Multi-linear Algebra
– Tucker Decomposition –
Lecture 16

F. M. Faulstich

22/03/2024

Recap

Recap

- Tensors:
  high-dimensional object
- Tensor diagrams:
  graphical representation of tensor operations
- Tensor decomposition
  CP decomposition
  low-rank approximation
  $\rightarrow$ tensor rank?

CP decomposition Format

Pros:

CP decomposition Format

Pros:

- Provides a notion of tensor rank
  CP-rank
- Very good compression
- Reduction of tensor algebra

Cons:

CP decomposition Format

Pros:

- Provides a notion of tensor rank
  CP-rank
- Very good compression
- Reduction of tensor algebra

Cons:

- Hard to find
  $\rightarrow$ No systematic way to compute *a* CP decomposition
- Difficult tensor sets:

  $$\mathcal{M}_r = \left\{ \mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d} \mid \text{ CP-rank}(\mathbf{X}) = r \right\}$$

  and

  $$\mathcal{M}_{\leq r} = \left\{ \mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d} \mid \text{ CP-rank}(\mathbf{X}) \leq r \right\}$$

  are not closed.
  $\rightarrow$ Not easy to optimize on.

Tucker Decomposition – matrices

Recall the rank characterization:

“ r is the smallest number, such that there exist r-dimensional subspaces $V \subseteq \mathbb{R}^m$ and $U \subseteq \mathbb{R}^n$, such that $\mathbf{A}$ is an element of the induced tensor space $V \otimes U \subseteq \mathbb{R}^{m \times n}$ ”

Tucker Decomposition – matrices

Recall the rank characterization:

“ r is the smallest number, such that there exist r-dimensional subspaces $V \subseteq \mathbb{R}^m$ and $U \subseteq \mathbb{R}^n$, such that $\mathbf{A}$ is an element of the induced tensor space $V \otimes U \subseteq \mathbb{R}^{m \times n}$ ”

What does this mean?
Let $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ be bases of U and V, respectively.

Tucker Decomposition – matrices

Recall the rank characterization:

" r is the smallest number, such that there exist r-dimensional subspaces $V \subseteq \mathbb{R}^m$ and $U \subseteq \mathbb{R}^n$, such that $\mathbf{A}$ is an element of the induced tensor space $V \otimes U \subseteq \mathbb{R}^{m \times n}$ "

What does this mean?
Let $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ be bases of U and V, respectively. Then

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \bar{\mathbf{v}}_i$$

Tucker Decomposition – matrices

Recall the rank characterization:

“ r is the smallest number, such that there exist r-dimensional subspaces $V \subseteq \mathbb{R}^m$ and $U \subseteq \mathbb{R}^n$, such that $\mathbf{A}$ is an element of the induced tensor space $V \otimes U \subseteq \mathbb{R}^{m \times n}$ ”

What does this mean?
Let $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ be bases of U and V, respectively. Then

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \bar{\mathbf{v}}_i = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^* = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

In the matrix case:

$$V = \mathrm{Im}(\mathbf{A}) \text{ and } U = \mathbb{R}^m / \ker(\mathbf{A})$$

$\rightarrow$ their dimension **always** coincide!

Tucker Decomposition – matrices

Tucker Decomposition – Tensors

How do we generalize this to tensors $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$?

- Find minimal subspaces s.t. $\mathbf{A}$ is an element of the induced tensor space
- The dimensions of these subspaces may not be equal

[1] $\dim(U_k) = r_k$

Tucker Decomposition – Tensors

How do we generalize this to tensors $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$?

- Find minimal subspaces s.t. $\mathbf{A}$ is an element of the induced tensor space
- The dimensions of these subspaces may not be equal

Let $\{U_k\}_{k=1}^{d}$ be a collection of subsets with $U_k \subseteq \mathbb{R}^{n_k}$.

[1] $\dim(U_k) = r_k$

Tucker Decomposition – Tensors

How do we generalize this to tensors $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$?

