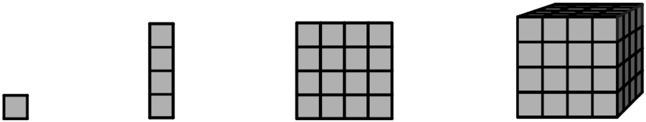
from linear algebra to multi-linear algebra



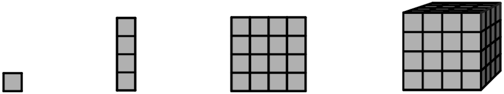
Anna Seigal

University of Oxford

9th March 2021

overview

- extending linear algebra to multi-linear algebra
  - challenges
  - algebraic properties of tensors



- focus: aspects of the singular value decomposition
  - singular vectors
  - notions of rank
  - low-rank approximation
  - analogues for symmetric tensors

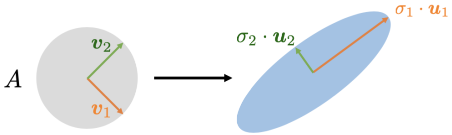
the singular value decomposition

decomposes a matrix $A \in \mathbb{R}^{n_1 n_2}$ as

$$A = U \Sigma V^{\mathsf{T}},$$

$U \in \mathbb{R}^{n_1 n_1}, V \in \mathbb{R}^{n_2 n_2}$ orthogonal, $\Sigma \in \mathbb{R}^{n_1 n_2}$ diagonal, non-negative.

singular vector pair (u_i, v_i) with singular value σ_i.



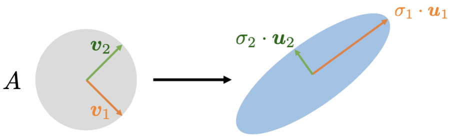
in tensor product notation: $A \in \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$

$$A = \sum_{i=1}^{n} \sigma_i u_i \otimes v_i^*.$$

[1] A. Edelman, Y. Wang (2020).

properties of the SVD

- finds (orthogonal) basis so that A is diagonal
- matrix A is linear map $\mathbb{R}^{n_2} \to \mathbb{R}^{n_1}$, $v \mapsto Av$.



$$v_i \mapsto \sigma_i u_i.$$

- rank of A is number of non-zero singular values $\sum_{i=1}^{n} \sigma_i u_i \otimes v_i^*$
- best rank r approximation is truncation to top r singular values[2]

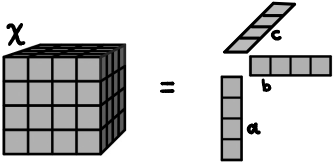
$$\sum_{i=1}^{r} \sigma_i u_i \otimes v_i^*$$

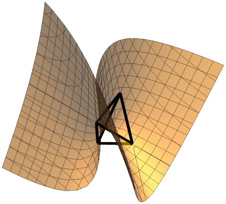
[2] C. Eckart, G. Young. "The approximation of one matrix by another of lower rank." (1936).

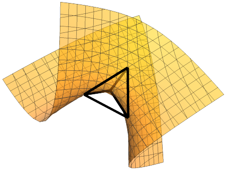
tensor rank

$X = (x_{ijk})$ order three tensor, $X \in V_1 \otimes V_2 \otimes V_3$.

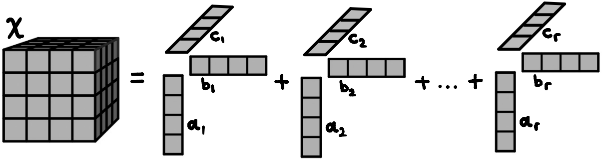
X has *rank* 1 if $\quad X = a \otimes b \otimes c \quad$ i.e. $x_{ijk} = a_i b_j c_k$.

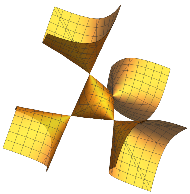






X has *rank r* if it is the sum of r rank 1 tensors, $\quad X = \sum_{i=1}^{r} a_i \otimes b_i \otimes c_i$.





