Tensors I: Basic Operations and Representations
Overview

Tensors: Vectors, matrices and so on …

Definition

Operators

Classical decompositions

PARAFAC/Candecomp, polyadic, CP

Tucker, HOSVD

Applications
Different Matrix Products

Kronecker product

\[ A = (a_1 \cdots a_n), \quad B = (b_1 \cdots b_m) \]

Matrix case:

\[ A \otimes B = \begin{pmatrix}
  a_{11}B & a_{12}B & \cdots & a_{1n}B \\
  a_{21}B & a_{22}B & \cdots & a_{2n}B \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1}B & a_{m2}B & \cdots & a_{mn}B
\end{pmatrix}
\]

\[ = (a_1 \otimes b_1 \ a_1 \otimes b_2 \ a_1 \otimes b_3 \ \cdots \ a_n \otimes b_{m-1} \ a_n \otimes b_m) \]

\[ A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix}, \Rightarrow \quad A \otimes B = \begin{pmatrix}
  5 & 7 & 15 & 21 \\
  6 & 8 & 18 & 24 \\
 10 & 14 & 20 & 28 \\
12 & 16 & 24 & 32
\end{pmatrix} \]
Different Matrix Products

Vector case (row or column form):

\[
a \otimes b = (a_1 \ldots a_n) \otimes (b_1 \ldots b_m) = \]
\[
= (a_1 b \ldots a_n b) = (a_1 b_1 \ldots a_1 b_m \ldots a_n b_1 \ldots a_n b_m)
\]

\[
a \otimes b = (a_1 \ldots a_n) \otimes \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = 
\]
\[
= (a_1 b \ldots a_n b) = \begin{pmatrix} a_1 b_1 & \ldots & a_n b_1 \\ \vdots & \ddots & \vdots \\ a_1 b_m & \ldots & a_n b_m \end{pmatrix}
\]
Different Matrix Products

Khatri-Rao product:

\[ A = (a_1 \quad \cdots \quad a_n), \quad B = (b_1 \quad \cdots \quad b_n); \]

\[ A \bullet B = (a_1 \otimes b_1 \quad a_2 \otimes b_2 \quad a_3 \otimes b_3 \quad \cdots \quad a_{n-1} \otimes b_{n-1} \quad a_n \otimes b_n) \]

= matching columnwise Kronecker product
only for matrices with the same number of columns!

Take product of both k-th columns to define new k-th column.

\[ A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix}, \Rightarrow \quad A \bullet B = \begin{pmatrix} 5 & 21 \\ 6 & 24 \\ 10 & 28 \\ 12 & 32 \end{pmatrix} \]
Different Matrix Products

Hadamard product:

\[
A = \begin{pmatrix}
a_{11} & \cdots & a_{1n} \\
\vdots & \ddots & \vdots \\
a_{m1} & \cdots & a_{mn}
\end{pmatrix}, \quad B = \begin{pmatrix}
b_{11} & \cdots & b_{1n} \\
\vdots & \ddots & \vdots \\
b_{m1} & \cdots & b_{mn}
\end{pmatrix},
\]

\[
A \ast B = \begin{pmatrix}
a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\
a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{nm}
\end{pmatrix}
\]

only for matrices of equal size!

\[
\|A\|_F^2 = \sum_{i,j} A_{i,j}^2 = \sum_{i,j} (A \ast A)_{i,j} = trace(A^T A) = \langle A, A \rangle
\]
**Definition**

Tensor as multi-indexed object:

One index: vector: \[ x = (x_i)_{i=1}^n = (x_i)_{i=1}^n = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \]

or \[ x = (x_1 \cdots x_n) \]

Two indices: matrix: \[ A = (A_{i,j})_{i=1,j=1}^{n,m} = (A_{i_1,i_2})_{i_1=1,i_2=1}^{n_1,n_2} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{pmatrix} \]

Three indices: cube: \[ A = (A_{i,j,k})_{i=1,j=1,k=1}^{n,m,l} = (A_{i_1,i_2,i_3})_{i_1=1,i_2=1,i_3=1}^{n_1,n_2,n_3} \]

Multi-index: \[ \chi = \left( x_{i_1i_2\cdots i_N} \right)_{i_1=1,i_2=1,\ldots,i_N=1}^{n_1,n_2,\cdots,n_N} \]
Motivation: Why tensors?

PDE for two-dimensional problems:

\[
\left( au_x \right)_x + \left( bu_y \right)_y = f(x, y)
\]

Discretization in 2D:

\[
\frac{au_{i-1,j} + au_{i+1,j} - (a+b)u_{i,j} + bu_{i,j-1} + bu_{i,j+1}}{h^2} = f_{i,j}
\]

\(u_{i,j}\) can be seen as a vector or as a 2-way tensor=matrix.

Linear system \(Au = f\) with block matrix \(A\):

\[
A_{ij,km}u_{km} = f_{ij}
\]

So matrix \(A_{ij,km}\) can be also seen as a 4-way tensor.
Motivation: Why tensors?

PDE with a number of additional parameters, high-dimensional problems:

\[ au_{xx} + bu_{yy} + cu_{zz} = f \]

for discrete sets of parameters \( a_{ijk} \)

Leads to linear system \( A_{mn} \) for each \( i,j,k \) \( \rightarrow A_{mn,ijk} \)

Classical matrix/vector problems but for huge problems:
Represent vector/matrix by tensor with efficient representation.

\[ x_i = x_{i_1...i_N} \]
Graphical Notation

Vector (1 leg): \( (x_i)_i \leftrightarrow \)

Matrix (2 legs): \( (a_{ij})_{i,j} \leftrightarrow \)

Cube (3 legs): \( (x_{ijk})_{i,j,k} \leftrightarrow \)

General tensor with N legs
\( (x_{i_1...i_N})_{i_1,...,i_N} \leftrightarrow \)
Graphical Notation

Matrix-vector product – contraction over index $j$:

$$\left( a_{ij} \right)_{i,j} \cdot \left( x_j \right)_j = \left( y_i \right)_i$$

$\sum_i a_{ij} x_j = a_{ij} x_j = y_i$

Einstein notation, shared indices are contracted via summation. No distinction between covariant and contravariant!
Basic Operations

Contraction \[ \sum_{i_1} x_{i_1} y_{i_1} \] gives scalar \( z \)

Tensor product \( x_{i_1} y_{i_2} \) gives 2-tensor \( z_{i_1 i_2} \)

More general:
\[
\sum_{i_n} x_{i_1 \cdots i_n \cdots i_N} y_{i_1' \cdots i_n' \cdots i_M} = z_{i_1 \cdots i_n-1 i_{n+1} i_n' \cdots i_M}
\]
Tensor as data hive of different form

\[ \text{kron}(x, y) = x \otimes y = (x_1 y \cdots x_n y)^T = (x_{i_1} y_{i_2})_{i_1, i_2} \]

seen as a column vector

\[ xy^T = \begin{pmatrix} 
  x_1 y_1 & \cdots & x_1 y_m \\
  \vdots & \ddots & \vdots \\
  x_n y_1 & \cdots & x_n y_m 
\end{pmatrix} = (x_{i_1} y_{i_2})_{i_1, i_2} \]

\[ = \text{kron}(y^T, x) = y^T \otimes x \]

seen as a matrix

\[ yx^T = \begin{pmatrix} 
  y_1 x_1 & \cdots & y_1 x_n \\
  \vdots & \ddots & \vdots \\
  y_m x_1 & \cdots & y_m x_n 
\end{pmatrix} = (x_{i_1} y_{i_2})_{i_1, i_2} \]

\[ = \text{kron}(x^T, y) = x^T \otimes y \]

seen as a matrix

\[ x \circ y = (x_{i_1} y_{i_2})_{i_1, i_2} \]

seen as a two-leg tensor
Matrix:

\[ A_{i_1i_2} \]

Matrix:

\[ A_{i_1i_2} \]

Operations: Contractions

\[ \sum_{i_1} x_{i_1} A_{i_1i_2} = z_{i_2} \]

\[ \sum_{i_2} A_{i_1i_2} y_{i_2} = z_{i_1} \]

\[ \sum_{i_1i_2} x_{i_1} A_{i_1i_2} y_{i_2} = z \]

\[ \sum_{i_2} A_{i_1i_2} B_{i_2i_3} = C_{i_1i_3} \]

Tensor product:

\[ A_{i_1i_2} x_{i_3} = C_{i_1i_2i_3} \]
Three Leg as Standardexample

Operations: Contraction in $i_1$, $i_2$, $i_3$ or combinations gives tensor with less legs.

