Tensor Ranks, and some Properties of Tensor Spaces

Pierre COMON

Lab. I3S, CNRS

and

Bernard MOURRAIN

INRIA institute

Collaborator: Lek-Heng LIM

☞ [1996], [June 2004 report] + results & proofs

see documents of the 2004 Tensor Decomposition Workshop:

http://csmr.ca.sandia.gov/~tgkolda/tdw2004

Tensors & Arrays

Definitions

Table \( T = \{ T_{ij..k} \} \)

- **Order** of \( T \) \( \overset{\text{def}}{=} \) \# of its ways = \# of its indices

- **Dimension** \( N_\ell \) \( \overset{\text{def}}{=} \) range of the \( \ell \)th index

- \( T \) is **Square** when all dimensions \( N_\ell = N \) are equal

- \( T \) is **Symmetric** when it is square and when its entries do not change by any permutation of indices
Tensors & Arrays

Properties

- Outer product $\mathbf{C} = \mathbf{A} \circ \mathbf{B}$:
  \[
  C_{ij,\ldots,ab,\ldots,d} = A_{ij,\ldots} B_{ab,\ldots,d}
  \]
  \textbf{Example:} outer product between 2 vectors: $\mathbf{u} \circ \mathbf{v} = \mathbf{u} \mathbf{v}^\mathsf{T}$

- Mode-1 inner product $\mathbf{A} \bullet_1 \mathbf{B}$:
  \[
  \{ \mathbf{A} \bullet_1 \mathbf{B} \}_{i_2,\ldots,i_M,j_2,\ldots,j_N} = \sum_k A_{k,i_2,\ldots,i_M} B_{k,j_2,\ldots,j_N}
  \]
  \textbf{Example:} matrix-vector product $\mathbf{A} \mathbf{u} = \mathbf{A}^\mathsf{T} \bullet_1 \mathbf{u}$

- Multilinearity. An order-3 tensor $\mathbf{T}$ is transformed by the multi-linear map $\{ \mathbf{A}, \mathbf{B}, \mathbf{C} \}$ into a tensor $\mathbf{T}'$:
  \[
  T'_{ijk} = \sum_{abc} A_{ia} B_{jb} C_{kc} T_{abc}
  \]
  Similarly: at any order.

Usefulness of $N$–way arrays

- Not much addressed in the literature before 1990
- Still hard (partly unsolved) numerical/theoretical problems
- Numerous areas of application
  - Speech
  - Mobile Communications
  - Machine Learning
  - Factor Analysis... $N$–way arrays
  - Biomedical Engineering
  - Psychometrics, Chemometrics...
Usefulness of symmetric arrays
Parafac vs ICA

PARAFAC:

\[
\begin{pmatrix}
\end{pmatrix} = \\
\begin{pmatrix}
\end{pmatrix} + \\
\begin{pmatrix}
\end{pmatrix} + \ldots
\]

PARAFAC cannot be used when:
- Lack of diversity
- Proportional slices
- Lack of physical meaning (e.g. video)

Then use Independent Component Analysis (ICA) [Comon’1991]

ICA: decompose a cumulant tensor instead of the data tensor

I3S

Edgeworth expansion
Approximation of a density

Francis Edgeworth (1845-1926).

\[
\frac{p_x(u)}{g_x(u)} = 1 + \frac{1}{3!} \kappa_3 h_3(v) + \frac{1}{4!} \kappa_4 h_4(v) + \frac{10}{6!} \kappa_3^2 h_6(v) + \frac{1}{5!} \kappa_5 h_5(v) + \frac{35}{7!} \kappa_3 \kappa_4 h_7(v) + \frac{280}{9!} \kappa_3^3 h_9(v) + \ldots
\]
ICA leads to tensor diagonalization

Cumulant tensors

Minimize statistical mutual dependence:

\[
I(p_x) = \int p_x(u) \log \frac{p_x(u)}{\prod_{i=1}^N p_{x_i}(u_i)} \, du.
\]

