Numerical methods for PDEs, IESC, Cargèse, 2016

Tensor numerical methods for high-dimensional problems

—–

Part 2

Tensors and low-rank tensor formats

Outline

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
5. Parametrization of low-rank tensor formats

Outline

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
5. Parametrization of low-rank tensor formats

Tensor product of vectors

For $I = \{1, \ldots, N\}$, an element v of the vector space $\mathbb{R}^I$ is identified with the set of its coefficients $(v_i)_{i \in I}$ on a certain basis $\{e_i\}_{i \in I}$ of $\mathbb{R}^I$,

$$v = \sum_{i \in I} v_i e_i .$$

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

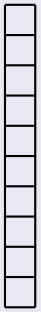
$$I = I_1 \times \ldots \times I_d .$$

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

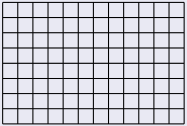
$$(v_i)_{i \in I} = (v_{i_1, \ldots, i_d})_{i_1 \in I_1, \ldots, i_d \in I_d}$$

which represents the coefficients of v on a certain basis of $\mathbb{R}^I$.

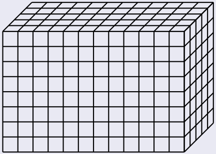
$d = 1$



$d = 2$



$d = 3$



Tensor product of vectors

The entries of the multidimensional array are equivalently denoted

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

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

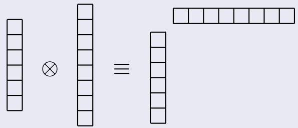
$$v := v^{(1)} \otimes \dots \otimes v^{(d)}$$

is defined by

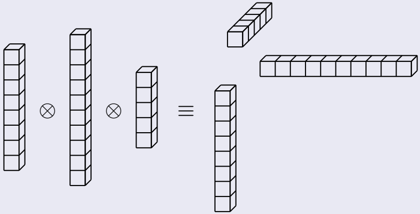
$$v(i) = v^{(1)}(i_1) \dots v^{(d)}(i_d).$$

Such a tensor is called an elementary tensor. For $d = 2$, using matrix notations, $v \otimes w$ is identified with the matrix vw^T.

$d = 2$



$d = 3$



Tensor product of vectors

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

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

Tensor product of functions

Let $\mathcal{X}_\nu \subset \mathbb{R}$, $1 \leq \nu \leq d$, be an interval and V_ν 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).

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

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

Example 1 (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 2 (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

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

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 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.

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

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

Therefore, the map Ψ which associates to a tensor $a \in \mathbb{R}^I$ the tensor $v = \Psi(a) := \sum_{i\in I} a_i \psi_i$ 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 $a \in \mathbb{R}^I = \mathbb{R}^{I_1 \times \dots \times I_d}$ 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

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

which grows exponentially with the dimension.

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

Outline

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
5. Parametrization of low-rank tensor formats

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

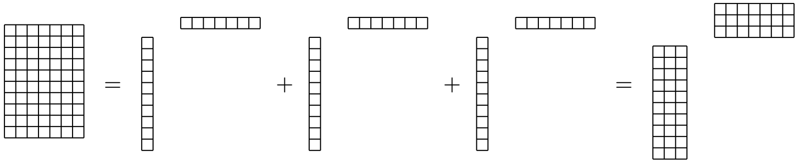
$$u = \sum_{k=1}^{r} v_k \otimes w_k$$

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



Singular value decomposition

Consider the case of a tensor space $\overline{V \otimes W}^{\|\cdot\|_\vee}$, where V and W are Hilbert spaces (e.g. spaces of functions), even infinite-dimensional, and where $\|\cdot\|_\vee$ denote the injective norm on $V \otimes W$ (the spectral norm for a matrix).

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

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

with $N = \min\{dim(V), dim(W)\} \in \mathbb{N} \cup \{\infty\}$, where v_k and w_k are orthonormal vectors.

The set of singular values of u is $\sigma(u) = \{\sigma_k(u)\}_{k=1}^{N}$.