Tensor product gives tensor with more legs.

See tensor as:
- collection of vectors $\rightarrow$ fiber
- collection of matrices $\rightarrow$ slices
- large matrix, unfolding

Operations between tensors are defined by contracting indices.
### Fibers

#### A: 3 x 4 x 2 – tensor

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>7</th>
<th>10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>14</th>
<th>15</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

#### Mode-1 fibers, $X_{:,j,k}$:

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>14</th>
<th>15</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

#### Mode-2 fibers, $X_{j,:,k}$:

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>16</th>
<th>19</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>17</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>18</td>
<td>21</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>14</th>
<th>15</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

#### Mode-3 fibers, $X_{j,k,:}$:

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>
Slices

A: 3 x 4 x 2 – tensor

Frontal slices, 1,2: X_{:,,:,k}

Lateral slices, 1,3: X_{:,k,:}

Horizontal sl. 2,3: X_{k,:,;}

1 4 7 10
2 5 8 11
3 6 9 12

13 16 19 22
14 17 20 23
15 18 21 24

1 4 7 10
2 5 8 11
3 6 9 12

13 16 19 22
14 17 20 23
15 18 21 24

1 4 7 10
2 5 8 11
3 6 9 12

13 16 19 22
14 17 20 23
15 18 21 24
Matricification

A: 3 x 4 x 2 – tensor

Mode-1 unfolding:

\[
A_{(1)} = \begin{bmatrix}
1 & 4 & 7 & 10 \\
2 & 5 & 8 & 11 \\
3 & 6 & 9 & 12 \\
\end{bmatrix}
\]

\[
A_{i_1\{i_2i_3\}} = A_{i_1j_1}
\]

\[
j_1 = i_2 + n_2(i_3 - 1)
\]

Mode-2 unfolding

\[
A_{(2)} = \begin{bmatrix}
1 & 2 & 3 & 13 & 14 & 15 \\
4 & 5 & 6 & 16 & 17 & 18 \\
7 & 8 & 9 & 19 & 20 & 21 \\
10 & 11 & 12 & 22 & 23 & 24 \\
\end{bmatrix}
\]

Mode-3 unfolding

\[
A_{(3)} = \begin{bmatrix}
1 & 2 & 3 & 13 & 14 & 15 \\
14 & 15 & 16 & 17 & 18 & 19 \\
20 & 21 & 22 & 23 & 24 \\
\end{bmatrix}
\]

Vectorization:

\[
vec(A) = \begin{bmatrix}
1 & 2 & \cdots & 23 & 24
\end{bmatrix}^T
\]
General Matricification

Tensor \( A_{i_1 \ldots i_n i_{n+1} \ldots i_N} \) \rightarrow \( A_{\{i_1 \ldots i_n\} \{i_{n+1} \ldots i_N\}} = A_{ij} \) Matrix

\[
i = i_1 + n_2 (i_2 - 1) + n_2 n_3 (i_3 - 1) + \ldots + n_2 \cdots n_n (i_n - 1),
\]
\[
j = i_{n+1} + n_{n+2} (i_{n+2} - 1) + n_{n+2} n_{n+3} (i_{n+3} - 1) + \ldots + n_{n+2} \cdots n_N (i_N - 1).
\]

or with any partitioning of the indices in two groups (rows/columns)

General remark on notation:
many properties/operations with tensors are formulated using totally different notations! ▶, ◁, ⊗, ⊙, ●, ○, ×, ▶◁
Basis Transformation

Tensor

\[ A = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} A_{ijk} \ e_i^{(1)} \otimes e_j^{(2)} \otimes e_k^{(3)} \]

Change of basis

\[ e_i^{(l)} = Q^{(l)} e_i^{(l)}, \ l = 1,2,3 \]

\[ A'_{pqr} = \left( \sum_{i} \sum_{j} \sum_{k} A_{ijk} \left( Q^{(1)} e_i^{(1)} \right) \otimes \left( Q^{(2)} e_j^{(2)} \right) \otimes \left( Q^{(3)} e_k^{(3)} \right) \right)_{pqr} = \]

\[ = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Q_{pi}^{(1)} Q_{qj}^{(2)} Q_{rk}^{(3)} A_{ijk} \]

Notation:

\[ A' = \left( Q^{(1)}, Q^{(2)}, Q^{(3)} \right) \cdot A \]
n-Mode Product of Tensor with Matrix

Tensor \( A_{i_1 \ldots i_n \ldots i_N} \), \( U_{ji_n} : (A \times_n U)_{i_1 \ldots i_{n-1} j_{i_{n+1}} \ldots i_N} = \sum_{i_n=1}^{I_n} a_{i_1 \ldots i_{n-1} i_n} \cdot u_{ji_n} =: B_{i_1 \ldots j \ldots i_N} \)

Contraction over \( i_n \), \( i_n \) replaced by index \( j := i_n' \)

In the n-mode product each mode-n fiber is multiplied by the matrix \( U \):

\[
B_{i_1 \ldots i_{n-1},i_{n+1} \ldots i_N} = U \cdot A_{i_1 \ldots i_{n-1} \ldots i_{n+1} \ldots i_N}
\]

Useful relation between n-mode product and mode-n-unfolding:

\[
B_{(n)} = U \cdot A_{(n)}
\]

Unfold tensor A to matrix, multiply by U, fold back to tensor B.

Unfolding in matrix with blue and red, matrix product, back
n-Mode Products

For multiple n-mode product the order is irrelevant:

\[
\begin{align*}
n \neq m & : A \times_n U \times_m V = A \times_m V \times_n U \\
\sum_{i_m} \left( \sum_{i_n} A_{i_1 \ldots i_n \ldots i_N} U_{ji_n} \right) V_{k_m} &= \\
= \sum_{i_n \neq i_m} A_{i_1 \ldots i_n \ldots i_m} U_{ji_n} V_{k_m} &= \sum_{i_m \neq i_n} A_{i_1 \ldots i_m \ldots i_N} V_{k_m} U_{ji_n} = \\
= \sum_{i_n} \left( \sum_{i_m} A_{i_1 \ldots i_n \ldots i_m \ldots i_N} V_{k_m} \right) U_{ji_n} \\

\text{A matrix: } B &= A \times_1 U \times_2 V \iff \\
\left( B_{jk} \right) &= \left( A_{i_1i_2} \right) \times_1 \left( U_{ji_1} \right) \times_2 \left( V_{ki_2} \right) = U \cdot A \cdot V^T = U_{j,} \cdot A \cdot \left( V_{k,} \right)^T \\
\text{especially } A \times_1 U &= U \cdot A, \quad A \times_2 V = A \cdot V^T
\end{align*}
\]
Proofs:  
\[ A \times_1 U = U \cdot A, \quad A \times_2 V = A \cdot V^T \]

\[
(A \times_1 U)_{ji_2} = \sum_{i_1} A_{i_1i_2} U_{ji_1} = \sum_{i_1} U_{ji_1} A_{i_1i_2} = (U \cdot A)_{ji_2}
\]

\[
(A \times_2 V)_{i_1j} = \sum_{i_2} A_{i_1i_2} V_{ji_2} = \sum_{i_2} A_{i_1i_2} (V^T)_{i_2j} = (A \cdot V^T)_{i_1j}
\]

\[
(A \times_1 U \times_2 V)_{jk} = \sum_{i_1,i_2} A_{i_1i_2} U_{ji_1} V_{ki_2} = \sum_{i_1,i_2} U_{ji_1} A_{i_1i_2} (V^T)_{ki_2} = (U \cdot A \cdot V^T)_{jk}
\]
n-Mode Products