- **Expansion of the Mutual information**

\[
I(p_z) \approx J(p_z) - \frac{1}{48} \sum_i 4 \kappa_{iii}^2 + \kappa_{iii}^2 + 7 \kappa_{iii}^4 - 6 \kappa_{iii}^2 \kappa_{iii}
\]

- **Approximate minimization of the Mutual information**

\[
\min I(p_z) \approx \max \sum_i \kappa_{iii}^2 \text{ or } \max \sum_i \kappa_{iii}^2
\]

Maximization of diagonal terms in **symmetric** tensors \( \kappa_{ijk} \) or \( \kappa_{ijkl} \)

Definition of Rank

**CAND**

- Any tensor can always be decomposed (possibly non uniquely) as:

\[
T = \sum_{i=1}^r u(i) \circ v(i) \circ \ldots w(i)
\]

- **Tensor rank** \( \text{def} \) minimal \# of terms necessary

- This **Canonical decomposition** (CAND) holds valid in a **ring**

The CAND of a multilinear transform = the multilinear transform of the CAND:

- If \( T \xrightarrow{\phi} T' = T \circ_1 A \circ_2 B \circ_3 C \),

- then \( (u,v,\ldots,w) \xrightarrow{\phi} (Au,Bv,\ldots,Cw) \)
Spaces of tensors

dimensions

- \( \mathcal{A}_N \): square asymmetric of dimensions \( N \) and order \( d \)
  \( \Rightarrow \) dimension \( N^d \)

- \( \mathcal{S}_N \): square symmetric of dimensions \( N \) and order \( d \)
  \( \Rightarrow \) dimension \( D(N, d) = \binom{N+d-1}{d} \)

<table>
<thead>
<tr>
<th>( N \setminus d )</th>
<th>quadric 2</th>
<th>cubic 3</th>
<th>quartic 4</th>
<th>quintic 5</th>
<th>sextic 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>21</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>20</td>
<td>35</td>
<td>56</td>
<td>84</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>35</td>
<td>70</td>
<td>126</td>
<td>210</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>56</td>
<td>126</td>
<td>252</td>
<td>462</td>
</tr>
</tbody>
</table>

Number of free parameters in a symmetric tensor as a function of order \( d \) and dimension \( N \)

Ranks are difficult to evaluate

Clebsch theorem

The generic ternary quartic cannot in general be written as the sum of 5 fourth powers

- \( D(3, 4) = 15 \)
- \( 3r \) free parameters in the CAND
- But \( r = 5 \) is not enough \( \rightarrow r = 6 \) is generic
Literature
Polynomials

Gauss’1825
Sylvester’1851
Cayley’1854
Clebsch’1861
Salmon’1874
Poincaré’1890
Hilbert’1900
Wakeford’1918
Grothendieck’1966
Dieudonné’1970
Shafarevich’1975

Ehrenborg, Mourrain, Kogan...

I3S

Literature
N-Way arrays

Tucker’1966
Harshman’1970
Caroll’1970
Kruskal’1977
Kroonenberg’1980
Leurgans’1993

Delathauwer, Sidiropoulos, ten Berge, Regalia, Bro, Stegeman, Golub, ...
Literature
Tensors & Polynomials

Howell’1978
Atkinson’1980
Strassen’1983
Rota’1984
Weinstein’1984
Lickteig’1985
Reznick’1992

Questions

1. Maximal rank in \( S_N \) or \( A_N \)
2. Generic rank \( S_N \)
3. Typical ranks of \( A_N \)
4. Bounds on ranks
5. Rank and CAND of a given tensor
6. Extract a large number of factors from a reduced-diversity array
7. Differences between \( \mathbb{R} \) and \( \mathbb{C} \)
Tensors and Polynomials

Bijection

- Symmetric tensor of order $d$ and dimension $N$ can be associated with a unique homogeneous polynomial of degree $d$ in $N$ variables:

$$p(x) = \sum_j T_j \cdot x^{f(j)}$$

1. integer vector $j$ of dimension $d \leftrightarrow$ integer vector $f(j)$ of dimension $N$
2. entry $f_k$ of $f(j)$ being $\text{def} = \#$ of times index $k$ appears in $j$
3. We have in particular $|f(j)| = d$.