Singular value decomposition of order-two tensors

Example 3 (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 4 (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.

Singular value decomposition

The canonical norm

$$\|u\| = \|\sigma(u)\|_2$$

is also called the Hilbert-Schmidt norm.

It is a particular case of Schatten p-norms which are defined for $1 \leq p \leq \infty$ by

$$\|u\|_{\sigma_p} = \|\sigma(u)\|_p.$$

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

$$\text{rank}(u) = \|\sigma(u)\|_0 = \#\{k : \sigma_k(u) \neq 0\}.$$

A tensor u has low rank if the vector of its singular values $\sigma(u)$ is sparse.

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.

In particular, since the application $v \mapsto \text{rank}(v)$ is lower semi-continuous, the set $\mathcal{R}_r$ is closed, which makes best approximation problems in $\mathcal{R}_r$ well posed.

Low-rank approximation of order-two tensors

For a Hilbert tensor space equipped with the canonical norm $\|\cdot\|$, the best approximation of a tensor u by an element of $\mathcal{R}_r$ is provided by the truncated singular value decomposition

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

where we only retain the r dominant singular values:

$$\min_{v \in \mathcal{R}_r} \|u - v\| = \|u - u_r\| = (\sum_{k > r+1} \sigma_k^2)^{1/2}.$$

Outline

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
5. Parametrization of low-rank tensor formats

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

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

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

A multivariate function $u(x_1, \ldots, x_d)$ with canonical rank bounded by r is such that

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

where the $v_k^{(\nu)}(x_\nu)$ are in the function space V_ν.

Canonical format

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

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

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

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

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

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

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.
- The set $\mathcal{R}_r$ is not an algebraic variety.
- The application $v \mapsto \text{rank}(v)$ is not lower semi-continuous and therefore, $\mathcal{R}_r$ is not closed. The consequence is that for most problems involving approximation in canonical format $\mathcal{R}_r$, there is no robust method when $d > 2$.

Example 5

Consider the 3-order 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 $n \to \infty$.

Outline

1 What are tensors ?

2 Low-rank order-two tensors

3 Canonical format

4 α-ranks and related low-rank tensor formats

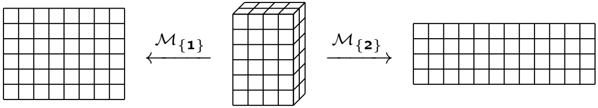
5 Parametrization of low-rank tensor formats

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

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

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 $\text{rank}_\alpha(u) = \text{rank}_{\alpha^c}(u)$.

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

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

Example 6

$u(x_1, \dots, x_d) = u_1(x_1) + \dots + u_d(x_d)$ where $u_1, \dots, u_d$ are non constant functions satisfies $\text{rank}_\alpha(u) = 2$ for all α.

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

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

Singular value decomposition

For Hilbert spaces the order-two tensor

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

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

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

A tensor u has low α-rank if the vector of its α-singular values $\sigma^\alpha(u)$ is sparse.

Subset of tensors with bounded α-rank

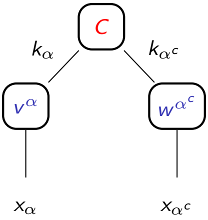
For a given subset $\alpha \subset D$, we define the subset of tensors with α-rank bounded by r_α as

$$\mathcal{T}_{r_\alpha}^{\{\alpha\}} = \{v \in V : \operatorname{rank}_\alpha(v) \leq r_\alpha\}.$$

Elements of $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$ admit the representation

$$v(x_\alpha, x_{\alpha^c}) = \sum_{k_\alpha=1}^{r_\alpha} \sum_{k_{\alpha^c}=1}^{r_\alpha} C(k_\alpha, k_{\alpha^c}) v^\alpha(x_\alpha, k_\alpha) w^{\alpha^c}(x_{\alpha^c}, k_{\alpha^c})$$

where $C \in \mathbb{R}^{r_\alpha \times r_\alpha}$ and v^α and w^{α^c} are order-two tensors.