For multiple n-mode product with the same n the order is relevant:

\[ A \times_n U \times_n V = A \times_n (VU) \]

\[
\sum_{i'_n} \left( \sum_{i_n} A_{i_1...i_n} U_{i'_n i_n} \right) V_{k'_n i_n} = \\
= \sum_{i_n} A_{i_1...i_n} \sum_{i'_n} U_{i'_n i_n} V_{k'_n i_n} = \sum_{i_n} A_{i_1...i_n} \left( \sum_{i'_n} V_{k'_n i_n} U_{i'_n i_n} \right) = \\
= \sum_{i_n} A_{i_1...i_n} W_{k_i n} = B_{i_1...k...i_N}
\]

Matrix case:  
\[ A \times_1 U \times_1 V = V \cdot U \cdot A = (VU) \cdot A, \]
\[ A \times_2 U \times_2 V = A \cdot U^T \cdot V^T = A \cdot (VU)^T \]
n-Mode Product with vector

n-mode vector product of tensor A with vector v: Compute all inner products of mode-n fibers with v.

\[ A \overline{\times}_n v = \left( \sum_{i_n=1}^{n_n} A_{i_1...i_n...i_N} v_{i_n} \right)_{i_1...i_{n-1}i_{n+1}...i_N} \]

\[ A \overline{\times}_n v \overline{\times}_m u = (A \overline{\times}_n v) \overline{\times}_{m-1} u = (A \overline{\times}_m u) \overline{\times}_n v = \]

\[ = \left( \sum_{i_n=1}^{n_n} \sum_{i_m=1}^{n_m} A_{i_1...i_n...i_m...i_N} v_{i_n} u_{i_m} \right)_{i_1...i_{n-1}i_{n+1}...i_{m-1}i_{m+1}...i_N} \]

for n<m because the order of the tensor is changed: After contracting i_n: m \rightarrow m-1

Matrix case: \[ A \overline{\times}_1 v = v^T \cdot A, \quad A \overline{\times}_2 v = A \cdot v \]
Properties

(1) \[(A \otimes B)(C \otimes D) = (AC) \otimes (BD),\]

(2) \[(A \otimes B)^{-T} = A^{-T} \otimes B^{-T}\]

(3) \[A \bullet B \bullet C = (A \bullet B) \bullet C = A \bullet (B \bullet C),\]

(4) \[(A \bullet B)^T (A \bullet B) = (A^T A) \ast (B^T B),\]
\[(A \bullet B)^{-1} = ((A^T A) \ast (B^T B))^{-1} (A \bullet B)^T\]
Proofs (1):

\[(A \otimes B)(C \otimes D) =\]

\[
\begin{pmatrix}
  a_{11}B & \cdots & a_{1n}B \\
  \vdots & \ddots & \vdots \\
  a_{m1}B & \cdots & a_{mn}B
\end{pmatrix}
\begin{pmatrix}
  c_{11}D & \cdots & c_{1k}D \\
  \vdots & \ddots & \vdots \\
  c_{n1}D & \cdots & c_{nk}D
\end{pmatrix}
=\]

\[
\begin{pmatrix}
  a_{11}c_{11}BD + \cdots + a_{1n}c_{n1}BD & \cdots \\
  \vdots & \ddots & \vdots \\
  \vdots & \ddots & \vdots
\end{pmatrix}
=\]

\[
\begin{pmatrix}
  (AC)_{11}BD & \cdots \\
  \vdots & \ddots \\
  \vdots & \ddots & \vdots
\end{pmatrix} = (AC) \otimes (BD),
\]
Proofs (2):

\[(A \otimes B)^{-T} = A^{-T} \otimes B^{-T}\]

\[
(A \otimes B)(A^{-1} \otimes B^{-1}) = \left((AA^{-1}) \otimes (BB^{-1})\right) = I \otimes I = I
\]

\[
(A \otimes B)^T = \begin{pmatrix}
  a_{11}B & \cdots & a_{1n}B \\
  \vdots & \ddots & \vdots \\
  a_{m1}B & \cdots & a_{mn}B
\end{pmatrix}^T
\]

\[
= \begin{pmatrix}
  a_{11}B^T & \cdots & a_{m1}B^T \\
  \vdots & \ddots & \vdots \\
  a_{1n}B^T & \cdots & a_{nm}B^T
\end{pmatrix} = A^T \otimes B^T
\]
Proofs (3):

\[ A \bullet B \bullet C = (A \bullet B) \bullet C = A \bullet (B \bullet C), \]

\[
(A \bullet B) \bullet C = \left( a_1 \otimes b_1 \cdots a_n \otimes b_n \right) \bullet C = \\
\left( (a_1 \otimes b_1) \otimes c_1 \cdots (a_n \otimes b_n) \otimes c_n \right) = \\
= \left( a_1 \otimes b_1 \otimes c_1 \cdots a_n \otimes b_n \otimes c_n \right) = \\
A \bullet (B \bullet C),
\]

because \( (A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C \)
Proofs (4):

\[(A \bullet B)^{-1} = ((A^T A) \ast (B^T B))^{-1} (A \bullet B)^T \]

\[((A^T A) \ast (B^T B)) = (A \bullet B)^T (A \bullet B)\]

\[
((A^T A) \ast (B^T B)) = \left(\begin{array}{c c c}
(a_i^T a_j)(b_i^T b_j)
\end{array}\right)
\]

\[
\left(\begin{array}{c c c}
(a_1^T a_1)(b_1^T b_1) & (a_1^T a_2)(b_1^T b_2) & \cdots \\
(a_2^T a_1)(b_2^T b_1) & \ddots & \\
\vdots & \ddots & \\
\end{array}\right)
\]

\[
\left(\begin{array}{c c c}
(a_1^T \otimes b_1^T)(a_1 \otimes b_1) & (a_1^T \otimes b_1^T)(a_2 \otimes b_2) & \cdots \\
(a_2^T \otimes b_2^T)(a_1 \otimes b_1) & \ddots & \\
\vdots & \ddots & \\
\end{array}\right)
\]

\[
= \left(\begin{array}{c c c}
(a_1^T \otimes b_1^T) & \cdots & (a_n^T \otimes b_n^T)
\end{array}\right)
\]

\[
= (a_1 \otimes b_1 \cdots a_n \otimes b_n)^T (a_1 \otimes b_1 \cdots a_n \otimes b_n) = (A \bullet B)^T (A \bullet B)
\]
n-Mode Products Tensor with Matrices

General relation between n-mode product, mode-n unfolding and Kronecker (tensor) product:

\[ Y = A \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)} \iff \]

\[ Y_{(n)} = U^{(n)} \cdot A_{(n)} \cdot \left( U^{(N)} \otimes \cdots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \cdots \otimes U^{(1)} \right)^T \]

\[ Y = A \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)} = \sum_{i_1, \ldots, i_N} A_{i_1 \ldots i_N} U^{(1)}_{j_1 i_1} \cdots U^{(N)}_{j_N i_N} = B_{j_1 \ldots j_N} \]