- Standard conventions $x^j \text{def} = \prod_{k=1}^N x_k^{j_k}$ and $|f| \text{def} = \sum_{k=1}^N f_k$, where $j$ and $f$ are integer vectors.

**Example:** $T =$ \begin{array}{ccc}v & & v \\
& v & \\
& & v \\
& & \\
& & & v \\
\end{array} $ \leftrightarrow p(x) = 3x^{[2,1]} = 3x_1^2x_2$

Orbits

Definition

- General Linear group $\mathcal{GL}$: group of invertible matrices
- Orbit of a polynomial $p$: all polynomials $q$ that can be transformed into $p$ by $A \in \mathcal{GL}$: $q(x) = p(Ax)$.
- Allows to classify polynomials
Quadrics

quadratic homogeneous polynomials

- Binary quadrics ($2 \times 2$ symmetric matrices)
  - Orbits in $\mathbb{R}$: $\{0, x^2, x^2 + y^2, x^2 - y^2\}$
    $\ni 2xy \in O(x^2 - y^2)$ in $\mathbb{R}[x, y]$
  - Orbits in $\mathbb{C}$: $\{0, x^2, x^2 + y^2\}$
    $\ni 2xy \in O(x^2 + y^2)$ in $\mathbb{C}[x, y]$

- Set of singular matrices is closed
- Set $\mathcal{Y}_r$ of matrices of at most rank $r$ is closed

Classification of ternary cubics

$3 \times 3 \times 3$

<table>
<thead>
<tr>
<th>$GL$-orbit</th>
<th>$\omega(p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^3$</td>
<td>1</td>
</tr>
<tr>
<td>$x^2y + xy^2$</td>
<td>2</td>
</tr>
<tr>
<td>$x^2y$</td>
<td>3</td>
</tr>
<tr>
<td>$x^3 + 3y^2z$</td>
<td>4</td>
</tr>
<tr>
<td>$x^3 + y^3 + 6xyz$</td>
<td>4</td>
</tr>
<tr>
<td>$x^3 + 6xyz$</td>
<td>4</td>
</tr>
<tr>
<td>$a(x^3 + y^3 + z^3) + 6bxyz$</td>
<td>4 ($\text{generic}$)</td>
</tr>
<tr>
<td>$xz^2 + y^2z$</td>
<td>5</td>
</tr>
</tbody>
</table>

Maximal rank George Salmon (1819-1904)
Topology of polynomials

**definition**

- Every elementary closed set $\overset{\text{def}}{=} \text{varieties, defined by } p(x) = 0$
- Closed sets = finite union of varieties
- Closure of a set $\mathcal{E}$: smallest closed set $\overline{\mathcal{E}}$ containing $\mathcal{E}$

◿ this is not Euclidian topology, called Zariski in $\mathbb{C}$

---

**Genericity**

Definitions proposed jointly with L-H.Lim

**Intuitive**

- A property is *typical* $\iff$ is is true on a non zero volume set
- A property is *generic* $\iff$ is is true almost everywhere

**Mathematical**

- $r$ is not typical if (zero volume):
  $\mathcal{Z}_r$ is contained in a non trivial closed set

  or

- $r$ is a typical rank if (density argument with Zariski):
  $\mathcal{Z}_r$ is the whole space

- Generic rank: *the typical rank when unique*
Tensor subsets

- Set of tensors of rank at most \( r \) with values in \( \mathbb{C} \):
  \[ \mathcal{Y}_r = \{ T \in T : r(T) \leq r \} \]
- Set of tensors of rank exactly \( r \): \( \mathcal{Z}_r = \{ T \in T : r(T) = r \} \)
  \[ \mathcal{Z} = \mathcal{Y}_r - \mathcal{Y}_{r-1}, \ r > 1 \]
- \( \mathcal{Z}_1 \) is closed but not \( \mathcal{Z}_r, \ r > 1 \)
- Zariski closures: \( \overline{\mathcal{Y}_r}, \overline{\mathcal{Z}_r} \).

