Handout for the mini-course
“Tensor models & applications”

J. Henrique de M. Goulart
IRIT / Toulouse INP
henrique.goulart@irit.fr

June 2024

Contents

1 Important properties & identities

1.1 Fundamental results from linear algebra

Definition 1. *The* ***kernel*** *of a matrix* $\mathbf{A} \in \mathbb{R}^{M\times N}$ *is a subspacce of* $\mathbb{R}^N$ *defined as*

$$\ker \mathbf{A} := \left\{ \boldsymbol{v} \in \mathbb{R}^N \,:\, \mathbf{A}\boldsymbol{v} = \mathbf{0} \right\}.$$

Its dimension $\dim\ker\mathbf{A}$ *is called the* ***nullity*** *of* $\mathbf{A}$.

Definition 2. *The rank of* $\mathbf{A} \in \mathbb{R}^{M\times N}$*, denoted* $\operatorname{rank}\mathbf{A}$*, is defined as the dimension of the image of* $\mathbb{R}^N$ *under the map* $\boldsymbol{v} \mapsto \mathbf{A}\boldsymbol{v}$*. In other words, it equals the dimension of the subspace spanned by the columns of* $\mathbf{A}$*, defined as* $\operatorname{colspan}\mathbf{A} := \left\{ \mathbf{A}\boldsymbol{v} \,:\, \boldsymbol{v} \in \mathbb{R}^N \right\}$.

Theorem 1 (Rank-nuillity)**.** *Let* $\mathbf{A} \in \mathbb{R}^{M\times N}$*. Then,*

$$\operatorname{rank}\mathbf{A} + \dim\ker\mathbf{A} = N.$$

Corollary 1. *Let* $\mathbf{A} \in \mathbb{R}^{M\times N}$ *and denote* $\operatorname{rowspan}\mathbf{A} := \left\{ \mathbf{A}^\mathsf{T}\boldsymbol{v} \,:\, \boldsymbol{v} \in \mathbb{R}^M \right\}$ *the subspace of* $\mathbb{R}^N$ *spanned by the rows of* $\mathbf{A}$*. Then,* $\dim\operatorname{colspan}\mathbf{A} = \dim\operatorname{rowspan}\mathbf{B}$ *and* $\operatorname{rank}\mathbf{A} \le \min\{M, N\}$.

Theorem 2 (Sylvester’s rank inequality)**.** *Let* $\mathbf{A} \in \mathbb{R}^{M\times R}$ *and* $\mathbf{B} \in \mathbb{R}^{R\times N}$*. Then,*

$$\operatorname{rank}\mathbf{AB} \ge \operatorname{rank}\mathbf{A} + \operatorname{rank}\mathbf{B} - R.$$

Proof. Follows from $\dim\ker\mathbf{AB} \leq \dim\ker\mathbf{A} + \dim\ker\mathbf{B}$ and the rank-nullity theorem. $\square$

Proposition 1. *Let $\mathbf{A} \in \mathbb{R}^{M\times R}$ and $\mathbf{B} \in \mathbb{R}^{R\times N}$. Then,*

$$\operatorname{rank}\mathbf{AB} \leq \min\{\operatorname{rank}\mathbf{A}, \operatorname{rank}\mathbf{B}\}.$$

Corollary 2. *Let $\mathbf{A} \in \mathbb{R}^{M\times R}$ and $\mathbf{B} \in \mathbb{R}^{R\times N}$, and suppose $\mathbf{A}$ has full column rank (that is, $\operatorname{rank}\mathbf{A} = R$). Then,*

$$\operatorname{rank}\mathbf{AB} = \operatorname{rank}\mathbf{B}$$

Similarly, if $\mathbf{B}$ has full row rank, then $\operatorname{rank}\mathbf{AB} = \operatorname{rank}\mathbf{A}$.

Theorem 3 (Singular value decomposition)**.** *Every $\mathbf{A} \in \mathbb{R}^{M\times N}$ admits a decomposition*

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

such that

- ***left singular vectors:*** *$\mathbf{U} \in \mathbb{R}^{M\times M}$ is an orthogonal matrix containing the eigenvectors of $\mathbf{AA}^\mathsf{T}$;*
- ***right singular vectors:*** *$\mathbf{V} \in \mathbb{R}^{N\times N}$ is an orthogonal matrix containing the eigenvectors of $\mathbf{A}^\mathsf{T}\mathbf{A}$;*
- ***singular values:*** *$\boldsymbol{\Sigma} \in \mathbb{R}^{M\times N}$ is a rectangular diagonal matrix given by*

$$\boldsymbol{\Sigma} = \begin{cases} \operatorname{Diag}(\sigma_1,\dots,\sigma_M), & M = N, \\ \begin{pmatrix} \operatorname{Diag}(\sigma_1,\dots,\sigma_M) & \mathbf{0}_{M,N-M} \end{pmatrix}, & M < N, \\ \begin{pmatrix} \operatorname{Diag}(\sigma_1,\dots,\sigma_N) \\ \mathbf{0}_{M-N,N} \end{pmatrix}, & N < M, \end{cases}$$

  with $\sigma_i = \sqrt{\lambda_i(\mathbf{AA}^\mathsf{T})} = \sqrt{\lambda_i(\mathbf{A}^\mathsf{T}\mathbf{A})}$ ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{M,N\}} \geq 0$.