N=2:

\[ Y = A \times_1 U^{(1)} \times_2 U^{(2)} = U^{(1)} A \left( U^{(2)} \right)^T \]

\[ Y_{(1)} = \left( U^{(1)} A \left( U^{(2)} \right)^T \right)_{(1)} = U^{(1)} A_{(1)} \left( U^{(2)} \right)^T \]

\[ Y_{(2)} = \left( U^{(1)} A \left( U^{(2)} \right)^T \right)_{(2)} = \left( U^{(1)} A \left( U^{(2)} \right)^T \right)^T = \]

\[ = U^{(2)} A^T \left( U^{(1)} \right)^T = U^{(2)} A_{(2)} \left( U^{(1)} \right)^T \]
n-Mode Products Tensor with Matrices

\[
Y_{(1)} = \left( A \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \right)_{(1)} =
\]

\[
= \left( \sum_{i_1,i_2,i_3} A_{i_1i_2i_3} U^{(1)}_{j_1i_1} U^{(2)}_{j_2i_2} U^{(N)}_{j_Ni_N} \right)_{(1)} =
\]

\[
= \left( \sum_{i_1} U^{(1)}_{j_1i_1} \left( \sum_{i_2i_3} A_{i_1i_2i_3} U^{(2)}_{j_2i_2} U^{(3)}_{j_3i_3} \right) \right)_{(1)} =
\]

\[
= \left( \sum_{i_1} B_{i_1j_2j_3} U^{(1)}_{j_1i_1} \right)_{(1)} = \left( B_{i_1j_2j_3} \times_1 U^{(1)}_{j_1i_1} \right)_{(1)} =
\]

\[
= U^{(1)} \left( B_{i_1j_2j_3} \right)_{(1)} = U^{(1)} \left( \sum_{i_2i_3} A_{i_1i_2i_3} U^{(3)}_{j_3i_3} U^{(2)}_{j_2i_2} \right)_{(1)} =
\]

\[
= U^{(1)} \sum_k A_{i_1k} \left( U^{(3)} \otimes U^{(2)} \right)_{r,k} = U^{(1)} A_{(1)} \left( U^{(3)} \otimes U^{(2)} \right)^T
\]

\[k = \{i_2i_3\}, r = \{j_2j_3\}\]
Rank of a tensor (3 leg case)

Rank-1 tensor:

\[
\begin{align*}
(A_{ijk}) &= (a \circ b \circ c) \\
(A_{ijk}) &= (a \otimes b \otimes c)
\end{align*}
\]

3 dimensional

as vector

with vectors a, b, and c

\[A_{ijk} = a_i b_j c_k\]
Rank-R tensor for 3-leg case:

PARAFAC (parallel factors)
Candecomp (canonical decomposition)
Polyadic form
→ CP (CANDECOMP/PARAFAC)

\[
\begin{align*}
A_{ijk} &= (u_1 \circ v_1 \circ w_1) + (u_2 \circ v_2 \circ w_2) + (u_3 \circ v_3 \circ w_3) + \cdots \\
A_{ijk} &= \sum_{r=1}^{R} (u_{ri} v_{rj} w_{rk})
\end{align*}
\]

Tensor rank R of tensor \((A_{ijk})\) is the number of rank-1 terms that are necessary for representing A.
Rank representation

\[ A = \sum_{r=1}^{R} u_r \circ v_r \circ w_r = \]

\[ = \sum_{r=1}^{R} \left( \sum_{i=1}^{I} u_{ir} e_i^{(1)} \circ \sum_{j=1}^{J} v_{jr} e_j^{(2)} \circ \sum_{k=1}^{K} w_{kr} e_k^{(3)} \right) = \]

\[ = \sum_{i,j,k}^{R} \left( \sum_{r=1}^{R} u_{ir} v_{jr} w_{kr} \right) e_i^{(1)} \circ e_j^{(2)} \circ e_k^{(3)} \]

With matrices U, V, and W we can write

\[ A_{ijk} = \sum_{r=1}^{R} \left( u_{ir} v_{jr} w_{kr} \right) = \sum_{p,q,t} \left( u_{ip} v_{jq} w_{kt} \right) \delta_{p,q,t} \]

\[ A = \left( U, V, W \right) \cdot I \]

with I the 3-way tensor with 1 on the main diagonal

U, V, W describe basis transformation with A \( \rightarrow \) I
Notation

Let $U$, $V$, and $W$ be the matrices built by the vectors $u_r$, $v_r$, and $w_r$. Then we can write

$$A_{(1)} = U (W \bullet V)^T,$$

$$A_{(2)} = V (W \bullet U)^T,$$

$$A_{(3)} = W (V \bullet U)^T.$$

Short notation: $A = [[U, V, W]] = \sum_{k=1}^{R} u_k \circ v_k \circ w_k$

Or more general with factor $\lambda$: $A = [[\lambda; U, V, W]] = \sum_{k=1}^{R} \lambda_k u_k \circ v_k \circ w_k$
Proof:

Two-leg tensor

\[(u \circ v)_{(1)} = u \cdot v^T = u_1 \cdots v_1 \cdots v_m\]

One 3-leg tensor:

\[(u \circ v \circ w)_{(1)} = \left(\left(u_i v_j w_k \right)_{ijk}\right)_{(1)} = \right\]

\[= \left(u_i (v_j w_k)\right)_{i\{jk\}} = \left(u_i (w \otimes v)_{\{jk\}}\right)_{i\{jk\}} \rightarrow u(w \otimes v)^T = u(w \cdot v)^T\]

General 3-leg case:

\[\left(\sum_{r=1}^{R} u_r \circ v_r \circ w_r\right)_{(1)} = \sum_{r=1}^{R} \left(u_r \circ (v_r \circ w_r)\right)_{(1)} = \]

\[= \sum_{r=1}^{R} u_r \cdot (w_r \cdot v_r)^T = \]

\[= (u_1 \cdots u_R) \cdot (w_1 \cdot v_1 \cdots w_R \cdot v_R)^T = \]

\[= U(W \cdot V)^T\]
General N-way tensor

\[ A = [[[U^{(1)}, U^{(2)}, \ldots, U^{(N)}]]] = \sum_{k=1}^{R} u_{1,k} \circ u_{2,k} \circ \ldots \circ u_{N,k} \]

\[ A = [[[\lambda; U^{(1)}, U^{(2)}, \ldots, U^{(N)}]]] = \sum_{k=1}^{R} \lambda_k u_{1,k} \circ u_{2,k} \circ \ldots \circ u_{N,k} \]

Mode-n matrix formula:

\[ A_{(n)} = U^{(n)} \Lambda \left( U^{(N)} \bullet \ldots \bullet U^{(n+1)} \bullet U^{(n-1)} \bullet \ldots \bullet U^{(1)} \right)^T \]

with \( \Lambda = \text{diag}(\lambda) \)
Proof:

3-leg tensor, proof like before:

\[
\sum_r \lambda_r u_r \circ v_r \circ w_r = (U\Lambda)(W \bullet V)^T
\]

In general:

\[
U^{(1)} \Lambda \left( U^{(N)} \bullet \cdots \bullet U^{(2)} \right)^T = \\
= U^{(1)} \left( \lambda_1 U^{(N)}_1 \otimes \cdots \otimes U^{(2)}_1 \right) \cdots \lambda_R U^{(N)}_R \otimes \cdots \otimes U^{(2)}_R)^T = \\
= \left( \sum_{r=1}^R \lambda_r U^{(1)}_r \otimes U^{(2)}_r \otimes \cdots \otimes U^{(N)}_r \right)_{(1)}
\]
Low rank approximation

\[ A_{i_1...i_N} = \sum_{k=1}^{R} a_{ki_1}...a_{ki_N} \approx \sum_{k=1}^{r} b_{ki_1}...b_{ki_N} \]

(1) For R large enough every A can be represented by CP

(2) For given A there is a minimum R with this property

(3) Approximate A as good as possible by r<R
PARAFAC, CP, CANDECOMP

Graphical

\[ A_{i_1,i_2,i_3,\ldots,i_N} \]

\[ \approx \]

\[ \sum_{r=1}^{R} U_{1,i_1,r} U_{2,i_2,r} \cdots U_{N,i_N,r} \]
Norm etc.