---

Example of sequence proving lack of closure of \( \mathcal{Y}_r \) for \( r > 1 \) outlined by L-H. Lim

Sequence of rank-2 tensors converging towards a rank-3:
\[
T_n = x_1 \circ x_2 \circ \left( \frac{1}{n} x_3 - y_3 \right) + (x_1 + \frac{1}{n} y_1) \circ (x_2 + \frac{1}{n} y_2) \circ y_3
\]
In fact:
\[
T_n = \frac{1}{n} \left[ x_1 \circ x_2 \circ x_3 + y_1 \circ x_2 \circ y_3 + x_1 \circ y_2 \circ y_3 \right] + \frac{1}{n^2} y_1 \circ y_2 \circ y_3
\]

**NB:** even possible to jump from rank \( r \) to rank \( r + 2 \)
(joint proof under development).
Generic rank in \( \mathbb{C} \)

joint work with B. Mourrain

- **Lemma** (in either \( \mathbb{R} \) of \( \mathbb{C} \), either symmetric or not)

  Strictly increasing series of \( \mathcal{Y}_k \) for \( k \leq \overline{R} \), then constant:

\[
\overline{\mathcal{Y}}_1 \subsetneq \overline{\mathcal{Y}}_2 \subsetneq \ldots \subsetneq \overline{\mathcal{Y}}_{\overline{R}} = \overline{\mathcal{Y}}_{\overline{R}+1} = \ldots = \mathcal{T}
\]

which guarantees the existence of a unique \( \overline{R} \)

- **Theorem 1** For tensors in \( \mathbb{C} \)

  If \( r_1 < r_2 < \overline{R} \), then

\[
\overline{Z}_{r_1} \subset \overline{Z}_{r_2} \subset \overline{Z}_{\overline{R}}
\]  

(2)

- **Theorem 2** For tensors in \( \mathbb{C} \)

  If \( \overline{R} < r_3 \leq R \), then

\[
\overline{Z}_{\overline{R}} \supset \overline{Z}_{r_3} \supset \overline{Z}_{R}
\]

**Prove that** \( \overline{R} \) **is the generic rank in** \( \mathbb{C} \)

---

**I3S**

2004 – 23/35 – P. Comon

Generic rank

e.g. binary quartics in \( \mathbb{C} \)

\[
\begin{align*}
\mathbb{I}_3 \mathbb{S} & \quad \mathcal{Z}_1 \\
\mathbb{I}_3 \mathbb{S} & \quad \mathcal{Z}_2 = \mathcal{Y}_2 - \mathcal{Z}_1 \\
\mathbb{I}_3 \mathbb{S} & \quad \mathcal{Z}_3 = \mathcal{Y}_3 - \mathcal{Z}_1 - \mathcal{Z}_2 = \mathcal{T} - \mathcal{Z}_1 - \mathcal{Z}_2 - \mathcal{Z}_4 \\
\mathbb{I}_3 \mathbb{S} & \quad \mathcal{Z}_4 = \mathcal{Y}_4 - \mathcal{Y}_3 
\end{align*}
\]
Generic rank in $\mathbb{C}$