Subset of tensors with bounded α-rank

The motivation behind the definition of tensor formats based on α-ranks is to benefit from the nice properties of the rank of order-two tensors.

The application $v \mapsto \text{rank}_\alpha(v)$ is lower semi-continuous and therefore, $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$ is closed.

For a Hilbert tensor space equipped with the canonical norm, a best approximation of a given tensor u in $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$ is provided by the truncated α-singular value decomposition where we retain the r_α largest α-singular values.

For a given $\alpha \subset D$, the determination of the α-rank of a tensor, which is equivalent to the determination of the rank an order-two tensor, is feasible.

Also, $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$ is a smooth manifold.

α-ranks and related low-rank formats

For T a collection of subsets of D, we define the T-rank of a tensor v, denoted $\text{rank}_T(u)$, 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

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

As a finite intersection of subsets $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$, $\mathcal{T}_r^T$ inherits from geometrical and topological properties of the subsets $\mathcal{T}_{r_\alpha}^{\{\alpha\}}$ which are favorable for numerical simulation. In particular, $\mathcal{T}_r^T$ is closed.

Higher-order singular value decompositions

For a Hilbert tensor space equipped with the canonical norm, and for a tree-structured set T, quasi-best approximations u_r of a given tensor u in $\mathcal{T}_r^T$ can be constructed from truncated singular value decompositions of α-matricisations of u (or of a sequence of approximations), $\alpha \in T$, such that

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

with $C(d) = O(\sqrt{d})$.

A possible algorithm

Let $T = \{\alpha_1, \ldots, \alpha_{\#T}\}$ with the sequence α_k ordered by decreasing level in the tree-structured set T.
Set $u^0 = u$, and for $k = 0, \ldots, \#T - 1$,

$$u^{k+1} \in \arg \min_{v \in \mathcal{T}_{r_{\alpha_k}}^{\alpha_k}} \|u^k - v\|,$$

and set $u_r = u^{\#T}$.

α-ranks and related low-rank formats

Different choices for T yield different tensor formats, the standard formats being

- the Tucker format,
- the Tensor Train format,
- and more general tree-based (or hierarchical) Tucker formats.

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
   - Tucker format
   - Tensor Train format
   - Tree-based (hierarchical) Tucker format
   - Tensor networks
5. Parametrization of low-rank tensor formats

Tucker format

For

$$T = \{\{1\}, \dots, \{d\}\},$$

the tuple

$$\text{rank}_T(v) = \{\text{rank}_{\{1\}}(v), \dots, \text{rank}_{\{d\}}(v)\}$$

is called the Tucker (or multilinear) rank of the tensor v.

The set of tensors with Tucker rank bounded by $r = (r_1, \dots, r_d)$, denoted

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

is such that

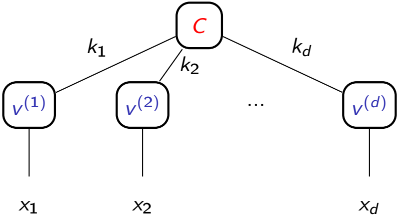
$$\mathcal{T}_r = \{v \in U_1 \otimes \dots \otimes U_d : \dim(U_\nu) = r_\nu, 1 \leq \nu \leq d\}.$$

Tucker format

A tensor in $v \in \mathcal{T}_r$ admits a representation

$$v(x_1, \ldots, x_d) = \sum_{k_1=1}^{r_1} \ldots \sum_{k_d=1}^{r_d} C(k_1, \ldots, k_d) v^{(1)}(x_1, k_1) \ldots v^{(d)}(x_d, k_d).$$

where $C \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ is an order-d tensor and the $v^{(\nu)}$ are order-two tensors.



The storage complexity is

$$\text{storage}(\mathcal{T}_r) = \prod_{\nu=1}^{d} r_\nu + \sum_{\nu=1}^{d} r_\nu \dim(V_\nu) = O(R^d + Rnd)$$

with $r_\nu = O(R)$ and $\dim(V_\nu) = O(n)$. This format still suffers from the curse of dimensionality.