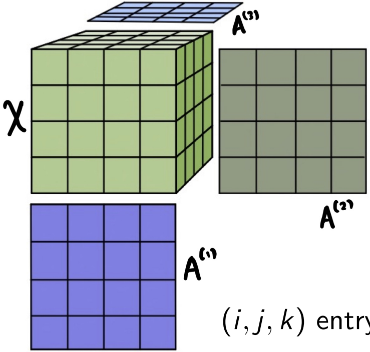
[later: different types of rank]

change of basis

tensor $X \in V_1 \otimes V_2 \otimes V_3$ (or $V_1 \otimes \cdots \otimes V_d$)

multiply by matrix $A^{(\ell)}$ in each vector space V_ℓ.

$[\![X; A^{(1)}, A^{(2)}, A^{(3)}]\!]$



(i, j, k) entry $\sum_{\alpha,\beta,\gamma} x_{\alpha\beta\gamma} a^{(1)}_{i\alpha} a^{(2)}_{j\beta} a^{(3)}_{k\gamma}$

multi-linear map

matrix $A \in V_1 \otimes V_2$ gives linear and bi-linear maps

$$\begin{array}{ccc} V_2^* & \to & V_1 \\ v & \mapsto & Av \\ & & \sum_j a_{ij} v_j \\ & & [\![A;\, \mathrm{I}\,,\, v]\!] \end{array} \qquad \begin{array}{ccc} V_1^* & \to & V_2 \\ w & \mapsto & A^{\mathsf{T}} w \\ & & \sum_i a_{ij} w_i \\ & & [\![A;\, w,\, \mathrm{I}\,]\!] \end{array} \qquad \begin{array}{ccc} V_1^* \times V_2^* & \to & \mathbb{K} \\ (w, v) & \mapsto & w^{\mathsf{T}} A v \\ & & \sum_{i,j} a_{ij} w_i v_j \\ & & [\![A;\, w,\, v]\!] \end{array}$$

tensor $X \in V_1 \otimes \cdots \otimes V_d$ gives multi-linear maps

e.g.

$$\begin{array}{ccc} V_d^* & \to & V_1 \otimes \cdots \otimes V_{d-1} \\ v & \mapsto & [\![X;\, \mathrm{I}, \ldots, \mathrm{I}, v]\!] \quad = \sum_j x_{i_1,\ldots,i_{d-1},j} v_j \end{array}$$

$$\begin{array}{ccc} V_2^* \times \cdots \times V_d^* & \to & V_1 \\ (v^{(2)}, \ldots, v^{(d)}) & \mapsto & [\![X;\, \mathrm{I}, v^{(2)}, \ldots, v^{(d)}]\!] \quad = \sum_{j_2,\ldots,j_d} x_{i,j_2,\ldots,j_d} v_{j_2}^{(2)} \cdots v_{j_d}^{(d)} \end{array}$$

singular vectors

matrix $A \in V_1 \otimes V_2$.
(w, v) singular vector pair: $[\![A; w, \mathrm{I}]\!] = \sigma v$ and $[\![A; \mathrm{I}, v]\!] = \sigma w$.

tensor $X \in V_1 \otimes \cdots \otimes V_d$.
$(v^{(1)}, \dots, v^{(d)})$ singular vectors tuple: critical point[3] of

$$\begin{aligned} \text{maximize} \quad & [\![X; v^{(1)}, \dots, v^{(d)}]\!] \\ \text{subject to} \quad & \|v^{(1)}\| = \cdots = \|v^{(d)}\| = 1. \end{aligned}$$

equivalently,

$$[\![X; v^{(1)}, \dots, v^{(k-1)}, \mathrm{I}_k, v^{(k+1)}, \dots, v^{(d)}]\!] = \sigma v^{(k)}, \quad \text{for all } k = 1, \dots, d.$$

singular value $\sigma = [\![X; v^{(1)}, \dots, v^{(d)}]\!]$.