Inner product: $\langle A_{i_1...i_N}, B_{i_1...i_N} \rangle = \sum_{i_1...i_N=1}^{n_1...n_N} A_{i_1...i_N} B_{i_1...i_N}$

Norm: $\| A_{i_1...i_N} \| = \sqrt{\sum_{i_1...i_N=1}^{n_1...n_N} A_{i_1...i_N}^2}$

Rank-One tensor: $A = a^{(1)} \circ a^{(2)} \circ \ldots \circ a^{(N)}$ with vectors $a^{(j)}$

Diagonal tensor: $A_{i_1...i_N} \neq 0 \iff i_1 = i_2 = \ldots = i_N$
Symmetry

A tensor is called cubical, if every mode is of the same size,
\( n_1=n_2=\ldots=n_N \)

A cubical tensor is called supersymmetric, if its elements
Remain constant under any permutation of the indices:

\[
A_{i_1\ldots i_N} = A_{i_{\pi(1)}\ldots i_{\pi(N)}}
\]

A tensor is partial symmetric, if it is symmetric in some modes,
e.g. three-way tensor, where all frontal slices are symmetric
matrices.
Example

$A_{111} = 1, A_{112} = A_{121} = A_{211} = 2, A_{122} = A_{212} = A_{221} = 3, A_{222} = 4.$
Results on tensor rank

\[ A_{i_1...i_N} = \sum_{k=1}^{R} a_{k_i_1}...a_{k_i_N} \] with minimum R, dimension \( n_1,...,n_N, \) \( n_j \leq n \)

For general N-way tensor it holds: \( R=\text{rank} \leq n^{N-1} \)

Proof: Assume \( n_N=n=\max n_j. \)

\[ A = \sum_{i_1,...,i_N} A_{i_1...i_N} e^{(1)}_{i_1} \otimes ... \otimes e^{(N)}_{i_N} = \]

\[ = \sum_{i_1,...,i_{N-1}} e^{(1)}_{i_1} \otimes ... \otimes e^{(N-1)}_{i_{N-1}} \otimes \left( \sum_{i_N} A_{i_1...i_N} e^{(N)}_{i_N} \right) \]

where the summation runs over maximum rank 1 terms.
Results on tensor rank

The true rank might be much smaller:

The maximum rank of a 3 leg tensor $3 \times 3 \times 3$ over IR is bounded by 5.

For general 3 leg $I \times J \times K$ tensor $A$ the maximum rank is bounded by $\text{rank}(A) \leq \min\{IJ, IK, JK\}$

For general 3 leg $I \times J \times 2$ tensor $A$ the maximum rank is bounded by $\text{rank}(A) \leq \min\{I, J\} + \min\{I, J, \frac{\max\{I, J\}}{2}\}$

The typical rank of a 3 leg tensor $5 \times 3 \times 3$ over IR is 5 or 6.
Results on tensor rank

Example: \[ A = a \otimes a + a \otimes b + b \otimes a + b \otimes b \]
with linearly independent \(a\) and \(b\), \(\text{rank} \leq 4\), with 4 linearly independent terms, but
\[ A = (a + b) \otimes (a + b) \quad \text{with rank 1.} \]

Theorem: \(\text{rank}(A)=3\) for
\[ A = v_1 \otimes v_2 \otimes w_3 + v_1 \otimes w_2 \otimes v_3 + w_1 \otimes v_2 \otimes v_3 \]
with linearly independent \(v_j, w_j\).

Proof: (1) \(\text{rank}(A)=0 \rightarrow A=0 \rightarrow v_1 \otimes a = w_1 \otimes b \quad !!!\)

(2) \(\text{rank}(A)=1 \rightarrow \)
\[ u \otimes v \otimes w = v_1 \otimes v_2 \otimes w_3 + v_1 \otimes w_2 \otimes v_3 + w_1 \otimes v_2 \otimes v_3 \]
Assume a linear functional with
\[ \varphi_1(v_1) = 1, \quad \varphi := \varphi_1 \otimes id \otimes id \]
and apply it on above equation:
\[ \varphi_1(u)v \otimes w = v_2 \otimes w_3 + w_2 \otimes v_3 + \varphi_1(w_1)v_2 \otimes v_3 = \]
\[ = v_2 \otimes w_3 + (w_2 + \varphi_1(w_1)v_2) \otimes v_3 \]
Left side rank 1 matrix, right side rank 2 matrix !!!

(3) Rank(A)=2:
\[ u \otimes v \otimes w + u' \otimes v' \otimes w' = v_1 \otimes v_2 \otimes w_3 + v_1 \otimes w_2 \otimes v_3 + w_1 \otimes v_2 \otimes v_3 \]
If \( u \) and \( u' \) are linearly dependent there is a functional
\[ \varphi_1(u) = \varphi_1(u') = 0, \quad \varphi_1(v_1) \neq 0 \quad \text{or} \quad \varphi_1(w_1) \neq 0 \]
\[ 0 = (\varphi_1 \otimes \text{id} \otimes \text{id})(A) = \varphi_1(v_1)(v_2 \otimes w_3 + w_2 \otimes v_3) + \varphi_1(w_1)v_2 \otimes v_3 \]

Linearly independent !!

Hence, \( u \) and \( u' \) have to be linearly independent, and one of the vectors \( u \) or \( u' \) must be linearly independent of \( v_1 \), say \( u' \) is l.i. of \( v_1 \).

Choose functional with \( \varphi_1(v_1) = 1, \quad \varphi_1(u') = 0 \).

\[ \varphi_1(u)v \otimes w = (v_2 \otimes w_3 + w_2 \otimes v_3) + \varphi_1(w_1)v_2 \otimes v_3 \]

Again, the left-hand-side matrix is rank \( \leq 1 \), the right-hand-side matrix has rank 2 !!!
For a supersymmetric tensor we can define the symmetric rank:

$$\text{rank}_S(A) = \min \left\{ r : \ A = \sum_{k=1}^{r} a_r \circ a_r \circ \ldots \circ a_r \right\}$$

Example: $$A = (1,2)^{\otimes 3} = (1,2) \otimes \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad B = (1,1)^{\otimes 3} = (1,1) \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Supersymmetric of symmetric rank 2.

Rank 1: $$ (a,b)^{\otimes 3} = A + B, \quad 4 \text{ equations for 2 unknowns } a,b.$$ 
$$a^3 = 2, b^3 = 9, a^2 b = 3, ab^2 = 5;$$
### Smallest Typical Rank 3-way T

<table>
<thead>
<tr>
<th>K</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>I</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

**DOF:** $R(I+J+K-2) \rightarrow$ Expected Rank: \[
\left[ \frac{IJK}{I + J + K - 2} \right]
\]
Examples

Strassen by considering a 3-leg tensor with rank 7
Hackbusch page 69

\[
\begin{pmatrix}
a_1 & a_2 \\
a_3 & a_4 \\
\end{pmatrix}
\begin{pmatrix}
b_1 & b_2 \\
b_3 & b_4 \\
\end{pmatrix}
= \begin{pmatrix}
c_1 & c_2 \\
c_3 & c_4 \\
\end{pmatrix}
\]

with submatrices \( a_j, b_j, c_j \)

\[c_v = \sum_{\mu,\lambda=1}^{4} t_{v,\mu,\lambda} a_\mu b_\lambda\]

\( t \) is of rank 7.
Matrix case: SVD

For a tensor that is a vector, the rank is 1.

For a tensor that is an nxm matrix, the rank is given by the singular value decomposition

\[ A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i (u_i v_i^T) = \sum_{i=1}^{r} \sigma_i (u_i \otimes v_i) \]

\[ r = \text{the number of nonzero singular values.} \]

For low rank approximation we can delete the small singular values.
Uniqueness of CP

Matrix case: A nxm matrix of rank r:

\[ A = U_{n,r} V_{r,m}^T = \sum_{k=1}^{r} u_r \circ v_r \]

Every matrix factorization of this form gives a CP representation.

