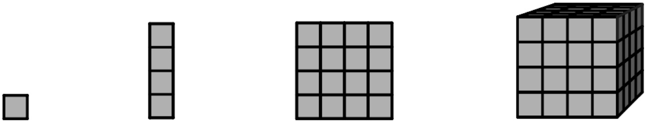
tensors in statistics and data analysis



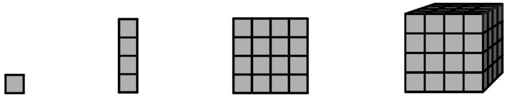
Anna Seigal

University of Oxford

17th March 2021

tutorial 1

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

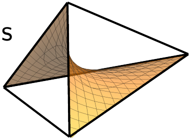


- aspects of the singular value decomposition
  - singular vectors
  - notions of rank
  - low-rank approximation
  - analogues for symmetric tensors

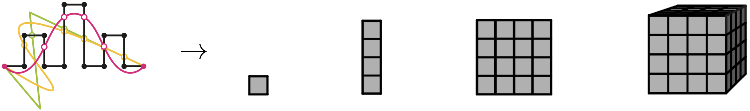
today

tensors in applications:

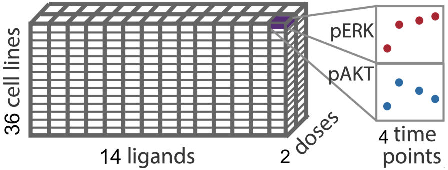
1. probabilities and statistical models



2. encodings (e.g. moments, signature)



3. data values



focus: tensor decompositions that arise, their algebraic properties

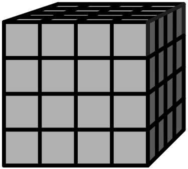
[1] AS, M. Beguerisse-Díaz, B. Schoeberl, M. Niepel, H. Harrington: *Tensor clustering with algebraic constraints gives interpretable groups of crosstalk mechanisms in breast cancer*, (2019).

probability tensors

random variable Y on $\{1, \dots, n\}$
probability mass function is vector

$$x_i = \mathbb{P}(Y = i).$$

random variables Y_j on $\{1, \dots, n_j\}$, $\;1 \le j \le d$.
joint probability mass function is $n_1 \times \cdots \times n_d$ tensor



$$x_{i_1 \dots i_d} = \mathbb{P}(Y_1 = i_1, \dots, Y_d = i_d).$$

properties of tensor $\leftrightarrow$ properties of distribution
family of tensors $\leftrightarrow$ statistical model

rank one

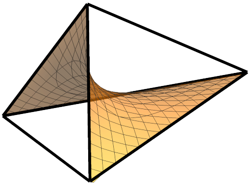
random variables Y_j on $\{1, \ldots, n_j\}$
probability mass function is $n_1 \times \cdots \times n_d$ tensor

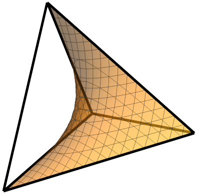
$$x_{i_1 \ldots i_d} = \mathbb{P}(Y_1 = i_1\,, \ldots, Y_d = i_d)$$

$Y_1, \ldots, Y_d$ independent means

$$\begin{aligned} x_{i_1 \ldots i_d} &= \mathbb{P}(Y_1 = i_1) \cdots \mathbb{P}(Y_d = i_d) \\ &= \quad x^{(1)}_{i_1} \quad \cdots \quad x^{(d)}_{i_d} \end{aligned}$$

(non-negative) rank one decomposition

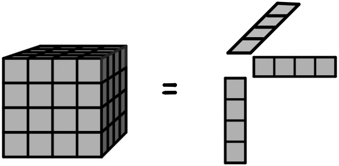




family of tensors $\leftrightarrow$ statistical model
rank one $\leftrightarrow$ independence model

beyond independence

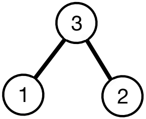
- rank one $\leftrightarrow$ independence model



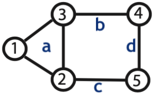
- conditional independence
  $(Y_1 \perp\!\!\!\perp Y_2) | Y_3$ means "Y_1 independent of Y_2 given Y_3"

$$x_{i_1 i_2 i_3} = \mathbb{P}(Y_1 = i_1, Y_2 = i_2, Y_3 = i_3)$$
$$= a_{i_1 i_3} b_{i_2 i_3}$$

slices rank one (for fixed i_3)

