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

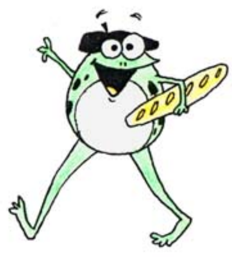
see documents of the 2004 Tensor Decomposition Workshop:
http://csmr.ca.sandia.gov/∼tgkolda/tdw2004

Tensors & Arrays

Definitions

Table $\boldsymbol{T} = \{T_{ij..k}\}$

- *Order* of $\boldsymbol{T} \stackrel{\text{def}}{=}$ # of its ways = # of its indices
- *Dimension* $N_\ell \stackrel{\text{def}}{=}$ range of the ℓth index
- $\boldsymbol{T}$ is *Square* when all dimensions $N_\ell = N$ are equal
- $\boldsymbol{T}$ is *Symmetric* when it is square and when its entries do not change by *any* permutation of indices



Tensors & Arrays

Properties

- Outer product $\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$:

$$C_{ij..\ell\,ab..d} = A_{ij..\ell}\, B_{ab..d}$$

➨ *Example:* outer product between 2 vectors: $\boldsymbol{u} \circ \boldsymbol{v} = \boldsymbol{u}\,\boldsymbol{v}^{\mathsf{T}}$

- Mode-1 inner product $\boldsymbol{A} \bullet_1 \boldsymbol{B}$:

$$\{\boldsymbol{A} \underset{1}{\bullet} \boldsymbol{B}\}_{i_2..i_M,j_2..j_N} = \sum_k A_{ki_2..i_M}\, B_{kj_2..j_N}$$

Similarly: mode$-p$ inner product $\boldsymbol{A} \bullet_p \boldsymbol{B}$

➨ *Example:* matrix-vector product $\boldsymbol{A}\,\boldsymbol{u} = \boldsymbol{A}^{\mathsf{T}} \bullet_1 \boldsymbol{u}$

- Multilinearity. An order-3 tensor $\boldsymbol{T}$ is transformed by the multi-linear map $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}\}$ into a tensor $\boldsymbol{T}'$:

$$T'_{ijk} = \sum_{abc} A_{ia} B_{jb} C_{kc} T_{abc}$$

Similarly: at any order.

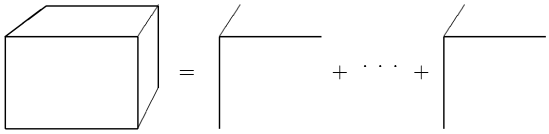
Usefulness of $N-$way arrays

Introduction

- 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...

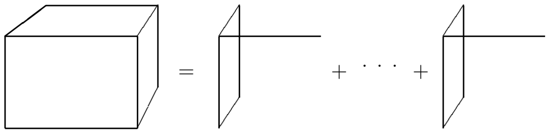
Usefulness of symmetric arrays

Parafac vs ICA

- **PARAFAC:** 

PARAFAC cannot be used when:

- Lack of diversity
- Proportional slices
- Lack of physical meaning (e.g.video)



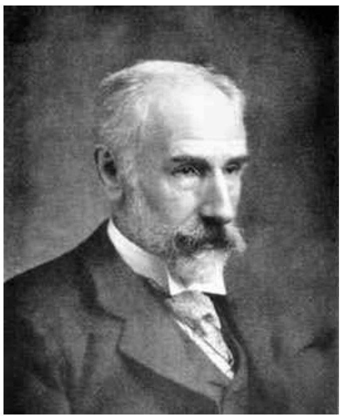
- Then use *Independent Component Analysis* (ICA) [Comon'1991]

ICA: decompose a *cumulant tensor* instead of the data tensor

- INDSCAL

Edgeworth expansion

Approximation of a density



Francis Edgeworth (1845-1926).

$$\begin{aligned}\frac{p_x(u)}{g_x(u)} &= 1 + \frac{1}{3!}\,\kappa_3\,h_3(v) + \frac{1}{4!}\,\kappa_4\,h_4(v) + \frac{10}{6!}\,\kappa_3^2\,h_6(v) \\ &\quad + \frac{1}{5!}\,\kappa_5\,h_5(v) + \frac{35}{7!}\,\kappa_3\kappa_4\,h_7(v) + \frac{280}{9!}\kappa_3^3\,h_9(v) + \dots\end{aligned}$$