how many singular vectors?[4]

[3] Lek-Heng Lim "Singular values and eigenvalues of tensors: a variational approach." (2005)

[4] S. Friedland, G. Ottaviani. "The number of singular vector tuples and uniqueness of best rank-one approximation of tensors." (2014)

best rank one approximation

matrix $A \in V_1 \otimes V_2$. best rank one approximation: $\sigma_1 w_1 \otimes v_1$
(w_1, v_1) singular vector pair with largest singular value σ_1.

tensor $X \in V_1 \otimes \cdots \otimes V_d$. singular vector tuple $(v^{(1)}, \ldots, v^{(d)})$ with singular value σ gives rank one tensor

$$\sigma v^{(1)} \otimes \cdots \otimes v^{(d)}.$$

Theorem: best rank one approximation of X is singular vector tuple with largest singular value

proof.
minimise $\|X - \sigma v^{(1)} \otimes \cdots \otimes v^{(d)}\|^2$
i.e. maximise $\langle X, v^{(1)} \otimes \cdots \otimes v^{(d)} \rangle = [\![X; v^{(1)}, \ldots, v^{(d)}]\!]$.

[5] L. De Lathauwer, B. De-Moor, J. Vandewalle. "On the Best Rank-1 and Rank-(R1 R2… RN) Approximation of Higher-order Tensors." (2000)

[6] Liqun Qi, Ziyan Luo. Tensor analysis: spectral theory and special tensors. (2017)

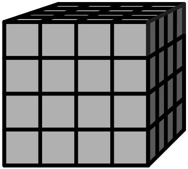
challenges for higher ranks

not true:

- best rank r approximation of tensor X is the sum of the singular vector tuples with top r singular values
- rank of tensor X is number of non-zero singular values

problems

1. *set of rank r tensors not closed.*
2. *tensor rarely expressibly as a sum of orthogonal rank one tensors*
   those that are: orthogonally decomposable (odeco)
3. *best rank r approximation not given by successive best rank one approximations*



low-rank approximation

recall: limit of rank r matrices has rank $\leq r$

limit of rank r tensors can have rank $> r$

example

$$\lim_{\epsilon\to 0}\frac{1}{2\epsilon}\left(\begin{bmatrix}1\\ \epsilon\end{bmatrix}\otimes\begin{bmatrix}1\\ \epsilon\end{bmatrix}\otimes\begin{bmatrix}1\\ \epsilon\end{bmatrix} \quad + \quad \begin{bmatrix}-1\\ \epsilon\end{bmatrix}\otimes\begin{bmatrix}-1\\ \epsilon\end{bmatrix}\otimes\begin{bmatrix}-1\\ \epsilon\end{bmatrix}\right).$$

$$= \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\left(\left[\begin{array}{cc||cc} 1 & \epsilon & \epsilon & \epsilon^2 \\ \epsilon & \epsilon^2 & \epsilon^2 & \epsilon^3 \end{array}\right] + \left[\begin{array}{cc||cc} -1 & \epsilon & \epsilon & -\epsilon^2 \\ \epsilon & -\epsilon^2 & -\epsilon^2 & \epsilon^3 \end{array}\right]\right)$$

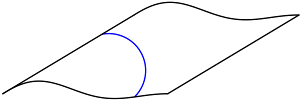
$$= \quad \left[\begin{array}{cc||cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right]$$

border rank r: limit of rank r tensors

set of border rank r tensors is an algebraic variety[7] (secant variety of Segre variety)

(given by vanishing of some polynomials)

dimension[8] and approximation[9]



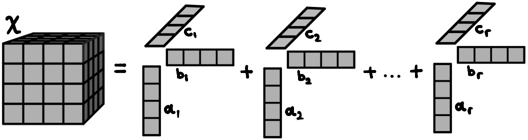
[7] JM Landsberg, "Tensors: geometry and applications." (2012).