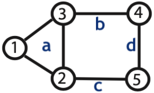


- more general graphs



$a_{i_1 i_2 i_3} b_{i_3 i_4} c_{i_2 i_5} d_{i_4 i_5}$

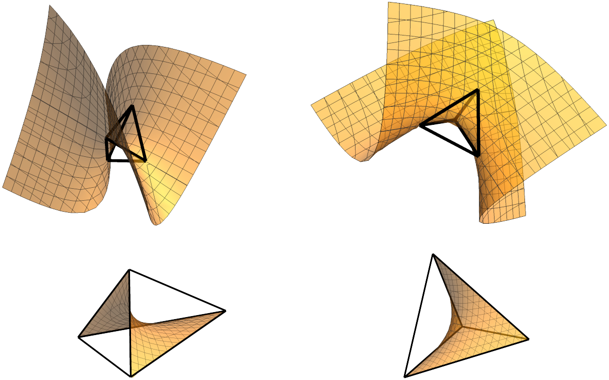
conditional independence equations



$$x_{i_1 \dots i_5} = a_{i_1 i_2 i_3} b_{i_3 i_4} c_{i_2 i_5} d_{i_4 i_5}$$

conditional independence $\leftrightarrow$ rank one conditions on various slices

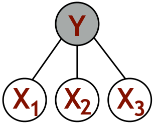
...conditional independence ideals[2]



[2] Seth Sullivant, Algebraic statistics. Vol. 194. American Mathematical Soc., 2018.

hidden variables

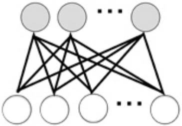
X_i on $\{1, \ldots, n_i\}$
Y on $\{1, \ldots, r\}$



$$x_{i_1 i_2 i_3} = \sum_{j=1}^{r} a_{i_1 j} b_{i_2 j} c_{i_3 j}$$

non-negative rank $\leq r$ $\quad\leftrightarrow\quad$ mixture model, r state hidden variable.

other graphs e.g. restricted Boltzmann machines[3]:



Hadamard product of non-negative rank r tensors
e.g.

$$x_{i_1 i_2 i_3} = \left(\sum_{j=1}^{r} a_{i_1 j} b_{i_2 j} c_{i_3 j} \right) \left(\sum_{k=1}^{r} u_{i_1 k} v_{i_2 k} w_{i_3 k} \right)$$

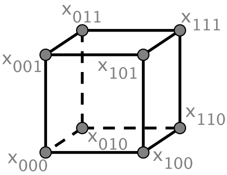
[3] G. Montúfar "Restricted Boltzmann machines: Introduction and review." (2018).

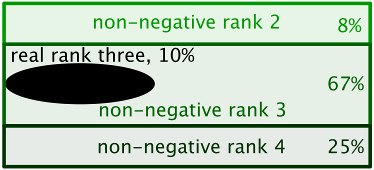
equations and inequalities

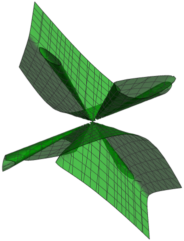
set of tensors in model give semi-algebraic set

what are equations and inequalities? which non-negative ranks occur?

e.g. $2 \times 2 \times 2$ tensors with $x_{ijk} > 0$







non-negative rank 2: four pieces[4], one is:

$$\left\{\begin{array}{ll} x_{000}x_{011} \geq x_{010}x_{001}, & x_{100}x_{111} \geq x_{110}x_{101}, \\ x_{000}x_{101} \geq x_{100}x_{001}, & x_{010}x_{111} \geq x_{110}x_{011}, \\ x_{000}x_{110} \geq x_{100}x_{010}, & x_{001}x_{111} \geq x_{101}x_{011}. \end{array}\right\}$$

non-negative rank ≤ 3: three pieces[5] , one is:

$$(x_{000}x_{011} - x_{001}x_{010})(x_{100}x_{111} - x_{101}x_{110}) \geq 0$$

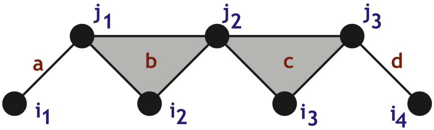
---

[4] E Allman, J Rhodes, B Sturmfels, P Zwiernik. Tensors of nonnegative rank two (2015).

[5] AS, G Montúfar "Mixtures and products in two graphical models" (2018).

