2x_3) = \cos(x_1)\cos(x_2) - \sin(x_1)\sin(x_2) + x_1x_2 + 2x_1x_3$$

has for minimal subspaces and ranks

- $U_1^{min}(u) = span\{\cos(x_1), \sin(x_1), x_1\}, \quad r_1 = 3$
- $U_2^{min}(u) = span\{\cos(x_2), \sin(x_2), x_2\}, \quad r_2 = 3$
- $U_3^{min}(u) = span\{1, x_3\}, \quad r_3 = 2$
- $U_{1,2}^{min}(u) = span\{\cos(x_1 + x_2), x_1x_2, x_1\}, \quad r_{1,2} = 3$
- $U_{2,3}^{min}(u) = span\{\cos(x_2), \sin(x_2), x_2 + 2x_3\}, \quad r_{2,3} = 3$
- $U_{1,3}^{min}(u) = span\{\cos(x_1), \sin(x_1), x_1, x_1x_3\}, \quad r_{1,3} = 4$

In particular, we can check that

$$U_{1,3}^{min}(u) \subset U_1^{min}(u) \otimes U_3^{min}(u) = span\{\cos(x_1), \sin(x_1), x_1, \cos(x_1)x_3, \sin(x_1)x_3, x_1x_3\}$$

$$r_{1,3} \leq r_1 r_3, \quad r_1 \leq r_{1,3} r_3, \quad r_3 \leq r_{1,3} r_1$$

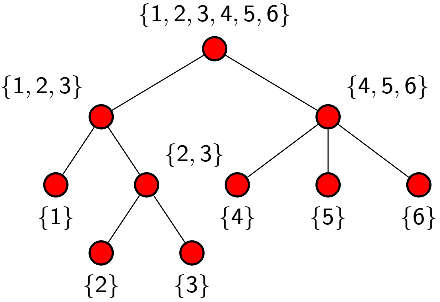
Outline

Tree-based tensor format

Tree-based (Hierarchical) tensor formats [Hackbusch-Kuhn'09] are subsets of tensors

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

where $r = (r_\alpha)_{\alpha \in T}$ and where T is a dimension partition tree T over $D = \{1, \dots, d\}$, with root D and leaves $\mathcal{L}(T) = \{\{\nu\} : 1 \leq \nu \leq d\}$. All nodes in T are non empty subsets of D. The set of children of $\alpha \in T$ is either empty (for a leaf node) or is a nontrivial partition of α (for an interior node).



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

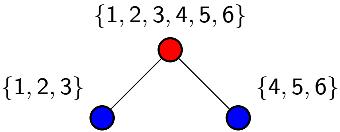
By convention, $\mathrm{rank}_D(v) = 1$.

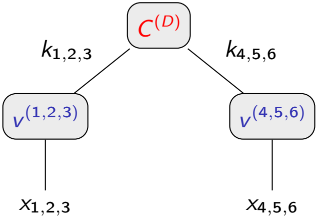
Tree-based tensor format

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 $U^{min}_\beta(v)$.