[8] James Alexander, André Hirschowitz: Polynomial interpolation in several variables. (1995).

[9] Y. Qi, M. Michałek, L-H. Lim. "Complex tensors almost always have best low-rank approximations." (2017)

real rank

for real problems, usually require real vectors in decomposition

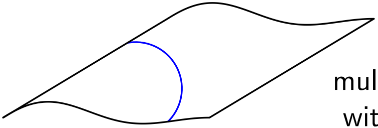


tensor has $\left\{\begin{matrix}\textit{complex}\\ \textit{real}\end{matrix}\right\}$ *rank r:* sum of r $\left\{\begin{matrix}\text{complex}\\ \text{real}\end{matrix}\right\}$ rank 1 tensors

example: complex rank 2, real rank 3

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 \end{array}\right] = \frac{1}{2}\left[\begin{array}{cc|cc} 1 & i & i & -1 \\ i & -1 & -1 & -i \end{array}\right] + \frac{1}{2}\left[\begin{array}{cc|cc} 1 & -i & -i & -1 \\ -i & -1 & -1 & i \end{array}\right]$$

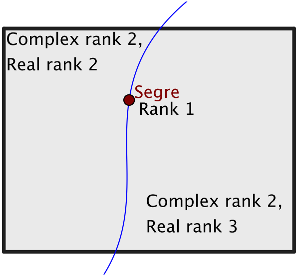
$$= \frac{1}{2}\begin{bmatrix}1\\ i\end{bmatrix} \otimes \begin{bmatrix}1\\ i\end{bmatrix} \otimes \begin{bmatrix}1\\ i\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\ -i\end{bmatrix} \otimes \begin{bmatrix}1\\ -i\end{bmatrix} \otimes \begin{bmatrix}1\\ -i\end{bmatrix}$$

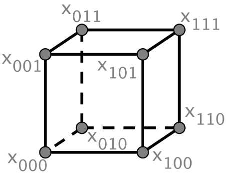


multiple real ranks occur
with positive probability

$2 \times 2 \times 2$ tensors

eight-dimensional space





real rank two $\iff$ *hyperdeterminant* $\mathbf{h}(X) \geq 0$.

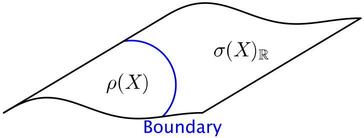
$$\begin{aligned}\mathbf{h}(X) \quad = \quad & x_{000}^2 x_{111}^2 + x_{001}^2 x_{110}^2 + x_{010}^2 x_{101}^2 + x_{011}^2 x_{100}^2 + 4x_{000}x_{011}x_{101}x_{110} + 4x_{001}x_{010}x_{100}x_{111} \\ & -2x_{000}x_{001}x_{110}x_{111} - 2x_{000}x_{010}x_{101}x_{111} - 2x_{000}x_{011}x_{100}x_{111} \\ & -2x_{001}x_{010}x_{101}x_{110} - 2x_{001}x_{011}x_{100}x_{110} - 2x_{010}x_{011}x_{100}x_{101}.\end{aligned}$$

$$\lim_{\epsilon \to 0} \frac{1}{2\epsilon}\left(\left[\begin{array}{cc|cc} 1 & \epsilon & \epsilon & \epsilon^2 \\ \epsilon & \epsilon^2 & \epsilon^2 & \epsilon^3 \end{array} \right] + \left[\begin{array}{cc|cc} -1 & \epsilon & \epsilon & -\epsilon^2 \\ \epsilon & -\epsilon^2 & -\epsilon^2 & \epsilon^3 \end{array} \right] \right).$$

stopping at ϵ: badly conditioned, no solution.

real rank approximation

set of tensors of real rank $\leq r$ is semi-algebraic set:
described by polynomial equations and inequalities.



- restrict set of tensors
- different notion of rank

- study closure of real rank r.
  best approximation is closest critical point to set and boundary.
  e.g. real rank two: secant variety and tangential variety[10].