symmetric tensors

<table>
<thead>
<tr>
<th>order $d$</th>
<th>$2$</th>
<th>$3$</th>
<th>$4$</th>
<th>$5$</th>
<th>$6$</th>
<th>$2$</th>
<th>$3$</th>
<th>$4$</th>
<th>$5$</th>
<th>$6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>dim. $N$</td>
<td>$2$</td>
<td>$3$</td>
<td>$3$</td>
<td>$3$</td>
<td>$4$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
</tr>
<tr>
<td></td>
<td>$3$</td>
<td>$4$</td>
<td>$6$</td>
<td>$7$</td>
<td>$10$</td>
<td>$3$</td>
<td>$2$</td>
<td>$3$</td>
<td>$0$</td>
<td>$2$</td>
</tr>
<tr>
<td></td>
<td>$4$</td>
<td>$5$</td>
<td>$10$</td>
<td>$14$</td>
<td>$22$</td>
<td>$6$</td>
<td>$0$</td>
<td>$5$</td>
<td>$0$</td>
<td>$4$</td>
</tr>
<tr>
<td></td>
<td>$5$</td>
<td>$6$</td>
<td>$15$</td>
<td>$26$</td>
<td>$42$</td>
<td>$10$</td>
<td>$5$</td>
<td>$5$</td>
<td>$4$</td>
<td>$0$</td>
</tr>
<tr>
<td></td>
<td>$6$</td>
<td>$7$</td>
<td>$22$</td>
<td>$42$</td>
<td>$77$</td>
<td>$15$</td>
<td>$4$</td>
<td>$6$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td></td>
<td>$7$</td>
<td>$8$</td>
<td>$30$</td>
<td>$66$</td>
<td>$132$</td>
<td>$21$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td></td>
<td>$8$</td>
<td>$9$</td>
<td>$42$</td>
<td>$99$</td>
<td>$215$</td>
<td>$28$</td>
<td>$0$</td>
<td>$6$</td>
<td>$0$</td>
<td>$4$</td>
</tr>
</tbody>
</table>

[Comon-Mourrain'1996]

Typical ranks in $\mathbb{R}$

Lack of uniqueness in $\mathbb{R}$

- Draw randomly entries of a tensor $\in \mathcal{T}(N, d)$ according to a distribution $q(t)$
- Typical ranks do not depend on $q(t)$, if c.d.f. absolutely continuous (no point-like mass). Only volumes of $Z_r$ do.
- Typical ranks depend on $(N, d)$

**Example:** $2 \times 2 \times 2$ asymmetric tensors
- drawn according to Gaussian symmetric $\Rightarrow \{2(57\%), 3(43\%)\}$
- drawn according to Gaussian asymmetric $\Rightarrow \{2(80\%), 3(20\%)\}$
Ranks in $\mathbb{R}$

vs rank in $\mathbb{C}$

- For any real tensor $T$, rank in $\mathbb{R}$ is always larger than rank in $\mathbb{C}$:
  
  \[ \text{rank}^\mathbb{C}(T) \leq \text{rank}^\mathbb{R}(T) \]

- In particular:
  
  generic rank $\leq$ typical ranks

**Example:**

\[
T(:, :, 1) = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad T(:, :, 2) = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix},
\]

- If decomposed in $\mathbb{R}$, it is of rank 3:

\[
T = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \circ^3 + \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \circ^3 - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \circ^3
\]

- Whereas it admits a CAND of rank 2 in $\mathbb{C}$:

\[
T = \frac{j}{2} \begin{pmatrix} j \\ 1 \end{pmatrix} \circ^3 - \frac{j}{2} \begin{pmatrix} -j \\ 1 \end{pmatrix} \circ^3
\]

### Symmetric vs Asymmetric rank

joint work with L-H.Lim

- Let $T \in \mathcal{S}$ symmetric tensor, and its CAND:

\[
T = \sum_{k=1}^{r} T_k
\]

where $T_k$ are rank-1.