ICA leads to tensor diagonalization

Cumulant tensors

Minimize statistical mutual dependence:

$$I(p_{\boldsymbol{x}}) = \int p_{\boldsymbol{x}}(\boldsymbol{u}) \log \frac{p_{\boldsymbol{x}}(\boldsymbol{u})}{\prod_{i=1}^{N} p_{x_i}(u_i)} \, d\boldsymbol{u}.$$

- **Expansion of the Mutual information**

$$I(p_{\boldsymbol{z}}) \approx J(p_{\boldsymbol{z}}) - \frac{1}{48} \sum_i 4\,\kappa_{iii}^2 + \kappa_{iiii}^2 + 7\,\kappa_{iii}^4 - 6\,\kappa_{iii}^2\,\kappa_{iiii}$$

- **Approximate minimization of the Mutual information**

$$\text{Min}\, I(p_z) \sim \text{Max} \sum_i \kappa_{iii}^2 \text{ or } \text{Max} \sum_i \kappa_{iiii}^2$$

➡ Maximization of diagonal terms in *symmetric* tensors κ_{ijk} or $\kappa_{ijk\ell}$

Definition of Rank

CAND

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

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

- *Tensor rank* $\stackrel{\text{def}}{=}$ 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 $\boldsymbol{T} \xrightarrow{\mathcal{L}} \boldsymbol{T}' = \boldsymbol{T} \bullet_1 \boldsymbol{A} \bullet_2 \boldsymbol{B} \bullet_3 \boldsymbol{C}$,
  - then $(\boldsymbol{u}, \boldsymbol{v}, ..\boldsymbol{w}) \xrightarrow{\mathcal{L}} (\boldsymbol{A}\boldsymbol{u}, \boldsymbol{B}\boldsymbol{v}, ..\boldsymbol{C}\boldsymbol{w})$

Spaces of tensors

dimensions

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

N\d	quadric 2	cubic 3	quartic 4	quintic 5	sextic 6
2	3	4	5	6	7
3	6	10	15	21	28
4	10	20	35	56	84
5	15	35	70	126	210
6	21	56	126	252	462

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

Ranks are difficult to evaluate

Clebsch theorem

Alfred Clebsch (1833-1872)

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

- $D(3,4) = 15$
- $3\,r$ 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...

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

Comon, Lim...

Questions

1. Maximal rank in $\mathcal{S}_N$ or $\mathcal{A}_N$
2. Generic rank $\mathcal{S}_N$
3. Typical ranks of $\mathcal{A}_N$
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

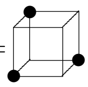
Bijection

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

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{j}} T_{\boldsymbol{j}}\, \boldsymbol{x}^{\boldsymbol{f}(\boldsymbol{j})} \tag{1}$$

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

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

➨ *Example:* $\boldsymbol{T} =$  $\leftrightarrow\ p(\boldsymbol{x}) = 3\,\boldsymbol{x}^{[2,1]} = 3\,x_1^2 x_2$

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 $\boldsymbol{A} \in \mathcal{GL}$: $q(\boldsymbol{x}) = p(\boldsymbol{A}\boldsymbol{x})$.
- Allows to classify polynomials

Quadrics

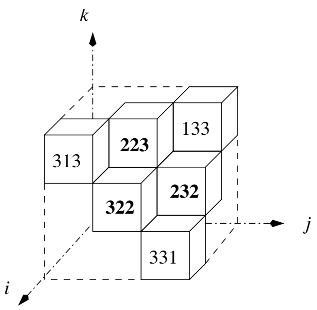
quadratic homogeneous polynomials

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

$3 \times 3 \times 3$

$\mathcal{GI}-$orbit	$\omega(p)$
x^3	1
$x^2y + xy^2$	2
x^2y	3
$x^3 + 3\,y^2z$	4
$x^3 + y^3 + 6\,xyz$	4
$x^3 + 6\,xyz$	4
$a\,(x^3+y^3+z^3) + 6b\,xyz$	4 (**generic**)
$xz^2 + y^2z$	5



Maximal rank

George Salmon (1819-1904)

Topology of polynomials

definition

- Every elementary closed set $\stackrel{\text{def}}{=}$ varieties, defined by $p(\boldsymbol{x}) = 0$
- 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* $\Leftrightarrow$ is is true on a non zero volume set
- A property is *generic* $\Leftrightarrow$ 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):