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
   - Tucker format
   - Tensor Train format
   - Tree-based (hierarchical) Tucker format
   - Tensor networks
5. Parametrization of low-rank tensor formats

Tensor train format

For

$$T = \{\{1\}, \{1,2\}, \dots, \{1, \dots, d-1\}\},$$

the tuple

$$\text{rank}_T(v) = \{\text{rank}_{\{1\}}(v), \text{rank}_{\{1,2\}}(v), \dots, \text{rank}_{\{1,\dots,d-1\}}(v)\}$$

is called the TT-rank of the tensor v.

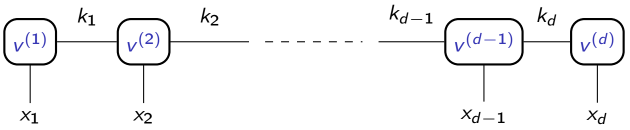
For a tuple $r = (r_1, \dots, r_{d-1})$, the set $\mathcal{T}_r^T$ of tensors with TT-rank bounded by r is denoted

$$\mathcal{TT}_r = \{v : \text{rank}_{\{1,\dots,\nu\}} = \text{rank}_{\{\nu+1,\dots,d\}}(v) \leq r_\nu, 1 \leq \nu \leq d-1\}.$$

Tensor train format

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

$$v(x) = \sum_{k_1=1}^{r_1} \dots \sum_{k_{d-1}=1}^{r_{d-1}} v^{(1)}(x_1, k_1) v^{(2)}(k_1, x_2, k_2) \dots v^{(d)}(k_{d-1}, x_d).$$



The storage complexity of an element in $\mathcal{TT}_r$ is

$$\text{storage}(\mathcal{TT}_r) = \sum_{\nu=1}^{d} r_{\nu-1} r_\nu \dim(V_\nu) = O(dnR^2)$$

with $\dim(V_\nu) = O(n)$, $r_\nu = O(R)$. Here we use the convention $r_0 = r_d = 1$.

1. What are tensors ?

2. Low-rank order-two tensors

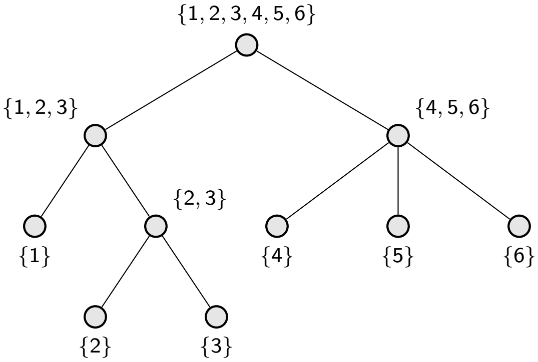
3. Canonical format

4. α-ranks and related low-rank tensor formats
   - Tucker format
   - Tensor Train format
   - Tree-based (hierarchical) Tucker format
   - Tensor networks

5. Parametrization of low-rank tensor formats

Tree-based (hierarchical) Tucker format

Tree-based (or hierarchical) Tucker formats are associated with a partition dimension tree T over $D = \{1, \dots, d\}$, with root D and leaves $\{\nu\}$, $1 \leq \nu \leq d$.



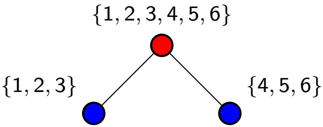
The tree-based rank of a tensor v is the tuple $\text{rank}_T(v) = (\text{rank}_\alpha(v))_{\alpha \in T}$.

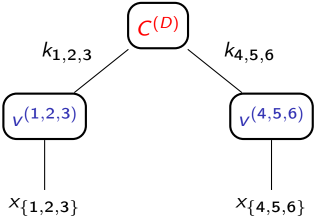
Tree-based (hierarchical) Tucker format

Let v be a tensor in $\mathcal{T}_r^T$ with $r = (r_\alpha)_{\alpha \in T}$. At the first level, v admits the representation

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

where $\{\beta_1, \dots, \beta_s\} = S(D)$ are the children of the root node D.