- Find minimal subspaces s.t. $\mathbf{A}$ is an element of the induced tensor space
- The dimensions of these subspaces may not be equal

Let $\{U_k\}_{k=1}^{d}$ be a collection of subsets with $U_k \subseteq \mathbb{R}^{n_k}$.
For each subspace U_k we have an orthonormal basis[1] $\{\mathbf{u}_{k,i}\}_{i=1}^{r_k}$
A tensor $\mathbf{A} \in \bigotimes_{k=1}^{d} U_k$ can then be expressed as

$$\mathbf{A} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{C}[i_1, ..., i_d] \cdot \mathbf{u}_{1,i_1} \otimes \mathbf{u}_{2,i_2} \otimes \cdots \otimes \mathbf{u}_{d,i_1}$$

[1] $\dim(U_k) = r_k$

Tucker Decomposition – Tensors

How do we generalize this to tensors $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$?

- Find minimal subspaces s.t. $\mathbf{A}$ is an element of the induced tensor space
- The dimensions of these subspaces may not be equal

Let $\{U_k\}_{k=1}^{d}$ be a collection of subsets with $U_k \subseteq \mathbb{R}^{n_k}$.
For each subspace U_k we have an orthonormal basis[1] $\{\mathbf{u}_{k,i}\}_{i=1}^{r_k}$
A tensor $\mathbf{A} \in \bigotimes_{k=1}^{d} U_k$ can then be expressed as

$$\mathbf{A} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{C}[i_1, ..., i_d] \cdot \mathbf{u}_{1,i_1} \otimes \mathbf{u}_{2,i_2} \otimes \cdots \otimes \mathbf{u}_{d,i_1}$$

$\Rightarrow \mathbf{C} \in \mathbb{R}^{r_1 \times \ldots \times r_d}$, called the *core tensor*

[1] $\dim(U_k) = r_k$

Tucker Decomposition – Tensors

We may interpret

$$\mathbf{U}_k = [\mathbf{u}_{k,1}|...|\mathbf{u}_{k,r_k}]^\top \in \mathbb{R}^{r_k \times n_k}$$

with elements

$$\mathbf{U}_k[i_k, j] = \mathbf{u}_{k,i_k}[j] \qquad \text{for } 1 \leq i_k \leq r_k \text{ and } 1 \leq j \leq n_k$$

Tucker Decomposition – Tensors

We may interpret

$$\mathbf{U}_k = [\mathbf{u}_{k,1}|...|\mathbf{u}_{k,r_k}]^\top \in \mathbb{R}^{r_k \times n_k}$$

with elements

$$\mathbf{U}_k[i_k, j] = \mathbf{u}_{k,i_k}[j] \qquad \text{for } 1 \leq i_k \leq r_k \text{ and } 1 \leq j \leq n_k$$

Then

$$\mathbf{A} = \mathbf{C} *_1 \mathbf{U}_1 *_2 \mathbf{U}_2 ... *_d \mathbf{U}_d$$

where $*_k$ is the kth mode contraction of $\mathbf{C}$ with $\mathbf{U}_k$.

Tucker Decomposition – Tensors

We may interpret

$$\mathbf{U}_k = [\mathbf{u}_{k,1}|...|\mathbf{u}_{k,r_k}]^\top \in \mathbb{R}^{r_k \times n_k}$$

with elements

$$\mathbf{U}_k[i_k, j] = \mathbf{u}_{k,i_k}[j] \qquad \text{for } 1 \leq i_k \leq r_k \text{ and } 1 \leq j \leq n_k$$

Then

$$\mathbf{A} = \mathbf{C} *_1 \mathbf{U}_1 *_2 \mathbf{U}_2 ... *_d \mathbf{U}_d$$

where $*_k$ is the kth mode contraction of $\mathbf{C}$ with $\mathbf{U}_k$.
Elementwise:

$$\mathbf{A}[j_1, ..., j_d] = \sum_{i_1,...,i_d} \mathbf{C}[i_1, ..., i_d] \mathbf{U}_1[i_1, j_1] \cdot \mathbf{U}_2[i_2, j_2] \cdots \mathbf{U}_d[i_d, j_d]$$

Tucker Decomposition – Tensors

Tucker rank

Given a Tucker decomposition

$$\mathbf{A} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{C}[i_1, ..., i_d] \cdot \mathbf{u}_{1,i_1} \otimes \mathbf{u}_{2,i_2} \otimes \cdots \otimes \mathbf{u}_{d,i_1}$$

We call the tuple

$$\mathbf{r} = (r_1, r_2, ..., r_d)$$

the rank of the associated decomposition.