[10] AS, Bernd Sturmfels. "Real rank two geometry." (2017)

orthogonally decomposable tensors

$X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is odeco[11] if

$$X = \sum_i \sigma_i v_i^{(1)} \otimes \cdots \otimes v_i^{(d)},$$

$v_1^{(j)}, \ldots, v_n^{(j)} \in \mathbb{R}^{n_j}$ orthonormal for every $1 \leq j \leq d$, $\sigma_i \in \mathbb{R}$.

which tensors are odeco?[12] what are singular vectors?[13]

best rank r approximation: keep the top r terms. generalisations:

- for which tensors does such a truncation property hold?[14]
- DSVD: weaker orthogonality of summands[15]
- X is a linear combination of its singular vectors[16]

---

[11] T. Zhang, G.H. Golub. Rank-one approximation to high order tensors (2001).

[12] A. Boralevi, et al. "Orthogonal and unitary tensor decomposition from an algebraic perspective." (2017).

[13] E. Robeva, AS. "Singular vectors of orthogonally decomposable tensors." (2017).

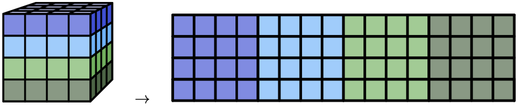
[14] N. Vannieuwenhoven, et al. "On generic nonexistence of the Schmidt–Eckart–Young decomposition for complex tensors." (2014).

[15] Derksen, Harm. "On the nuclear norm and the singular value decomposition of tensors." (2016).

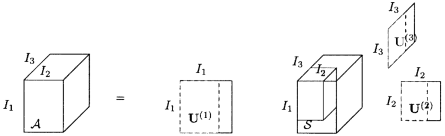
[16] J. Draisma, G. Ottaviani, A. Tocino. "Best rank-k approximations for tensors: generalizing Eckart–Young." (2018)

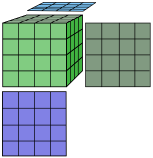
other extensions of the SVD

flattening

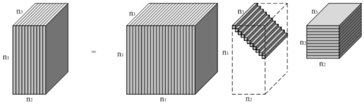


higher-order singular value decomposition[17] (special Tucker decomposition)





t-SVD (tubal rank)[18]



...tensor networks

[17] L. De Lathauwer, B. De Moor, and J. Vandewalle. "A multilinear singular value decomposition." (2000)

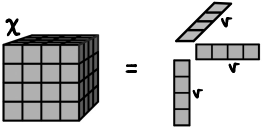
[18] Z. Zhang, S. Aeron. "Exact tensor completion using t-SVD." (2016).

symmetric tensors

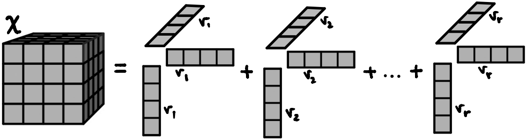
matrix A is symmetric if $A = A^\mathsf{T}$ (i.e. $a_{ij} = a_{ji}$).
tensor X symmetric if its entries are the same under permuting indices:

$$x_{ijk} = x_{ikj} = x_{jik} = x_{jki} = x_{kij} = x_{kji}.$$

e.g. moments, cumulants, partial derivatives of smooth functions.
symmetric rank 1: $X = v \otimes v \otimes v$ i.e. $x_{ijk} = v_i v_j v_k$.



symmetric rank r: sum of r tensors of symmetric rank 1.

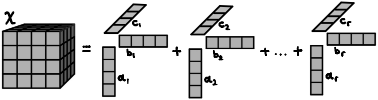


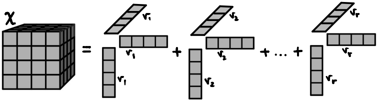
eigenvectors: critical points of $[\![X; v, \ldots, v]\!]$, $\|v\|^2 = 1$.

rank vs. symmetric rank

recall: symmetric matrix of rank r has symmetric rank r

Comon's conjecture[19]: symmetric tensor of rank r has symmetric rank r.





true in some cases
but false in general:

example: rank 903, symmetric rank $\geq$ 904 (complex)
rank 761, symmetric rank 762 (real)

---

[19] Comon P, Golub G, Lim LH, Mourrain B. Symmetric tensors and symmetric tensor rank. SIAM Journal on Matrix Analysis and Applications. 2008;30(3):1254-79.