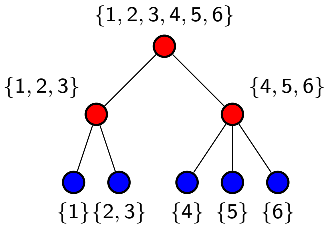


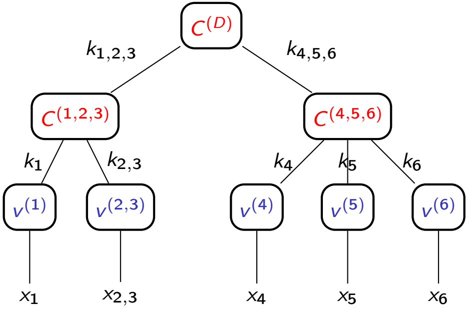


Tree-based (hierarchical) Tucker format

Then, for an interior node α of the tree, with children $S(\alpha) = \{\beta_1, \dots, \beta_s\}$, the tensor v^α admits the representation

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



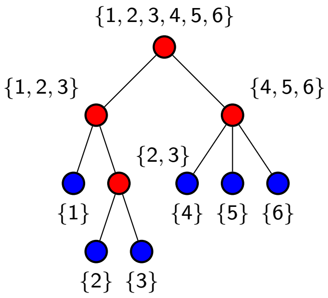


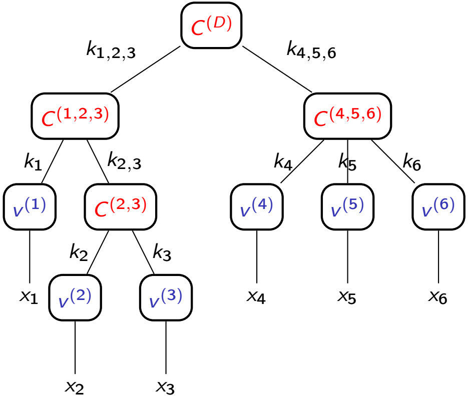
Tree-based (hierarchical) Tucker format

Finally, denoting by $\mathcal{L}(T) = \{\{\nu\} : \nu \in D\}$ the leaves of the tree, the tensor v admits the Tucker-like representation

$$v(x) = \sum_{\substack{1\le k_\nu \le r_\nu \\ \nu\in\{1,\dots,d\}}} \left(\sum_{\substack{1\le k_\alpha \le r_\alpha \\ \alpha\in T\setminus\mathcal{L}(T)}} \prod_{\mu\in T\setminus\mathcal{L}(T)} C^{(\mu)}(k_\mu,(k_\beta)_{\beta\in S(\alpha)}) \right) v^{(1)}(x_1,k_1)\dots v^{(d)}(x_d,k_d)$$

where we use the convention $C^{(D)}_{(k_\beta)_{\beta\in S(D)}} = C^{(D)}_{1,(k_\beta)_{\beta\in S(D)}}$ and $r_D = 1$.

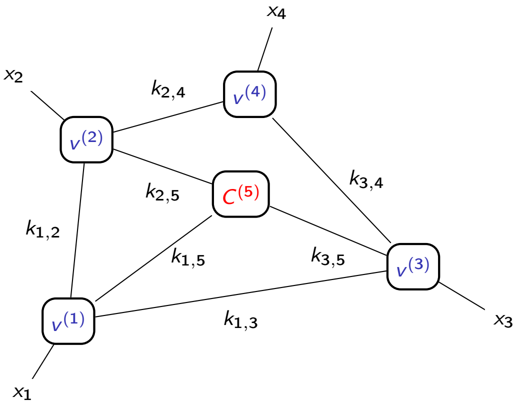




1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
   - Tucker format
   - Tensor Train format
   - Tree-based (hierarchical) Tucker format
   - Tensor networks
5. Parametrization of low-rank tensor formats

Tensor networks

More general tensor formats, called tensor networks, are associated with graphs $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with nodes $\mathcal{N}$ and edges $\mathcal{E}$.