The *Tucker rank* (T-rank) $\mathbf{r}$ is the minimal d-tuple such that there exists a Tucker representation of $\mathbf{A}$.

What does minimal mean?

Tucker rank

Given a Tucker decomposition

$$\mathbf{A} = \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \mathbf{C}[i_1, ..., i_d] \cdot \mathbf{u}_{1,i_1} \otimes \mathbf{u}_{2,i_2} \otimes \cdots \otimes \mathbf{u}_{d,i_1}$$

We call the tuple

$$\mathbf{r} = (r_1, r_2, ..., r_d)$$

the rank of the associated decomposition.

The *Tucker rank* (T-rank) $\mathbf{r}$ is the minimal d-tuple such that there exists a Tucker representation of $\mathbf{A}$.

What does minimal mean? Partial order of d-tuples:
On the set of d-tuples we define the partial order $\preccurlyeq$ as

$$(x_1, ..., x_d) = \mathbf{x} \preccurlyeq \mathbf{y} = (y_1, ..., y_d) \quad \Leftrightarrow \quad x_i \leq y_i \ \ \forall i$$

Why Tucker?

- Looks a bit more complicated than CP decomposition
  $\rightarrow$ How to find the subspaces?
  $\rightarrow$ Rank definition aligns less with what we know from matrices!
- Can be computed constructively
  $\rightarrow$ higher-order SVD (HOSVD)
- Tensors of bounded or fixed Tucker rank are much nicer
  $\rightarrow$ Manifold structure $\rightarrow$ Closed set

high-order SVD

Algorithm:

Input: Target tensor $\mathbf{A}$

Output: Core tensor $\mathbf{C}$, basis matrices $\mathbf{U}_k \in \mathbb{R}^{r_k \times n_k}$ for $1 \leq k \leq d$

for $k = 1 : d$

Calculate $\tilde{\mathbf{U}}_k \mathbf{\Sigma}_k \mathbf{V}^* = \mathrm{SVD}(\mathbf{A}^{(k)})$

[Recall notation: $\mathbf{A}^{(k)}$ was the k-mode matricization]

$\mathbf{U}_k = \tilde{\mathbf{U}}_k^\top$

Calculate $\mathbf{C} = \mathbf{A} *_1 \mathbf{U}_1^\top *_2 \mathbf{U}_2^\top \ldots *_d \mathbf{U}_d^\top$

high-order SVD

The HOSVD yields Tucker decomposition of minimal rank:

Theorem
Given a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$, let $\mathbf{C} \in \mathbb{R}^{r_1 \times \ldots \times r_d}$ and $\mathbf{U}_k \in \mathbb{R}^{r_k \times n_k}$ for $1 \leq k \leq d$ be the core and basis matrices obtained with the HOSVD. Then

$$\mathbf{A} = \mathbf{C} *_1 \mathbf{U}_1 *_2 \mathbf{U}_2 \ldots *_d \mathbf{U}_d$$

is a Tucker representation of $\mathbf{A}$ of minimal rank. The obtained representation rank is also the Tucker rank of $\mathbf{A}$, is related to the matrix rank of the kth mode matricizations via

$$\text{T-rank}(\mathbf{A}) = \left(\text{rank}(\mathbf{A}^{(1)}), \text{rank}(\mathbf{A}^{(2)}), \ldots, \text{rank}(\mathbf{A}^{(d)})\right)$$