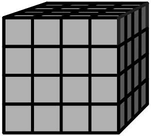
[20] Yaroslaw Shitov: A counterexample to Comon's conjecture, SIAM J. Appl. Algebra Geometry, 2 (2018) no. 3, 428–443.

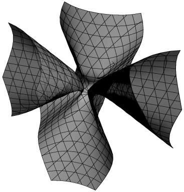
tensors and polynomials

symmetric tensors	$\leftrightarrow$	homogeneous polynomials
size $n \times \cdots \times n$ (d times)		degree d in n variables

$$T \leftrightarrow \sum_{i,j,\ldots,k=1}^{n} T_{ij\ldots k} x_i x_j \cdots x_k.$$

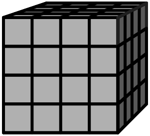
symmetric matrix M	$\leftrightarrow$	quadratic form $\mathbf{x}^T M \mathbf{x}$, $\mathbf{x} = (x_1, \ldots, x_n)$
order three sym tensor	$\leftrightarrow$	cubic polynomial
rank one tensor $\mathbf{v}^{\otimes d}$	$\leftrightarrow$	linear power $\ell^d = (v_1 x_1 + \cdots + v_n x_n)^d$
$\sum_{i=1}^r \mathbf{v_i}^{\otimes d}$	$\leftrightarrow$	Waring rank decomposition $\sum_{i=1}^r \ell_i^d$

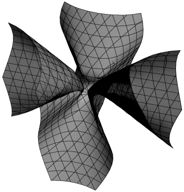




$4 \times 4 \times 4$ sym tensors $\leftrightarrow$ cubic surfaces.

classical tensor decomposition

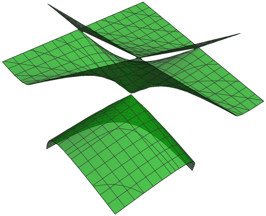




$4 \times 4 \times 4$ sym tensors $\leftrightarrow$ cubic surfaces.

Theorem (Sylvester's Pentahedral Theorem, 1851)

a generic cubic surface can be decomposed uniquely as the sum of five cubes of linear forms, $f = \ell_1^3 + \ell_2^3 + \ell_3^3 + \ell_4^3 + \ell_5^3$.



proof. $\mathrm{Hessian}(f) = \det \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}$

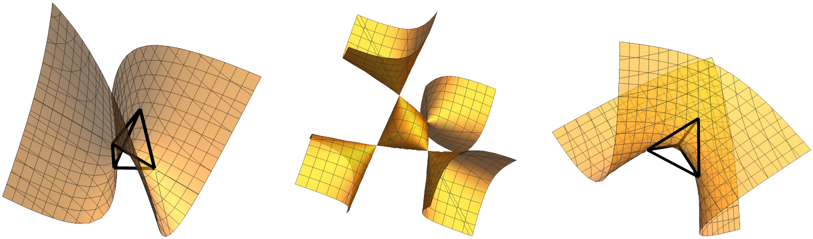
find ℓ_i from Hessian.

$\rightarrow$ 10 lines, where $\ell_i = \ell_j = 0$

$\rightarrow$ 10 singular points, $\ell_i = \ell_j = \ell_k = 0$.

outlook

- *can* extend aspects of linear algebra to tensors
- *but* need to choose most important properties for some context (orthogonality, low rank, . . .)
- encourages us to think about the structure we really need
- vast space of tensors divides into semi-algebraic subsets on which properties of interest hold



Thank you!