QR-factorizations, SVD.

In the matrix case (2-leg-case) the rank representations are not unique!
Uniqueness 3 leg case

Let $A$ be a three-way tensor of rank $R$:

$$A = [[U, V, W]] = \sum_{k=1}^{R} u_k \circ v_k \circ w_k$$

In the following uniqueness is considered with respect to other rank $R$ representations up to scaling and permutations:

$$A = [[U, V, W]] = [[U\Pi, V\Pi, W\Pi]] \quad \text{for any } R \times R \text{ permutation } \Pi$$

$$A = \sum_{k=1}^{R} (\alpha_k u_k) \circ (\beta_k v_k) \circ (\gamma_k w_k) \quad \text{with } \alpha_k \beta_k \gamma_k = 1, \text{ for } k=1, \ldots, R$$
k-rank of a matrix

The k-rank of a matrix $B$ - denoted by $k_B$ – is the maximum number $k$ such that any $k$ columns of $B$ are linearly independent.

Example: Consider, matrix built by vectors $u$, $v$, and $\alpha u + \beta v$. What is the rank? What is the k-rank?

$$A = [[U, V, W]]$$

Then the CP representation of $A$ is unique if

$$k_U + k_V + k_W \geq 2R + 2$$
Let $A$ an $I \times J \times K$-Tensor:

Then the CP representation of $A$ is unique if

$$\min\{I, R\} + \min\{J, R\} + \min\{K, R\} \geq 2R + 2$$

For $R \leq K$ the CP representation of $A$ is unique if

$$2R(R - 1) \leq I(I - 1)J(J - 1)$$

The CP representation is unique for an $N$-way rank $R$ tensor

$$A = \left[[U^{(1)}, U^{(2)}, \ldots, U^{(N)}]\right] = \sum_{k=1}^{R} u_{k}^{(1)} \circ u_{k}^{(2)} \circ \cdots \circ u_{k}^{(N)}$$

if

$$\sum_{n=1}^{N} k_{U^{(n)}} \geq 2R + (N - 1)$$
Approximation of tensor by CP

Matrix case trivial via SVD: keep larger singular values and replace smaller ones by 0.

For 3-way tensors this is not so easy. Especially for

\[ A = \sum_{k=1}^{R} \lambda_k u_k \circ v_k \circ w_k \]

summing up \( r \) of these terms will not give a good rank-\( r \) approximation.

For finding the best rank-\( r \) approximation we have to determine all factors simultaneously!
Rank-r approximation

The situation is even worse: the best rank-r approximation might even not exist!

Consider

$$A = u_1 \circ v_1 \circ w_2 + u_1 \circ v_2 \circ w_1 + u_2 \circ v_1 \circ w_1$$

where the matrices U, V, and W have linearly independent Columns.

Approximation by rank-2 tensors:

$$B_\alpha = \alpha \left( u_1 + \frac{1}{\alpha} u_2 \right) \circ \left( v_1 + \frac{1}{\alpha} v_2 \right) \circ \left( w_1 + \frac{1}{\alpha} w_2 \right) - \alpha (u_1 \circ v_1 \circ w_1)$$

$$\|A - B_\alpha\| = \frac{1}{\alpha} \left\| u_2 \circ v_2 \circ w_1 + u_2 \circ v_1 \circ w_2 + u_1 \circ v_2 \circ w_2 + \frac{1}{\alpha} u_2 \circ v_2 \circ w_2 \right\| \to 0$$

Example for degeneracy!
Another example:

\[ A(n) = n^2 \left( x + \frac{1}{n^2} y + \frac{1}{n} z \right) \otimes^3 + n^2 \left( x + \frac{1}{n^2} y - \frac{1}{n} z \right) \otimes^3 - 2n^2 x \otimes^3 \]

with linearly independent \( x, y, z \).

The sequence of rank 3 tensors converges for \( n \to \infty \) to the rank 5 tensor:

\[ A(\infty) = x \otimes x \otimes y + x \otimes y \otimes x + y \otimes x \otimes x + x \otimes z \otimes z + z \otimes x \otimes z + z \otimes z \otimes x \]
Rank spaces

Hence a sequence of rank-2 tensors converges against a rank-3 tensor:
The space of rank-2 tensors is not closed!

We can approximate the 3-way tensor as good as we want by rank-2 tensors, but the sequence of approximations does not converge in the rank-2 space.
Computing the CP

Standard method: **Alternating Least Squares method (ALS)**  
**Alternating Linear Scheme method (ALS)**

Given any (high-rank) tensor \( A \)  
Compute \( r \)-rank approximation in tensor \( B \)

\[
\min_B \| A - B \| 
\text{ with } B = \sum_{k=1}^{r} \lambda_k u_k \circ v_k \circ w_k = [[\lambda; U, V, W]]
\]

**ALS approach**: fix two matrices, e.g. \( V \) and \( W \), and solve for \( U \).  
This leads to the matrix minimization

\[
\min_{\hat{U}} \| A^{(1)} - \hat{U}(W \cdot V)^T \|_F
\]

with solution

\[
\hat{U} = A^{(1)}((W \cdot V)^T)^{-1} = A^{(1)}(W \cdot V)(W^T W \ast V^T V)^{-1}
\]
Computational Aspects:

Advantage: Compute pseudoinverse of small $r \times r$-matrix

Afterwards, $\lambda$ is defined by normalization

$$\lambda_k = \|\hat{u}_k\|, \quad u_k = \hat{u}_k / \lambda_k, \quad k = 1, \ldots, r$$

In this way we update $U$, then $V$, then $W$, then again $U$ and so on until convergence.

Costs per step: $IJKr + JKr^2 + (J+K)r^2 + r^3$

$$1 \left\{ A_{(1)} \left( W \bullet V \right) \left( W^T W \ast V^T V \right)^{-1} \right\}^r - r$$
ELS
ALS with enhanced line search

Assume, ALS has computed new $U_{new}$ replacing $U_{old}$. Hence, we have a change in the direction $\Delta = U_{new} - U_{old}$ in the form $U_{new} = U_{old} + \Delta$.

We generalize this by introducing line search and step size $\mu$ in the form $U_{new} = U_{old} + \mu \Delta$ looking for an optimal value of $\mu$.

$$\min_{\mu} \left\| A - \sum_{k=1}^{R} (u_k + \mu \delta_k) \circ v_k \circ w_k \right\|^2$$

$$= \min_{\mu} \left\| (A - \sum_{k=1}^{R} u_k \circ v_k \circ w_k) - \mu \sum_{k=1}^{R} \delta_k \circ v_k \circ w_k \right\|^2$$

$$= \min_{\mu} \left\| B - \mu C \right\|^2 \rightarrow \mu \rightarrow U_{new} = U_{old} + \mu \Delta$$
ELS general

\[ U_{\text{new}} = U_{\text{old}} + \mu \Delta_U, \quad V_{\text{new}} = V_{\text{old}} + \mu \Delta_V, \quad W_{\text{new}} = W_{\text{old}} + \mu \Delta_W, \]

\[
\min_{\mu} \left\| A - \sum_{k=1}^{R} \left( u_k + \mu \delta_{u,k} \right) \circ \left( v_k + \mu \delta_{v,k} \right) \circ \left( w_k + \mu \delta_{w,k} \right) \right\|^2
\]

\[
= \min_{\mu} \left\| B - \mu^3 C - \mu^2 D - \mu E \right\|^2
\]

\[
= \min_{\mu} a_0 + a_1 \mu + a_2 \mu^2 + a_3 \mu^3 + a_4 \mu^4 + a_5 \mu^5 + a_6 \mu^6
\]

Find the 5 roots of the derivative and choose the root with minimum value of the objective function. Gives new U, V, and W. Use ALS for new search directions and repeat.
Application of the CP

Starting point: 3-leg tensors often have small rank and the low-rank approximation is unique.