**Theorem**

If the constraint $T_k \in \mathcal{S}$ is relaxed, then the rank is still the same

- But $T_k$'s need not be each symmetric when solution is not essentially unique
Bounds (1)

asymmetric $\mathbb{C}$

- Tensors of order $d$ and dimensions $(N_1, \ldots, N_d)$:
  - Upper bound
    \[
    \left\lfloor \frac{\prod_{i=1}^{d} N_i}{1 + \sum_{i=1}^{d} (N_i - 1)} \right\rfloor \leq \overline{R}
    \]
  - Square case $K_i = N$:
    \[N^d/(dN - d + 1) \leq \overline{R}\]

- Lower bound (Square case):
  \[N^d/(dN - d + 1) \leq \overline{R}\]

---

Bounds (2)

Symmetric $\mathbb{C}$

- Lower bound
  \[\left\lfloor \frac{\binom{N + d - 1}{d}}{N} \right\rfloor \leq \overline{R}\]

- Upper bound [Reznick’92]
  \[\overline{R} \leq \binom{N + d - 2}{d - 1}\]
Construction of the CAND (1)

**2x2x...x2**

**Sylvester’s theorem in** $\mathbb{R}$

- A binary quantic $p(x, y) = \sum_{i=0}^{d} \gamma_i c(i) x^i y^{d-i}$ can be decomposed in $\mathbb{R}[x, y]$ into a sum of $r$ powers as $p(x, y) = \sum_{j=1}^{r} \lambda_j (\alpha_j x + \beta_j y)^d$ if and only if the form

$$q_c(x, y) = \prod_{j=1}^{r} (\beta_j x - \alpha_j y) = \sum_{l=0}^{r} g_l x^l y^{r-l}$$

satisfies

$$\begin{bmatrix}
\gamma_0 & \gamma_1 & \cdots & \gamma_r \\
\gamma_1 & \gamma_2 & \cdots & \gamma_{r+1} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{d-r} & \cdots & \gamma_d
\end{bmatrix}
\begin{bmatrix}
g_0 \\
g_1 \\
\vdots \\
g_r
\end{bmatrix} = 0.$$ 

- Valid even in non generic cases.
- Similar theorem in $\mathbb{C}$ (cf. appendix)

---

Construction of the CAND (2)

**2x2x...x2**

- Start with $r = 1$ ($d \times 2$ matrix)
  and increase $r$ until it looses its column rank

```
1 2  
2 3  
3 4  
4 5  
5 6  
6 7  
7 8  
```

$\rightarrow$

```
1 2 3  
2 3 4  
3 4 5  
4 5 6  
5 6 7  
6 7 8  
```

```
1 2 3 4  
2 3 4 5  
3 4 5 6  
4 5 6 7  
5 6 7 8  
```

---

I3S
Algorithms
Large rank cases

- If rank sub-generic: use ALS or accelerations
- Otherwise, build another tensor of sub-generic rank: use BIOME algorithm

Future works
Open questions

- How many typical ranks can exist for \( \mathbb{R} \) tensors?
  Conjecture: at most 2
- Algorithm to compute generic rank for \( \mathbb{C} \) asymmetric tensors
- Maximal achievable ranks?
- What does "low-rank approximation" means for tensors when rank > 1?
- General algorithm for computing a CAND
- Definition of eigen-uplets of tensors
**Appendix**

2x2x...x2

Sylvester’s theorem in $\mathbb{C}$

A binary quantic $p(x, y) = \sum_{i=0}^{d} c(i) \gamma_{i} x^{i} y^{d-i}$ can be written as a sum of $d$th powers of $r$ distinct linear forms:

$$p(x, y) = \sum_{j=1}^{r} \lambda_{j} (\alpha_{j} x + \beta_{j} y)^{d},$$

(3)

if and only if **(i)** there exists a vector $\mathbf{g}$ of dimension $r + 1$, with components $g_{\ell}$, such that

$$\begin{bmatrix}
\gamma_{0} & \gamma_{1} & \cdots & \gamma_{r} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{d-r} & \cdots & \gamma_{d-1} & \gamma_{d}
\end{bmatrix} \mathbf{g}^{\ast} = \mathbf{0}.$$  

(4)

and **(ii)** the polynomial $q(x, y) \overset{\text{def}}{=} \sum_{\ell=0}^{r} g_{\ell} x^{\ell} y^{r-\ell}$ admits $r$ distinct roots.