Low T-rank approximation

Ee can adjust the HOSVD to get a T-rank approximation of $\mathbf{A}$ of rank $\mathbf{r}'$

Input: Target tensor $\mathbf{A}$

Output: Core tensor $\mathbf{C}$, basis matrices $\mathbf{U}_k \in \mathbb{R}^{r_k \times n_k}$ for $1 \leq k \leq d$

Set $\mathbf{A}_0 = \mathbf{A}$

for $k = 1 : d$

Calculate rank r'_k-SVD: $\tilde{\mathbf{U}}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^\top = \mathrm{SVD}_{r'_k}(\mathbf{A}_{k-1}^{(k)})$

Set $\mathbf{A}_k^{(k)} = \boldsymbol{\Sigma}_k \mathbf{V}_k^\top$

$\mathbf{U}_k = \tilde{\mathbf{U}}_k^\top$

$\mathbf{C} = \mathbf{A} *_1 \mathbf{U}_1^\top *_2 \mathbf{U}_2^\top \ldots *_d \mathbf{U}_d^\top$

Quasi Best Approximation

There is no Eckart-Young theorem for low T-rank approximations!

Theorem:
Given a tensor $\mathbf{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$, let $\mathbf{C} \in \mathbb{R}^{r_1 \times \ldots \times r_d}$ and $\mathbf{U}_k \in \mathbb{R}^{r_k \times n_k}$ for $1 \leq k \leq d$ be the core and basis matrices obtained with the rank $\mathbf{r}'$-truncated HOSVD. These define a rank $\mathbf{r}'$ Tucker approximation

$$\mathbf{A}' = \mathbf{C} *_1 \mathbf{U}_1 *_2 \mathbf{U}_2 \ldots *_d \mathbf{U}_d.$$

The tensor $\mathbf{A}'$ is a quasi-best rank $\mathbf{r}'$ approximation to $\mathbf{A}$, i.e.,

$$\|\mathbf{A} - \mathbf{A}'\|_F \leq \sqrt{d}\ \min\left\{\|\mathbf{A} - \mathbf{Y}\|_F \mid \mathbf{Y} \in \mathbb{R}^{n_1 \times \ldots \times n_d},\ \text{T-rank}(\mathbf{Y}) \preccurlyeq \mathbf{r}'\right\}$$

The Tucker manifold

Theorem [Tucker manifold]:
The set

$$\mathcal{M}_r^{\mathrm{T}} = \left\{ \mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d} \;\middle|\; \text{T-rank}(\mathbf{X}) = r \right\}$$

admits a manifold structure[2].

[2] Uschmajew & Vandereycken. Linear Algebra and its Applications (2013)

The Tucker manifold

Theorem [Tucker manifold]:
The set

$$\mathcal{M}_r^{\mathrm{T}} = \left\{ \mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d} \mid \text{ T-rank}(\mathbf{X}) = r \right\}$$

admits a manifold structure[2].

Proposition:
The set

$$\mathcal{M}_{\preccurlyeq r}^{\mathrm{T}} = \left\{ \mathbf{X} \in \mathbb{R}^{n_1 \times \ldots \times n_d} \mid \text{ T-rank}(\mathbf{X}) \preccurlyeq r \right\}$$

is closed.

[2] Uschmajew & Vandereycken. Linear Algebra and its Applications (2013)

Storage of Tucker decomposition

We store:

- the core tensor $\mathbf{C} \in \mathbb{R}^{r_1 \times \ldots \times r_d}$
- the basis matrices $\mathbf{U}_k \in \mathbb{R}^{r_k \times n_k}$ for $1 \leq k \leq d$

This scales as

$$\mathcal{O}(r^d + dnr)$$

where $r = \max_i(r_i)$ and $n = \max_i(n_i)$

Note that for $r \ll n$ this is a significant reduction over $\mathcal{O}(n^d)$

Accessing entries

One has to compute

$$\mathbf{A}[j_1, ..., j_d] = (\mathbf{C} *_1 \mathbf{U}_1 *_2 \mathbf{U}_2 ... *_d \mathbf{U}_d)\,[j_1, ..., j_d]$$
$$= \sum_{i_1=1}^{r_1} ... \sum_{i_d=1}^{r_d} \mathbf{C}[i_1, ..., i_d]\mathbf{U}_1[i_1, j_1]\mathbf{U}_2[i_2, j_2]...\mathbf{U}_d[i_d, j_d]$$

set $r = \max_i(r_i)$.