Therefore, the best approximating rank-1 term can give useful information on the data:
- Mixtures of analytes can be separated
- Concentrations can be measured
- Pure spectra and profiles can be estimated

Typical example: 3-way data in time, space, frequency

Translate matrix case by additional index in 3-leg tensor to achieve uniqueness!
Application of the CP

Van Huffel: PARAFAC in EEG monitoring

EEG data as 3-way tensor

\[
X = A_1 \circ B_1 + \ldots + A_R \circ B_R + \varepsilon
\]

Split EEG in different frequencies using wavelets.

\[\rightarrow \text{Analysis in 3 dimensions instead of just 2}\]
EEG Monitoring

PARAFAC: Example extracting 1 component

\[ X = A_1 C_1 B_1 \]

- \( B_1 \): time course
- \( A_1 \): distribution over channels
- \( C_1 \): frequency content
  (distribution across scales).

\[ \begin{align*}
\text{frequency} & \quad \text{time} \\
\text{place} & \quad \text{place}
\end{align*} \]
EEG rank terms

\[ \chi = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{A}_1 \end{bmatrix} + \ldots + \begin{bmatrix} \mathbf{B}_R \\ \mathbf{A}_R \end{bmatrix} + \mathbf{E} \]

\[ \mathbf{A}_1 \]

\[ \mathbf{A}_2 \]
EEG: epileptic seizure onset localization
Better localization by CP than visually or by other matrix techniques.
Block PARAFAC (L,L,1)

Consider more general higher rank terms (L,L,1) Because larger blocks might be necessary for accurate representation of the data.

\[ A = \sum_{r=1}^{R} E_r \otimes w_r, \quad E_r: I \times J - matrix, \quad \text{rank}(E_r) = L \]

\[ A = \sum_{r=1}^{R} \left( U_r \cdot V_r^T \right) \otimes w_r \]

Also block representations are often unique, e.g. for \( RL \leq \min(I,J) \) and \( W \) without proportional columns.

„Essentially unique“, upto - permutations, - factor between \( U \) and \( V \) - scaling
Visualization

\[ A = (U, V, w) \cdot D = [[D; U, V, w]] = \sum_{r=1}^{R} D_r \times_1 U_r \times_2 V_r \times_3 w_r \]
Mode n-Rank of a Tensor

View the tensor as collection of vectors in the n-th index (fibers)
The rank of these collection of vectors is the mode n-rank.

Example with $R_1=R_2=2$, $R_3=1$

Mode n=3:
Vectors (0,0), (1,1), (1,1), (0,0)

$$R_3 = \text{rank}\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} = 1$$

Mode n-rank is the rank of the mode-n unfolding matrix $A_{(n)}$
Tucker Decomposition

(three-mode) factor analysis (Tucker, 1966)
N-mode PCA (principal component analysis)
Higher-order SVD (HOSVD) (De Lathauwer, 2000)
N-mode SVD

Idea: decompose given N-way tensor into a core
N-way tensor with less entries in each dimension
multiplied by a matrix along each mode.

\[ A = G \times_1 U \times_2 V \times_3 W = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{k=1}^{K} g_{pqk} u_p \circ v_q \circ w_k = \]

\[ = [[G; U, V, W]] = (U, V, W) \cdot G \]

with core tensor G and U,V,W matrices relative to each mode,
Basis transformation in each direction.
Tucker Decomposition

\[
A_{ijk} = (G_{ijk}) \times_1 U \times_2 V \times_3 W
\]

Unfolding in \(i\) and \(\{jk\}\),

\[
\text{SVD} \rightarrow U
\]

Backfolding \(\Lambda V\)

Repeat for other unfoldings.

\[
(A_{ijk}) = \sum_{\mu_1=1}^{k_1} \sum_{\mu_2=1}^{k_2} \sum_{\mu_3=1}^{k_3} G_{\mu_1\mu_2\mu_3} \cdot u_{\mu_1} \circ v_{\mu_2} \circ w_{\mu_3}
\]

\[
A_{ijk} = \sum_{\mu_1}^{k_1} \sum_{\mu_2}^{k_2} \sum_{\mu_3}^{k_3} G_{\mu_1\mu_2\mu_3} u_{\mu_1} v_{\mu_2} w_{\mu_3}, i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K
\]

**Multilinear rank** \((k_1, k_2, k_3)\)
Computation

\[ A(1) : I_1 \times I_2 I_3; \quad SVD : \quad A(1) = U^{(1)} \Sigma^{(1)} V^{(1)^T} \]
\[ A(2) : I_2 \times I_1 I_3; \quad SVD : \quad A(2) = U^{(2)} \Sigma^{(2)} V^{(2)^T} \]
\[ A(3) : I_3 \times I_1 I_2; \quad SVD : \quad A(3) = U^{(3)} \Sigma^{(3)} V^{(3)^T} \]

\[
G = A \times_1 U^{(1)^T} \times_2 U^{(2)^T} \times_3 U^{(3)^T}
\]

Result:

\[
G_{(1)} = U^{(1)^T} A_{(1)} \left( U^{(3)^T} \otimes U^{(2)^T} \right) = \Sigma^{(1)} V^{(1)^T} \left( U^{(3)^T} \otimes U^{(2)^T} \right),
\]

\[
G_{(2)} = U^{(2)^T} A_{(2)} \left( U^{(3)^T} \otimes U^{(1)^T} \right) = \Sigma^{(2)} V^{(2)^T} \left( U^{(3)^T} \otimes U^{(1)^T} \right),
\]

\[
G_{(3)} = U^{(3)^T} A_{(3)} \left( U^{(2)^T} \otimes U^{(1)^T} \right) = \Sigma^{(3)} V^{(3)^T} \left( U^{(2)^T} \otimes U^{(1)^T} \right).
\]
Reduction

Allows reduction by deleting zero singular values in $\Sigma^{(i)}$, and resp. columns in $U^{(i)}$, and rows $V^{(i)}$.

Furthermore, we can get back the original $A$ via

$$A = G \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}$$

$U$ with orthonormal columns
$G$ is „all-orthogonal and ordered“
Proof:

\[ n \neq m : A \times_m U \times_n V = A \times_n V \times_m U \]

\[ A \times_n U \times_n V = A \times_n (VU) \]

\[ B = A \times_n U \iff B_{(n)} = U \cdot A_{(n)} \]

\[ G = A \times_1 U^{(1)^T} \times_2 U^{(2)^T} \times_3 U^{(3)^T} \]

\[ G \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} = \left( A \times_1 U^{(1)^T} \times_2 U^{(2)^T} \times_3 U^{(3)^T} \right) \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \]

\[ = A \times_1 \left( U^{(1)^T} U^{(1)} \right) \times_2 \left( U^{(2)^T} U^{(2)} \right) \times_3 \left( U^{(3)^T} U^{(3)} \right) \]

\[ = A \times_1 \mathbf{I} \times_2 \mathbf{I} \times_3 \mathbf{I} = A \]
Core Tensor $G$ "all-orthogonal":

$$A = G \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}$$

with the additional property

$$\langle G_{i,\ldots}, G_{j,\ldots} \rangle = 0 \quad \text{for} \quad i \neq j$$