Tree-based tensor formats are particular cases of tensor networks, called tree tensor networks, where $\mathcal{G}$ is a dimension partition tree.

Outline

1. What are tensors ?
2. Low-rank order-two tensors
3. Canonical format
4. α-ranks and related low-rank tensor formats
5. Parametrization of low-rank tensor formats

Parametrization and storage of low-rank tensor formats

Ultimately, a tensor in a certain low-rank tensor format $\mathcal{M}_r$ admits a multilinear parametrization of the form

$$v(x_1, \dots, x_d) = \sum_{k_1=1}^{r_1} \dots \sum_{k_L=1}^{r_L} \prod_{\nu=1}^{d} p^{(\nu)}(x_\nu, (k_i)_{i \in S_\nu}) \prod_{\nu=d+1}^{M} p^{(\nu)}((k_i)_{i \in S_\nu})$$

where the parameter $p^{(\nu)}$ is an element of a tensor space $P^{(\nu)}$ which depends on a subset of summation variables $(k_i)_{i \in S_\nu} := k_{S_\nu}$.

For a low-rank tensor format $\mathcal{M}_r$, there exists a multilinear map

$$\Psi : P^{(1)} \times \dots \times P^{(M)} \to V$$

which associates to a set of parameters $\{p^{(1)}, \dots, p^{(M)}\}$ the tensor

$$v = \Psi(p^{(1)}, \dots, p^{(M)}).$$

Approximation in low-rank tensor formats is the first step between linear approximation and nonlinear approximation.

Parametrization and storage of low-rank tensor formats

The storage complexity is

$$\text{storage}(\mathcal{M}_r) = \sum_{\nu=1}^{d} \dim(V_\nu) \prod_{i \in S_\nu} r_i + \sum_{\nu=d+1}^{L} \prod_{i \in S_\nu} r_i.$$

If $r_i = O(R)$, $\dim(V_\nu) = O(n)$, $\#S_\nu = O(s)$ for $\nu \leq d$ and $\#S_\nu = O(s')$ for $\nu > d$, then

$$\text{storage}(\mathcal{M}_r) = O(dnR^s + (M-d)R^{s'}).$$

The key to break the curse of dimensionality is to consider low-rank formats with $s = O(1)$ and $s' = O(1)$.

Parametrization and storage of low-rank tensor formats

Examples

- Canonical format: $L = 1$, $M = d$, $S_\nu = \{1\}$ for all ν.

$$\text{storage}(\mathcal{R}_r) = O(ndR)$$

- Tucker format: $L = d$, $M = d+1$, $S_\nu = \{\nu\}$ for $1 \leq \nu \leq d$, and $S_{d+1} = \{1, \dots, d\}$.

$$\text{storage}(\mathcal{T}_r) = O(ndR + R^d)$$

- Tensor train format: $L = d-1$, $M = d$, $S_1 = \{1\}$, $S_d = \{d-1\}$ and $S_\nu = \{\nu-1, \nu\}$ for $2 \leq \nu \leq d-1$.

$$\text{storage}(\mathcal{T}\mathcal{T}_r) = O(ndR^2)$$

- Tree-based tensor format (for a dimension partition tree T): $L = \#T - 1$, $M = \#T$, $S_\nu = \{\nu\}$ for $1 \leq \nu \leq d$ and S_ν cointains the sons of the node $\{\nu\}$ for $\nu > d$.

$$\text{storage}(\mathcal{T}_r^T) = O(ndR + dR^{k+1})$$

  where k is the maximal number of sons of the nodes ($k = 2$ for a binary tree).

- Tensor networks: arbitrary L and M and $\#\{\nu : i \in S_\nu\} = 2$ for all $1 \leq i \leq L$.