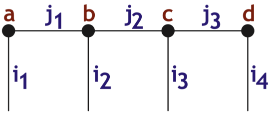
connection to tensor networks

graphical model:



$$x_{i_1 i_2 i_3 i_4} = \sum_{j_1, j_2, j_3} \mathbf{a}_{i_1 j_1} \mathbf{b}_{j_1 i_2 j_2} \mathbf{c}_{j_2 i_3 j_3} \mathbf{d}_{j_3 i_4}$$

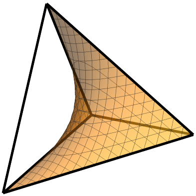
tensor network:

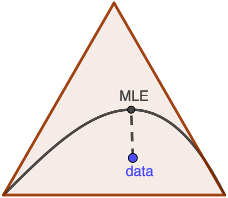


[6] E Robeva, AS "Duality of graphical models and tensor networks" (2019).

estimating parameters

maximum likelihood estimate: point(s) on model most likely to give data





data: $u_{i_1 \dots i_d}$ = fraction of times $\mathbf{i} = (i_1, \dots, i_d)$ occurs.

likelihood: $L(p) = \prod_{\mathbf{i}} p_{\mathbf{i}}^{u_{\mathbf{i}}}$

or log-likelihood: $\ell(p) = \sum_{\mathbf{i}} u_{\mathbf{i}} \log p_{\mathbf{i}}$

maximize ℓ over the model

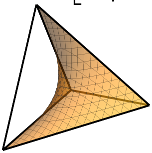
(minimize Kullback-Leibler divergence $KL(p \| u)$)

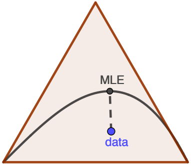
[7] LH Lim, P Comon "Nonnegative approximations of nonnegative tensors" (2009)

'best' rank one approximation

independence model $\leftrightarrow$ rank one

two binary random variables $\begin{bmatrix} \lambda_0\mu_0 & \lambda_0\mu_1 \\ \lambda_1\mu_0 & \lambda_1\mu_1 \end{bmatrix}$, $\lambda_0 + \lambda_1 = 1$, $\mu_0 + \mu_1 = 1$





data: $\begin{bmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{bmatrix}$, u_{ij} = fraction of times (i, j) occurs.

likelihood: $(\lambda_0\mu_0)^{u_{00}}(\lambda_0\mu_1)^{u_{01}}(\lambda_1\mu_0)^{u_{10}}(\lambda_1\mu_1)^{u_{11}} = \lambda_0^{u_{0+}}\lambda_1^{u_{1+}}\mu_0^{u_{+0}}\mu_1^{u_{+1}}$

maximum likelihood estimate

$$\begin{array}{ll} \hat{\lambda}_0 = u_{0+} & \hat{\mu}_0 = u_{+0} \\ \hat{\lambda}_1 = u_{1+} & \hat{\mu}_1 = u_{+1} \end{array} \qquad \begin{bmatrix} u_{0+}u_{+0} & u_{1+}u_{+0} \\ u_{1+}u_{+0} & u_{1+}u_{+1} \end{bmatrix}$$

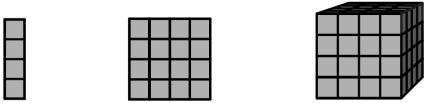
...hidden variables[8]

[8] E Allman et al. "Maximum likelihood estimation of latent class model through model boundary decomposition" (2017).

moments and cumulants

random vector X taking values in $\mathbb{R}^d$
e.g. d-dimensional Gaussian, $X \sim N(\mu, \Sigma)$

moment tensors $\mathbb{E}[X \otimes \cdots \otimes X]$ (or cumulant tensors)



$\ldots$

e.g. moments: $(\mu_i,\ \mu_i\mu_j + \sigma_{ij},\ \mu_i\mu_j\mu_k + \mu_i\sigma_{jk} + \mu_j\sigma_{ik} + \mu_k\sigma_{ij}, \ldots)$
cumulants: $(\mu_i,\ \sigma_{ij},\ 0,\ \ldots,)$

properties of tensors $\leftrightarrow$ **properties of distribution**

- tensor decomposition $\leftrightarrow$ estimate parameters[9]
- dimension of tensors $\leftrightarrow$ identifiability of model[10]
- equations in tensor entries $\leftrightarrow$ model membership[11]

[9] A. Anandkumar et al. "Tensor decompositions for learning latent variable models." (2012).
[10] C. Améndola, K. Ranestad, B. Sturmfels. "Algebraic identifiability of Gaussian mixtures." (2018).
[11] E. Robeva, J.B. Seby. "Multi-trek separation in Linear Structural Equation Models." (2020).