  $\overline{\mathcal{Z}}_r$ is the whole space

- Generic rank: *the typical rank when unique*

Tensor subsets

- Set of tensors of rank *at most* r with values in $\mathbb{C}$:

$$\mathcal{Y}_r = \{\boldsymbol{T} \in \mathcal{T} : \; r(\boldsymbol{T}) \leq r\}$$

- Set of tensors of rank *exactly* r: $\mathcal{Z}_r = \{\boldsymbol{T} \in \mathcal{T} : \; r(\boldsymbol{T}) = r\}$

$$\mathcal{Z} = \mathcal{Y}_r - \mathcal{Y}_{r-1}, \; r > 1$$

- $\mathcal{Z}_1$ 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:

$$\boldsymbol{T}_n = \boldsymbol{x}_1 \circ \boldsymbol{x}_2 \circ (\frac{1}{n}\boldsymbol{x}_3 - \boldsymbol{y}_3) + (\boldsymbol{x}_1 + \frac{1}{n}\boldsymbol{y}_1) \circ (\boldsymbol{x}_2 + \frac{1}{n}\boldsymbol{y}_2) \circ \boldsymbol{y}_3$$

In fact:

$$\boldsymbol{T}_n = \frac{1}{n}[\boldsymbol{x}_1 \circ \boldsymbol{x}_2 \circ \boldsymbol{x}_3 + \boldsymbol{y}_1 \circ \boldsymbol{x}_2 \circ \boldsymbol{y}_3 + \boldsymbol{x}_1 \circ \boldsymbol{y}_2 \circ \boldsymbol{y}_3] + \frac{1}{n^2}\boldsymbol{y}_1 \circ \boldsymbol{y}_2 \circ \boldsymbol{y}_3$$

Lek-Heng Lim

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 $\overline{\mathcal{Y}}_k$ for $k \leq \overline{R}$, then constant:

$$\overline{\mathcal{Y}}_1 \subset_{\neq} \overline{\mathcal{Y}}_2 \subset_{\neq} \dots \subset_{\neq} \overline{\mathcal{Y}}_{\overline{R}} = \overline{\mathcal{Y}}_{\overline{R}+1} = \dots \mathcal{T}$$

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

- **Theorem 1** For tensors in $\mathbb{C}$
  If $r_1 < r_2 < \overline{R}$, then

$$\overline{\mathcal{Z}}_{r_1} \subset \overline{\mathcal{Z}}_{r_2} \subset \overline{\mathcal{Z}}_{\overline{R}} \qquad (2)$$

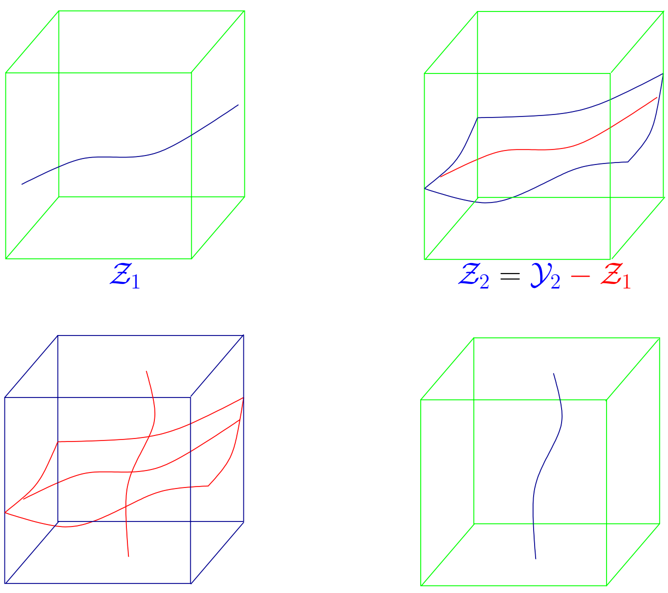
- **Theorem 2** For tensors in $\mathbb{C}$
  If $\overline{R} < r_3 \leq R$, then

$$\overline{\mathcal{Z}}_{\overline{R}} \supset \overline{\mathcal{Z}}_{r_3} \supseteq \overline{\mathcal{Z}}_R$$