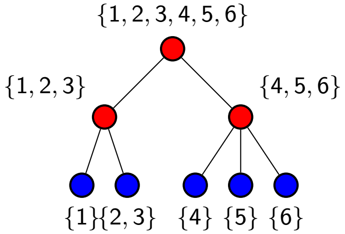


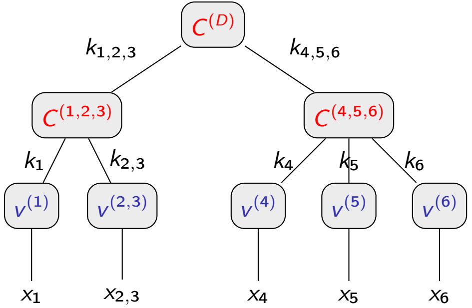


Tree-based tensor format

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

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



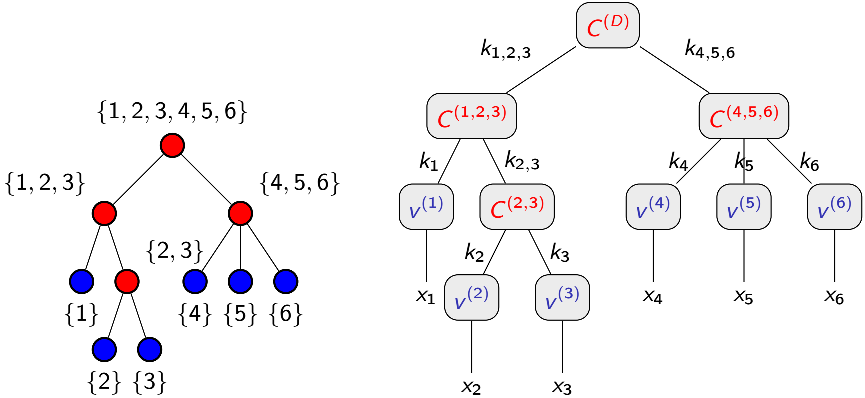


Tree-based tensor format as a tree tensor network

Finally, the tensor v admits the representation

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

where the parameters C^α and $v^{(\nu)}$ form a tree tensor network.

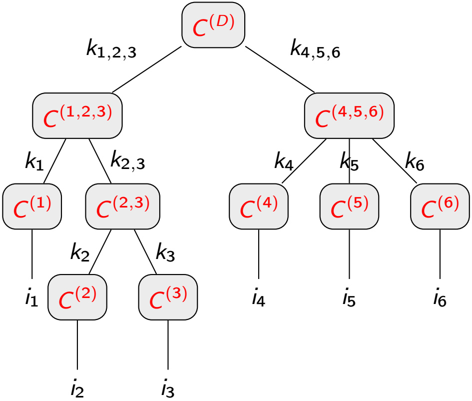


Tree-based tensor format as a tree tensor network

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

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

with $a(i_1, \dots, i_d) = \sum_{\substack{1 \le k_\beta \le r_\beta \\ \beta \in T}} \prod_{\alpha \in T \setminus \mathcal{L}(T)} C^{(\alpha)}_{(k_\beta)_{\beta \in S(\alpha)}, k_\alpha} \prod_{\alpha \in \mathcal{L}(T)} C^{(\alpha)}_{i_\alpha, k_\alpha}$ or using tensor diagram notations

$a(i_1, \dots, i_d) =$ 

Representation complexity

The representation complexity for the representation of a tensor in $\mathcal{T}_r^T(V)$ 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 $\# I^\alpha = O(N)$,

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

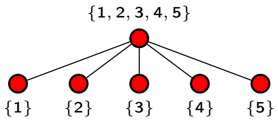
where $s = \max_{\alpha \in T \setminus \mathcal{L}(T)} \# S(\alpha)$ is the arity of the tree.

Since $\# T \leq 2d + 1$,

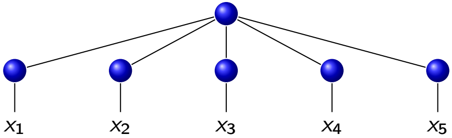
$$C(T,r) = O(dNR + dR^{s+1} + R^s)$$

Tucker format

The Tucker format [Hitchcock'27] corresponds to a trivial tree with one level, arity $s = d$, $\#T = d + 1$,



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

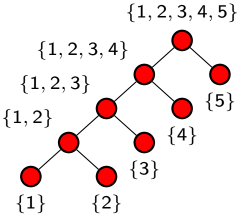


The representation complexity

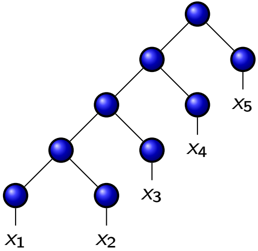
$$C(T, r) = O(dNR + R^d)$$

Tensor train Tucker format

The tensor train Tucker format corresponds to a linear binary tree



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

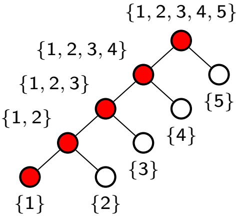


The representation complexity $C(T, r) = O(dNR + (d-2)R^3 + R^2)$.

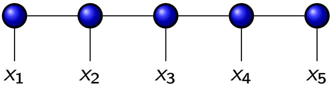
Tensor train format

The tensor train format [Oseledets-Tyrtyshnikov'09] was discovered independently in quantum physics [Baxter'68 , Affleck'87] and coined Matrix Product State (MPS).
It corresponds to a degenerate tree-based format where T is a subset of a linear tree

$$T = \{\{1\}, \{1,2\}, \{1,2,3\}, \ldots, \{1,\ldots,d\}\}$$



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



or explicitly

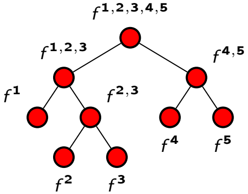
$$u(x_1,\ldots,x_d) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{d-1}=1}^{r_{1,\ldots,d-1}} v^{(1)}_{k_1}(x_1) v^{(2)}_{k_1,k_2}(x_2) \ldots v^{(d-1)}_{k_{d-2},k_{d-1}}(x_{d-1}) v^{(d)}_{k_{d-1}}(x_d)$$

The complexity is $C(T,r) = O(dNR^2)$.