Proof:

$$G_{(1)} = U^{(1)} A_{(1)} \left( U^{(3)} \otimes U^{(2)} \right)$$

$$G_{(1),i} = U^{(1)}_i A_{(1)} \left( U^{(3)} \otimes U^{(2)} \right)$$

$$\langle G_{(1),i}, G_{(1),j} \rangle = \left( U^{(3)} \otimes U^{(2)} \right)^T A_{(1)}^T \left( U^{(1)T}_i U^{(1)}_j \right) A_{(1)} \left( U^{(3)} \otimes U^{(2)} \right)$$

and similarly for index 2 and 3.
Properties

Mode-n singular values = norms of slices = sing.v. of $A_n$

Truncate by deleting small singular values/vectors

$$G = A \times_1 U^{(1)T} \times_2 U^{(2)T} \times_3 U^{(3)T} \rightarrow$$

$$\tilde{G} = A \times_1 \tilde{U}^{(1)T} \times_2 \tilde{U}^{(2)T} \times_3 \tilde{U}^{(3)T}$$

$$A \rightarrow \tilde{A} = \tilde{G} \times_1 \tilde{U}^{(1)} \times_2 \tilde{U}^{(2)} \times_3 \tilde{U}^{(3)}$$
Tucker Graphical

\[ A_{i_1, i_2, i_3, \ldots, i_N} \]

\[ \approx \]

\[ G_{m_1, m_2, m_3, \ldots, m_N} \]

\[ D_{small} \sum_{m_1, \ldots, m_N} G_{m_1, \ldots, m_N} U_{m_1i_1} \cdots U_{m_Ni_N}, \]
Three-way Tucker

\[ A = G \times_1 U \times_2 V \times_3 W = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{k=1}^{K} G_{pqk} u_p \odot v_q \odot w_k = [[G;U,V,W]] \]

\[ A_{ijm} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{k=1}^{K} G_{pqk} u_{ip} v_{jq} w_{mk}, \]

See slide 31+32:

\[ A_{(1)} = U \cdot G_{(1)} (W \otimes V)^T \]

\[ A_{(2)} = V \cdot G_{(2)} (W \otimes U)^T \]

\[ A_{(3)} = W \cdot G_{(3)} (V \otimes U)^T \]
N-way Tucker

\[ A = G \times_1 U^{(1)} \times_2 U^{(2)} \ldots \times_N U^{(N)} = \left[ [G; U^{(1)}, U^{(2)}, \ldots, U^{(N)}] \right] \]

\[ A_{i_1 i_2 \ldots i_N} = \sum_{k_1=1}^{R_1} \sum_{k_2=1}^{R_2} \ldots \sum_{k_N=1}^{R_N} G_{k_1 k_2 \ldots k_N} u^{(1)}_{i_1 k_1} u^{(2)}_{i_2 k_2} \ldots u^{(N)}_{i_N k_N}, \quad i_n = 1, \ldots, I_n \]

\[ A_{(n)} = U^{(n)} \cdot G_{(n)} \left( U^{(N)} \otimes \ldots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \ldots \otimes U^{(1)} \right)^T \]

Further definition:
Tucker1 is decomposition relative to only one index, Tucker2 relative to 2 indices, and Tucker relative to all indices.

\[ A_1 = G \times_1 U^{(1)} \times_2 I \ldots \times_N I = \left[ [G; U^{(1)}, I, \ldots, I] \right] \]
Computing the Tucker Dec.

For \( n=1, \ldots, N \)

\[ U^{(n)} := \text{matrix of left singular vectors of } A^{(n)} \]

\[ G := A \times_1 U^{(1)T} \times_2 U^{(2)T} \times_3 \cdots \times_N U^{(N)T} \]

Output: \( G, U^{(1)}, \ldots, U^{(N)} \).

We can use this algorithm also for approximating \( A \) by choosing in \( U^{(n)} \) only the dominant left singular vectors!

\[ R_n \rightarrow r_n \]
Approximating Tucker dec.

$$\min_{G,U^{(1)},...,U^{(N)}} \left\| A - [[[G;U^{(1)},...,U^{(N)}]]]\right\|$$

subject to $G \in IR^{r_1 \times \cdots \times r_N}$, $U^{(n)} \in IR^{I_n \times r_n}$ columnwise orthogonal

Rewrite as minimizing $\left\| vec(A) - \left(U^{(N)} \otimes \cdots \otimes U^{(1)}\right) vec(G)\right\|$

with solution $G := A \times_1 U^{(1)^T} \times_2 U^{(2)^T} \times_3 \cdots \times_N U^{(N)^T}$

Minimize: $\left\| A - [[[G;U^{(1)},...,U^{(N)}]]]\right\|^2$

$$= \left\| A \right\|^2 - 2\left\langle A, [[[G;U^{(1)},...,U^{(N)}]]]\right\rangle + \left\| [[[G;U^{(1)},...,U^{(N)}]]]\right\|^2$$

$$= \left\| A \right\|^2 - 2\left\langle A \times_1 U^{(1)^T} \times_2 \cdots \times_N U^{(N)^T}, G\right\rangle + \left\| G \right\|^2$$

$$= \left\| A \right\|^2 - 2\left\langle G, G\right\rangle + \left\| G \right\|^2 = \left\| A \right\|^2 - \left\| G \right\|^2$$

$$= \left\| A \right\|^2 - \left\| A \times_1 U^{(1)^T} \times_2 \cdots \times_N U^{(N)^T}\right\|^2$$

via ALS
ALS for Tucker

\[
\max_{U^{(n)}} \left\| A \times_1 U^{(1)T} \times_2 \cdots \times_N U^{(N)T} \right\|
\]

subject to \( U^{(n)} \in IR^{I \times r_n} \) columnwise orthogonal

\[
\max_{U^{(n)}} \left\| U^{(n)T} W \right\| \text{ with } W = A_{(n)} \left( U^{(N)} \otimes \cdots \otimes U^{(n+1)} \otimes U^{(n-1)} \otimes \cdots \otimes U^{(1)} \right)
\]

ALS method:
For \( n=1,\ldots,N \):
choose \( U^{(n)} \) the \( r_n \) dominant singular vectors of \( W \)
Repeat until convergence
Uniqueness

Tucker is not unique:

$$[[G; A, B, C]] = [[G \times_1 U \times_2 V \times_3 W; AU^{-1}, BV^{-1}, CW^{-1}]]$$
Application: Tensorfaces

Given a database of images of different persons, e.g. with different looks=expressions, illumination, positions=views. We can collect all the images in a big 5-leg tensor

\[ A = \left( a_{i_{\text{people}}, i_{\text{views}}, i_{\text{illum}}, i_{\text{express}}, i_{\text{pixel}}} \right) \]

In the example there is a database of 28 male persons with 5 poses, 3 illuminations, 3 expressions, each image 512x352 pixels. Hence, A is a 28x5x3x3x7943 tensor.
Database

28 subjects with 45 images per person.

Expression: smile

Illuminations

45 images for one person:
Principal Component Analysis
PCA

Use eigenfaces to capture the important features in a compact form. Often eigendecomposition and eigenvectors are used in PCA.

Here we use the Tucker decomposition:

\[ A = G \times_1 U_{\text{people}} \times_2 U_{\text{views}} \times_3 U_{\text{illum}} \times_4 U_{\text{express}} \times_5 U_{\text{pixels}} \]

resulting in matricification relative to pixel index

\[ A_{(\text{pixels})} = U_{\text{pixels}} G_{(\text{pixels})} \left( U_{\text{express}} \otimes U_{\text{illum}} \otimes U_{\text{views}} \otimes U_{\text{people}} \right)^T \]

\[ \text{image data} \quad \text{basis vectors} \quad \text{coefficients} \]
Interpretation

\[ A_{\text{pixels}} = U_{\text{pixels}} \cdot G_{\text{pixels}} \cdot (U_{\text{express}} \otimes U_{\text{illum}} \otimes U_{\text{views}} \otimes U_{\text{people}})^T \]

The mode matrix \( U_{\text{pixels}} \) can be interpreted as PCA.

By the core tensor \( G \) we can transform the eigenimages present in \( U_{\text{pixels}} \) into eigenmodes representing the principal axes of variation across the various factors (people, viewpoints, illuminations, expressions) by forming \( G \times^5 U_{\text{pixels}} \).
The first 10 PCA eigenvectors (eigenfaces) contained in the mode matrix $U_{\text{pixels}}$

„Multilinear Analysis of Image Ensembles: TensorFaces“ by M.A.O. Vasilescu and D. Terzopoulos

Similar paper on PCA on human motion via Tensors.