the signature

a path is a function $\psi : [0,1] \to \mathbb{R}^d$, $\quad t \mapsto \Big(\psi_1(t), \psi_2(t), \dots, \psi_d(t)\Big)$

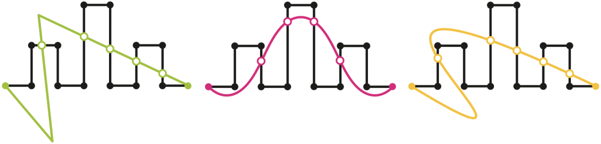
the signature[12] is a sequence of tensors

$d, \qquad d^2, \qquad d^3, \qquad \cdots$

...whose entries are iterated integrals. e.g.

$$c_i = \int_0^1 \mathrm{d}\psi_i(t) = \psi_i(1) - \psi_i(0), \qquad c_{ijk} = \int_0^1 \left(\int_0^v \left(\int_0^u \mathrm{d}\psi_i(t) \right) \mathrm{d}\psi_j(u) \right) \mathrm{d}\psi_k(v)$$

properties of tensors $\leftrightarrow$ **properties of path**

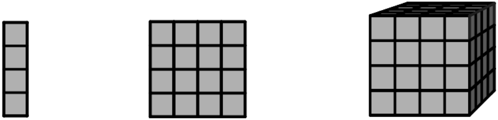


[12] KT Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula (1957)

signatures of paths

linear path $\psi(t) = tv$ has signature

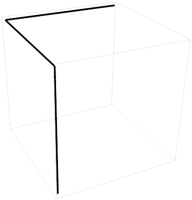
v, $\quad v \otimes v$, $\quad v \otimes v \otimes v$, $\quad \cdots$

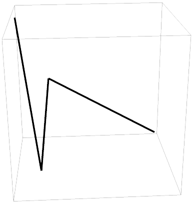


$\cdots$

$v \in V$, can write signature as $\exp(v) \in T(V) = \bigoplus_{d \geq 0} V^{\otimes d}$

piecewise linear path with pieces $v_1, \ldots, v_n$





has signature

$$\exp(v_1) \otimes \cdots \otimes \exp(v_n) = (1 + v_1 + \frac{1}{2} v_1^{\otimes 2} + \cdots) \otimes \cdots \otimes (1 + v_n + \frac{1}{2} v_n^{\otimes 2} + \cdots)$$

tensor product of signatures $\leftrightarrow$ **concatenation of paths**

tensor decomposition of signatures

piecewise linear path with pieces $v_1, \ldots, v_n \in \mathbb{R}^d$ has signature

$$\exp(v_1) \otimes \cdots \otimes \exp(v_n)$$

exp-rank one decomposition fits piecewise linear path to signature

i.e. given signature $S \in T(\mathbb{R}^d)$, seek $v_1, \ldots, v_n \in \mathbb{R}^d$ such that

$$S \approx \exp(v_1) \otimes \cdots \otimes \exp(v_n).$$

- recovery from infinite signature[13]
- finite truncation of signature, connection to statistical moments[14]
- third order signature[15]

---

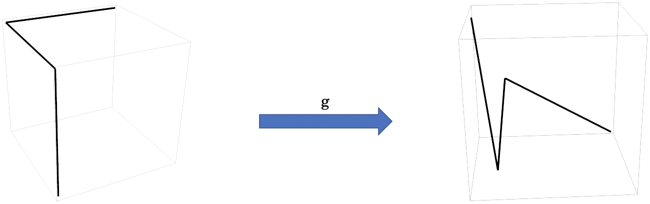
[13] T Lyons, W Xu "Inverting the signature of a path" (2014).

[14] C Améndola, P Friz, B Sturmfels. "Varieties of signature tensors." (2019).