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

Tree tensor networks as feed-forward neural networks

It corresponds to a sum-product feed forward neural network with a sparse architecture (given by T), a number of hidden layers equal to $depth(T) + 1$ (including a featuring layer), and width at level ℓ related to the α-ranks of the nodes α of level ℓ.

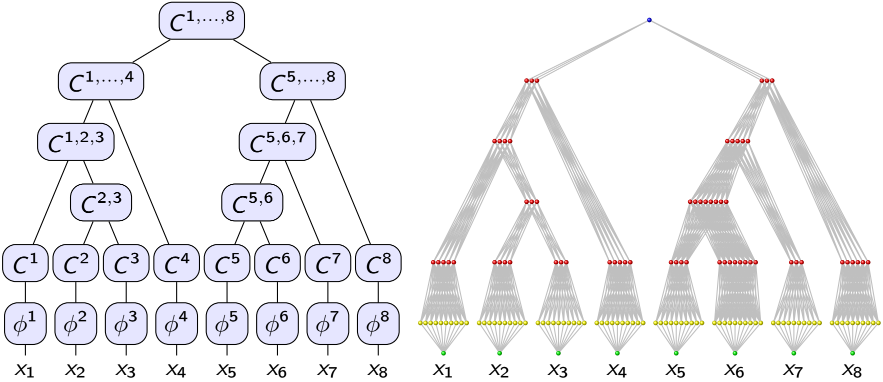


Figure: Tree tensor network and corresponding feed-forward sum-product neural network with 10 features per variable x_ν (right)

Properties of tree-based tensor formats

Many favorable properties inherited from the matrix case.

- Complexity is linear in d and polynomial in the rank for storage, evaluation, differentiation, integration...

Properties of tree-based tensor formats

Many favorable properties inherited from the matrix case.

- Complexity is linear in d and polynomial in the rank for storage, evaluation, differentiation, integration...

- Not so nonlinear approximation tool. A tensor u in tree-based format admits a multilinear parametrization with parameters $(C_\alpha)_{\alpha \in T}$ forming a tree tensor network, i.e.

  $$u = R((C_\alpha)_{\alpha \in T})$$

  with R a multilinear map.

Properties of tree-based tensor formats

Many favorable properties inherited from the matrix case.

- Complexity is linear in d and polynomial in the rank for storage, evaluation, differentiation, integration...
- Not so nonlinear approximation tool. A tensor u in tree-based format admits a multilinear parametrization with parameters $(C_\alpha)_{\alpha \in T}$ forming a tree tensor network, i.e.

  $$u = R((C_\alpha)_{\alpha \in T})$$

  with R a multilinear map.
- Topological properties ensure the well-posedness of optimization problems and existence of stable algorithms

Properties of tree-based tensor formats

Many favorable properties inherited from the matrix case.

- Complexity is linear in d and polynomial in the rank for storage, evaluation, differentiation, integration...
- Not so nonlinear approximation tool. A tensor u in tree-based format admits a multilinear parametrization with parameters $(C_\alpha)_{\alpha \in T}$ forming a tree tensor network, i.e.

$$u = R((C_\alpha)_{\alpha \in T})$$

  with R a multilinear map.
- Topological properties ensure the well-posedness of optimization problems and existence of stable algorithms
- Geometrical properties can be exploited for optimization and dynamical approximation.

Properties of tree-based tensor formats

Many favorable properties inherited from the matrix case.

- Complexity is linear in d and polynomial in the rank for storage, evaluation, differentiation, integration...
- Not so nonlinear approximation tool. A tensor u in tree-based format admits a multilinear parametrization with parameters $(C_\alpha)_{\alpha \in T}$ forming a tree tensor network, i.e.