➨ Prove that $\overline{R}$ is the generic rank in $\mathbb{C}$

Generic rank

e.g. binary quartics in $\mathbb{C}$



$\mathcal{Z}_1$

$\mathcal{Z}_2 = \mathcal{Y}_2 - \mathcal{Z}_1$

$$\begin{aligned} \mathcal{Z}_3 &= \mathcal{Y}_3 - \mathcal{Z}_1 - \mathcal{Z}_2 \\ &= \mathcal{T} - \mathcal{Z}_1 - \mathcal{Z}_2 - \mathcal{Z}_4 \end{aligned}$$

$\mathcal{Z}_4 = \mathcal{Y}_4 - \mathcal{Y}_3$

Generic rank in $\mathbb{C}$

symmetric tensors

order d / dim. N	Generic rank $\overline{R}$: 2	3	4	5	6	Dim. of solution: 2	3	4	5	6
2	2	2	3	3	4	1	0	1	0	1
3	3	4	6	7	10	3	2	3	0	2
4	4	5	10	14	22	6	0	5	0	4
5	5	8	15	26	42	10	5	5	4	0
6	6	10	22	42	77	15	4	6	0	0
7	7	12	30	66	132	21	0	0	0	0
8	8	15	42	99	215	28	0	6	0	4

[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 $\mathbb{C}$

- $\forall \boldsymbol{T}$ real tensor, rank in $\mathbb{R}$ always larger than rank in $\mathbb{C}$:

$$rank^{\mathbb{C}}(\boldsymbol{T}) \leq rank^{\mathbb{R}}(\boldsymbol{T})$$

- In particular:

generic rank $\leq$ typical ranks

Example:

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

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

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

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

$$\boldsymbol{T} = \frac{\jmath}{2}\begin{pmatrix} -\jmath \\ 1 \end{pmatrix}^{\circ 3} - \frac{\jmath}{2}\begin{pmatrix} \jmath \\ 1 \end{pmatrix}^{\circ 3}$$

Symmetric vs Asymmetric rank

joint work with L-H.Lim

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

$$\boldsymbol{T} = \sum_{k=1}^{r} \boldsymbol{T}_k$$

where $\boldsymbol{T}_k$ are rank-1.

- **Theorem**

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

- But $\boldsymbol{T}_k$'s need not be each symmetric when solution is not essentially unique

Bounds (1)

asymmetric $\mathbb{C}$

- Tensors of order d and dimensions $(N_1, ..N_d)$:
  - Upper bound

$$\left\lceil \frac{\prod_{i=1}^{d} N_i}{1 + \sum_{i=1}^{d}(N_i - 1)} \right\rceil \leq \overline{R}$$

  - Square case $K_i = N$:

$$N^d/(dN - d + 1) \leq \overline{R}$$

- Lower bound (Square case):

$$N^d/(d\,N - d + 1) \leq \overline{R}$$

Bounds (2)

Symmetric $\mathbb{C}$

- Lower bound

$$\left\lceil \frac{\binom{N+d-1}{d}}{N} \right\rceil \leq \overline{R}$$

- 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

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

and has distinct real roots.

- 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

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 2 & 3 \\ \hline 3 & 4 \\ \hline 4 & 5 \\ \hline 5 & 6 \\ \hline 6 & 7 \\ \hline 7 & 8 \\ \hline \end{array} \longrightarrow \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 2 & 3 & 4 \\ \hline 3 & 4 & 5 \\ \hline 4 & 5 & 6 \\ \hline 5 & 6 & 7 \\ \hline 6 & 7 & 8 \\ \hline \end{array} \longrightarrow \begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline 2 & 3 & 4 & 5 \\ \hline 3 & 4 & 5 & 6 \\ \hline 4 & 5 & 6 & 7 \\ \hline 5 & 6 & 7 & 8 \\ \hline \end{array}$$

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 dth powers of r distinct linear forms:

$$p(x,y) = \sum_{j=1}^{r} \lambda_j\, (\alpha_j\, x + \beta_j\, y)^d, \tag{3}$$

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

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_r \\ \vdots & & & \vdots \\ \gamma_{d-r} & \cdots & \gamma_{d-1} & \gamma_d \end{bmatrix} \boldsymbol{g}^* = \mathbf{0}. \tag{4}$$

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