[15] M Pfeffer, AS, B Sturmfels. "Learning paths from signature tensors." (2019).

equivariance of the signature

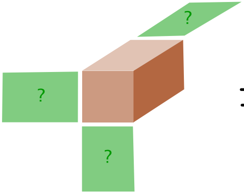
linear transformation of path



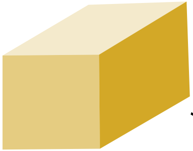
transforms signature under congruence

$$v \mapsto gv, \qquad M \mapsto gMg^{\mathsf{T}}, \qquad T \mapsto [\![T; g]\!] = \sum_{\alpha\beta\gamma} t_{\alpha\beta\gamma} g_{i\alpha} g_{j\beta} g_{k\gamma}.$$

$[\![T; g, g, g]\!]$



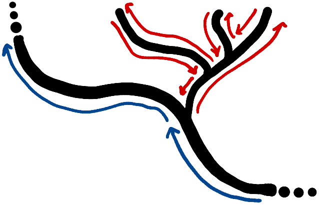
=



S

properties and applications

- signature invariant under (exactly) re-parametrisation and tree like equivalence[16] [17]



- applications:
  machine learning pipelines[18] [19] biological time-series data[20] [21]

[16] KT Chen: Integration of paths–A faithful representation of paths by noncommutative formal power series, Transactions of the AMS 89.2 (1958): 395-407.

[17] Ben Hambly, Terry Lyons. "Uniqueness for the signature of a path of bounded variation and the reduced path group." Annals of Mathematics (2010): 109-167.

[18] I Chevyrev, A Kormilitzin. A primer on the signature method in machine learning. arXiv preprint arXiv:1603.03788. 2016 Mar 11.

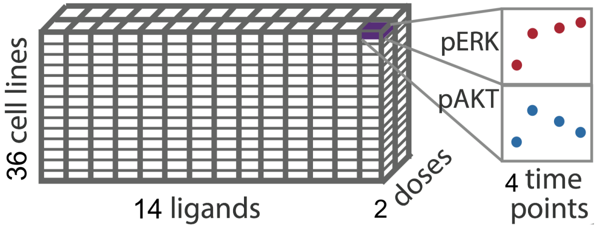
[19] P Bonnier, et al. "Deep signatures." (2019) .

[20] I. P. Arribas, et al. "A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder." (2018).

[21] J. Morrill, et al. "The signature-based model for early detection of sepsis from electronic health records in the intensive care unit." (2019).

tensors of biological data

each entry of tensor is a data value
indices can be ordered (e.g. spatial, temporal) or unordered
e.g. signaling pathways in cancer cells[22]



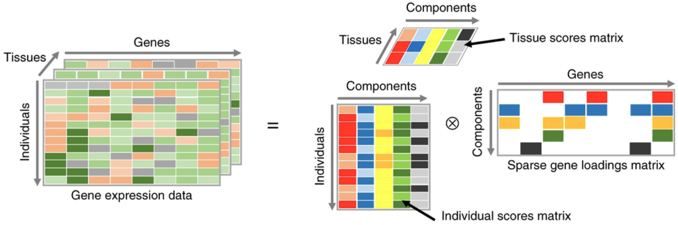
how are the 5 variables related?
if we flatten to matrix, what do we lose?

- cell lines $\times$ everything else $= 36 \times 224$
- cell lines & ligands $\times$ everything else $= 504 \times 16$

[22] AS, M. Beguerisse-Díaz, B. Schoeberl, M. Niepel, H. Harrington: *Tensor clustering with algebraic constraints gives interpretable groups of crosstalk mechanisms in breast cancer*, (2019).

tensor decompositions of biological data

- low rank CP decomposition[23]
- high rank CP with sparsity[24]



- higher order singular value decomposition[25]
- semi non-negative Tucker decomposition"[26]
- non-negative Tucker decomposition[27]

[23] A Williams, et al. "Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis" (2018)

[24] V. Hore, et al. "Tensor decomposition for multiple-tissue gene expression experiments" (2016)

[25] L Omberg, G H Golub, O Alter. "A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies" (2007)

[26] M Wang, J Fischer, Y S Song, "Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition (2019)

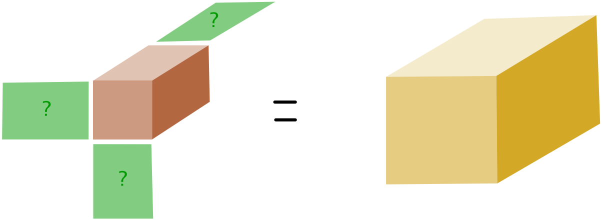
[27] C M Schürch, S S Bhate et al. "Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front" (2020)

outlook

- connection between tensor structure and application

$$\textbf{property of tensors} \leftrightarrow \textbf{property of} \begin{bmatrix} \textbf{probability distribution} \\ \textbf{path} \\ \textbf{biological setting?} \end{bmatrix}$$

- new tensor decompositions from applications
- new applications of tensor decompositions



Thank you!