$$u = R((C_\alpha)_{\alpha \in T})$$

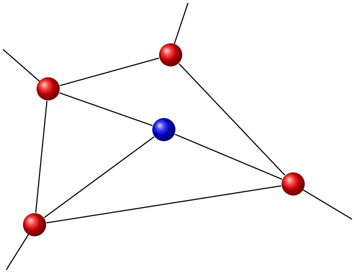
  with R a multilinear map.
- Topological properties ensure the well-posedness of optimization problems and existence of stable algorithms
- Geometrical properties can be exploited for optimization and dynamical approximation.
- Possible extensions of singular value decomposition for u in a Hilbert tensor space V, and a way to obtain approximations u_r in $\mathcal{T}_r^T(V)$ such that

$$\|u - u_r\| \leq C_d \inf_{v \in \mathcal{T}_r^T(V)} \|u - v\|$$

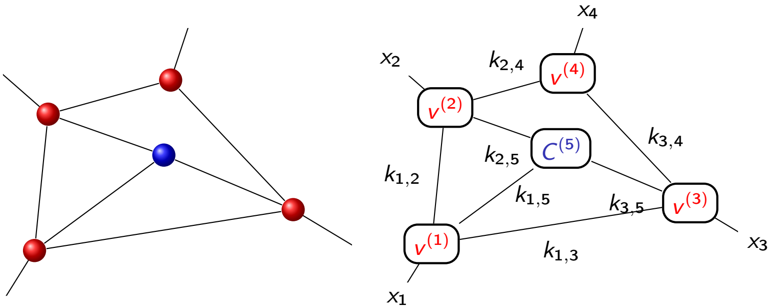
  with $C_d \sim \sqrt{d}$.

General tensor networks

More general tensor networks are associated with graphs $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with nodes (vertices) $\mathcal{N}$ and edges $\mathcal{E}$, d of the nodes being associated with variables x_ν, $1 \leq \nu \leq d$



More general tensor networks are associated with graphs $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with nodes (vertices) $\mathcal{N}$ and edges $\mathcal{E}$, d of the nodes being associated with variables x_ν, $1 \le \nu \le d$



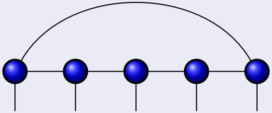
They have a multilinear parametrization of the form

$$v(x_1, \dots, x_d) = \sum_{\substack{1 \le k_e \le r_e \\ e \in \mathcal{E}}} \prod_{\nu=1}^{d} v^{(\nu)}(x_\nu, (k_e)_{e \in E_\nu}) \prod_{\nu=d+1}^{N} C^{(\nu)}((k_e)_{e \in E_\nu})$$

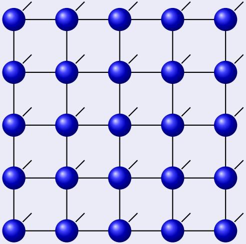
Tree tensor networks is a particular case where $\mathcal{G}$ is a tree.

Examples of tensor networks

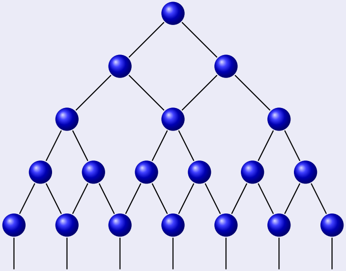
Tensor ring (MPS with periodic boundary conditions)



PEPS



MERA



General tensor networks

When the graph contains cycles,

- integers r_e (bond dimensions) may not have an interpretation as α-ranks,
- no notion of singular value decomposition,
- loss of nice geometrical and topological properties,
- computational complexity increases,
- but yet powerful for some high-dimensional applications.

Outline

Tensorization of vectors

A vector $v \in \mathbb{R}^N$ with $N = b^L$ can be identified with a tensor of order L

$$\boldsymbol{v} \in \mathbb{R}^b \otimes \ldots \otimes \mathbb{R}^b = (\mathbb{R}^b)^{\otimes L}$$

such that for $i \in \{0, \ldots, N-1\}$

$$v(i) = \boldsymbol{v}(i_1, \ldots, i_L)$$

where $(i_1, \ldots, i_L) \in \{0, \ldots, b-1\}$ are the integers of the representation of i in base b

$$i = \sum_{k=1}^{d} i_k b^{L-k} = [i_1, \ldots, i_L]_b.$$

The map which associates to v its tensorization $\boldsymbol{v}$ is a linear isometry from $\ell_2(\mathbb{R}^N)$ to $\ell_2(\mathbb{R}^b)^{\otimes L}$.