- Per for-loop we have $\mathcal{O}(r)$ operations
- For d nested for-loops this yields: $\mathcal{O}(r^d)$
- Recall: $\mathcal{O}(dr)$ for CP-decomposition

Adding Tucker decompositions

Consider

$$\mathbf{X} = \mathbf{C} *_1 \mathbf{U}_1 *_2 \mathbf{U}_2 ... *_d \mathbf{U}_d$$

and

$$\bar{\mathbf{X}} = \bar{\mathbf{C}} *_1 \bar{\mathbf{U}}_1 *_2 \bar{\mathbf{U}}_2 ... *_d \bar{\mathbf{U}}_d$$

Define

$$\mathbf{V}_k = \begin{bmatrix} \mathbf{U}_k \\ \bar{\mathbf{U}}_k \end{bmatrix} \in \mathbb{R}^{(r_k + \bar{r}_k) \times n_k}$$

and

$$\mathbf{D}[i_1, ..., i_d] = \begin{cases} \mathbf{C}[i_1, ..., i_d] & \text{if } i_\ell \leq r_\ell \ \forall \ell \\ \bar{\mathbf{C}}[i_1 - r_1, ..., i_d - r_d] & \text{if } i_\ell > r_\ell \ \forall \ell \\ 0 & \text{else} \end{cases}$$

Adding Tucker decompositions

Then

$$
\begin{aligned}
(\mathbf{X} + \bar{\mathbf{X}})[j_1, ..., j_d] &= \sum_{i_1=1}^{r_1} ... \sum_{i_d=1}^{r_d} \mathbf{C}[i_1, ..., i_d]\mathbf{U}_1[i_1, j_1] \cdots \mathbf{U}_d[i_d, j_d] \\
&\quad + \sum_{i_1=1}^{\bar{r}_1} ... \sum_{i_d=1}^{\bar{r}_d} \bar{\mathbf{C}}[i_1, ..., i_d]\bar{\mathbf{U}}_1[i_1, j_1] \cdots \bar{\mathbf{U}}_d[i_d, j_d] \\
&= \sum_{i_1=1}^{r_1+\bar{r}_1} ... \sum_{i_d=1}^{r_d+\bar{r}_d} \mathbf{D}[i_1, ..., i_d]\mathbf{V}_1[i_1, j_1] \cdots \mathbf{V}_d[i_d, j_d]
\end{aligned}
$$

Setting $r = \max_i(r_i)$ and $\bar{r} = \max_i(\bar{r}_i)$
The storage scales as

$$
\mathcal{O}\left((r + \bar{r})^d + dn(r + \bar{r})\right)
$$

and evaluating elements scales

$$
\mathcal{O}\left((r + \bar{r})^d\right)
$$

Other operations

Operation	Tucker	CP	Tensor
Had. Prod.	$\mathcal{O}(ndr\bar{r} + r^d\bar{r}^d)$	$\mathcal{O}(ndr\bar{r})$	$\mathcal{O}(n^d)$
Frob. In. Prod.	$\mathcal{O}(ndr\bar{r} + dr\bar{r}^d + r^d)$	$\mathcal{O}(ndr\bar{r})$	$\mathcal{O}(n^d)$
Frob. Norm	$\mathcal{O}(r^d)$	$\mathcal{O}(ndr^2)$	$\mathcal{O}(n^d)$
k-mode Prod.	$\mathcal{O}(mnr + mr^2 + r^{d+1})$	$\mathcal{O}((d+m)nr)$	$\mathcal{O}(n^dm)$