Some matrix-vector operations can be efficiently implemented using tensor algebra, such as the Hadamard transform

$$H_L v \equiv (H_1 \otimes \ldots \otimes H_1)\boldsymbol{v}$$

Tensorization of tensors

A tensor $v \in \mathbb{R}^N \otimes \ldots \otimes \mathbb{R}^N = (\mathbb{R}^N)^{\otimes d}$ with $N = b^L$ can be identified with a tensor of order dL

$$\boldsymbol{v} \in (\mathbb{R}^b)^{\otimes dL}$$

with

$$v(i_1, \ldots, i_d) = \boldsymbol{v}(i_1^1, \ldots, i_1^L, \ldots, i_d^1, \ldots, i_d^L)$$

where

$$i_\nu = [i_\nu^1 \ldots i_\nu^{L_\nu}]_b$$

Other orderings of variables can be considered, such as

$$v(i_1, \ldots, i_d) = \boldsymbol{v}(i_1^1, \ldots, i_d^1, \ldots, i_1^L, \ldots, i_d^L)$$

Tensors with different dimensions can be considered, i.e.

$$v \in \mathbb{R}^{N_1} \otimes \ldots \otimes \mathbb{R}^{N_d}, \quad N_\nu = b_\nu^{L_\nu}$$

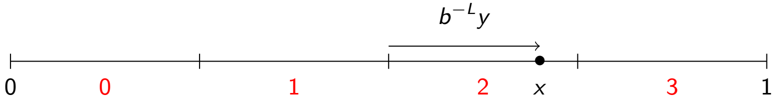
is identified with a tensor of order $\sum_{\nu=1}^d L_\nu$.

Tensorization of univariate functions

Consider a function $f \in \mathbb{R}^{[0,1)}$ defined on the interval $[0,1)$.

- For $b, L \in \mathbb{N}$, we subdivide uniformly the interval $[0,1)$ into b^L intervals. Any $x \in [0,1)$ can be written

$$x = b^{-L}(i + y), \quad i \in \{0, \dots, b^L - 1\}, \quad y \in [0,1).$$

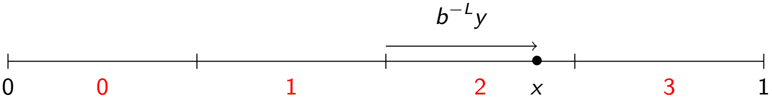


Tensorization of univariate functions

Consider a function $f \in \mathbb{R}^{[0,1)}$ defined on the interval $[0,1)$.

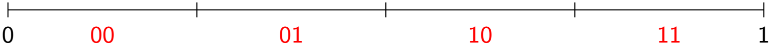
- For $b, L \in \mathbb{N}$, we subdivide uniformly the interval $[0,1)$ into b^L intervals. Any $x \in [0,1)$ can be written

$$x = b^{-L}(i + y), \quad i \in \{0, \dots, b^L - 1\}, \quad y \in [0,1).$$



- The integer i admits a representation in base b

$$i = \sum_{k=1}^{L} i_k b^{L-k} = [i_1 \dots i_L]_b, \quad i_k \in \{0, \dots, b-1\}$$

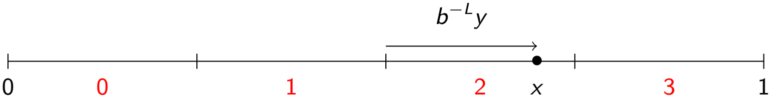


Tensorization of univariate functions

Consider a function $f \in \mathbb{R}^{[0,1)}$ defined on the interval $[0,1)$.

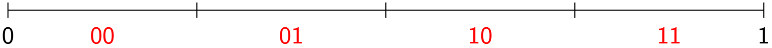
- For $b, L \in \mathbb{N}$, we subdivide uniformly the interval $[0,1)$ into b^L intervals. Any $x \in [0,1)$ can be written

$$x = b^{-L}(i + y), \quad i \in \{0, \dots, b^L - 1\}, \quad y \in [0,1).$$



- The integer i admits a representation in base b

$$i = \sum_{k=1}^{L} i_k b^{L-k} = [i_1 \dots i_L]_b, \quad i_k \in \{0, \dots, b-1\}$$

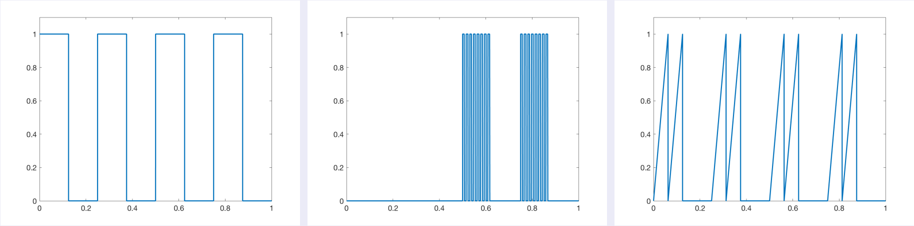


- f is thus identified with a multivariate function (tensor of order $L+1$)

$$\boldsymbol{f} \in (\mathbb{R}^b)^{\otimes L} \otimes \mathbb{R}^{[0,1)} \quad \text{such that} \quad f(x) = \boldsymbol{f}(i_1, \dots, i_L, y)$$

Tensorization of univariate functions

Examples of elementary tensors $f(x) = v^1(i_1)...v^L(i_L)v^{L+1}(y)$ $(b=2)$



(a) $\delta_{\mathbf{0}}(i_{\mathbf{3}})$ (b) $\delta_{\mathbf{1}}(i_{\mathbf{1}})\delta_{\mathbf{0}}(i_{\mathbf{3}})\delta_{\mathbf{0}}(i_{\mathbf{7}})$ (c) $\delta_{\mathbf{0}}(i_{\mathbf{1}})y$ $(L=4)$

Ranks of polynomials and splines

Polynomials

Consider a polynomial $q(x)$ of degree p. For any $\alpha \subset \{1, \dots, L\}$,

$$q(x) = q(b^{-L}(\sum_{k=1}^{L} i_k b^{L-k} + y)) = q(g(i_\alpha) + \tilde{g}(i_{\alpha^c})) = \sum_{j=0}^{p} g(i_\alpha)^j h_j(i_{\alpha^c})$$

so that $\mathrm{rank}_\alpha(\boldsymbol{q}) \leq p + 1$.

Ranks of polynomials and splines

Polynomials

Consider a polynomial $q(x)$ of degree p. For any $\alpha \subset \{1, \dots, L\}$,

$$q(x) = q(b^{-L}(\sum_{k=1}^{L} i_k b^{L-k} + y)) = q(g(i_\alpha) + \tilde{g}(i_{\alpha^c})) = \sum_{j=0}^{p} g(i_\alpha)^j h_j(i_{\alpha^c})$$

so that $\text{rank}_\alpha(\boldsymbol{q}) \leq p + 1$.

Trigonometric polynomials

The tensorization of function $\cos(\omega x + \varphi)$ at resolution L has all ranks equal to 2.

Then a trigonometric polynomial $q(x)$ of degree p is such that for any $\alpha \subset \{1, \dots, L\}$,

$$\text{rank}_\alpha(\boldsymbol{q}) \leq 2p + 1.$$

Splines

A spline φ_N of degree p over N b-adic intervals forming a partition of $[0,1)$ is such that

$$\mathrm{rank}_{\{1,\ldots,\nu\}}(\varphi_N) \leq \begin{cases} p+N, & 1 \leq \nu < \ell. \\ p+1, & \ell \leq \nu \leq L. \end{cases}$$

where $b^{-\ell}$ is the minimal length of intervals.

Tensorization of multivariate functions

A function $f(x_1, \dots, x_d)$ defined on $[0,1)^d$ can be similarly identified with a tensor of order $(L+1)d$

$$\boldsymbol{f} \in (\mathbb{R}^b)^{\otimes Ld} \otimes (\mathbb{R}^{[0,1)})^{\otimes d}$$

such that

$$f(x_1, \dots, x_d) = \boldsymbol{f}(i_1^1, \dots, i_d^1, \dots, i_1^L, \dots, i_d^L, y_1, \dots, y_d)$$

$$\text{where} \quad x_\nu = b^{-L}(\sum_{k=1}^{L} i_\nu^k b^{L-k} + y_\nu) = b^{-L}([i_\nu^1 \dots i_\nu^L]_b + y_\nu)$$

Tensorization of multivariate functions

The map $T_{b,d}$ which associates to a function f its tensorization $\boldsymbol{f}$ is a linear isometry from $L^p([0,1)^d)$ to $L^p(\{0,\dots,b-1\}^{Ld} \times [0,1)^d)$ for any $0 < p \leq \infty$.

References I

W. Hackbusch.
Tensor Spaces and Numerical Tensor Calculus, volume 56.
Springer Nature, 2019.

T. G. Kolda and B. W. Bader.
Tensor decompositions and applications.
SIAM Review, 51(3):455–500, September 2009.

L.-H. Lim.
Tensors in computations.
arXiv e-prints, page arXiv:2106.08090, June 2021.

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.

R. Orus.
A practical introduction to tensor networks: Matrix product states and projected entangled pair states.
Annals of Physics, 349:117 – 158, 2014.