Furthermore, the number of nonzero singular values of $\mathbf{A}$ equals the rank of $\mathbf{A}$, which can thus be written as

$$\mathbf{A} = \sum_{r=1}^{\operatorname{rank}\mathbf{A}} \sigma_r\, \boldsymbol{u}_r \boldsymbol{v}_r^\mathsf{T}.$$

1.2 Kronecker & Khatri-Rao products

Definition 3. *The **Kronecker product** of two matrices $\mathbf{A} \in \mathbb{R}^{M\times N}$ and $\mathbf{B} \in \mathbb{R}^{L\times P}$ yields a matrix $\mathbf{A} \boxtimes \mathbf{B} \in \mathbb{R}^{ML\times NP}$ given by*

$$\mathbf{A} \boxtimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \dots & a_{MN}\mathbf{B} \end{pmatrix}.$$

.

Definition 4. *The vectorization of a matrix* $\mathbf{A} = \begin{pmatrix} \boldsymbol{a}_1 & \dots & \boldsymbol{a}_N \end{pmatrix} \in \mathbb{R}^{M\times N}$, *dentoted* $\operatorname{vec}\mathbf{A} \in \mathbb{R}^{NM}$, *yields a vector obtained by stacking the columns of* $\mathbf{A}$ *as:*

$$\operatorname{vec}\mathbf{A} = \begin{pmatrix} \boldsymbol{a}_1^\mathsf{T} & \dots & \boldsymbol{a}_N^\mathsf{T} \end{pmatrix}^\mathsf{T}.$$

Proposition 2. *The Kronecker product satisfies:*

- $\mathbf{A} \boxtimes \mathbf{B} \neq \mathbf{B} \boxtimes \mathbf{A}$ *in general;*
- $\mathbf{A} \boxtimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \boxtimes \mathbf{B} + \mathbf{A} \boxtimes \mathbf{C}$ *and* $(\mathbf{A} + \mathbf{B}) \boxtimes \mathbf{C} = \mathbf{A} \boxtimes \mathbf{C} + \mathbf{B} \boxtimes \mathbf{C}$*;*
- $\mathbf{A} \boxtimes (\mathbf{B} \boxtimes \mathbf{C}) = (\mathbf{A} \boxtimes \mathbf{B}) \boxtimes \mathbf{C}$*;*
- $(\mathbf{A} \boxtimes \mathbf{B})^\mathsf{T} = \mathbf{A}^\mathsf{T} \boxtimes \mathbf{B}^\mathsf{T}$*;*
- $(\mathbf{A} \boxtimes \mathbf{B})(\mathbf{C} \boxtimes \mathbf{D}) = (\mathbf{AC}) \boxtimes (\mathbf{BD})$ *(for matrices of compatible dimensions);*
- $\operatorname{vec}(\mathbf{ABC}^\mathsf{T}) = (\mathbf{C} \boxtimes \mathbf{A}) \operatorname{vec}\mathbf{B}$*;*
- $\operatorname{rank}(\mathbf{A} \boxtimes \mathbf{B}) = \operatorname{rank}\mathbf{A} \operatorname{rank}\mathbf{B}$.

If, in particular, $\mathbf{A} \in \mathbb{R}^{M\times M}$ *and* $\mathbf{B} \in \mathbb{R}^{N\times N}$*, then*

- $(\mathbf{A} \boxtimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \boxtimes \mathbf{B}^{-1}$ *if the inverses exist;*
- $\det(\mathbf{A} \boxtimes \mathbf{B}) = (\det\mathbf{A})^N (\det\mathbf{B})^M$.

Definition 5. *The* ***Khatri-Rao*** *or* ***columnwise Kronecker product*** *of two matrices* $\mathbf{A} \in \mathbb{R}^{M\times R}$ *and* $\mathbf{B} \in \mathbb{R}^{N\times R}$ *yields a matrix* $\mathbf{A} \odot \mathbf{B} \in \mathbb{R}^{MN\times R}$ *given by*

$$\mathbf{A} \odot \mathbf{B} = \begin{pmatrix} \boldsymbol{a}_1 \boxtimes \boldsymbol{b}_1 & \dots & \boldsymbol{a}_R \boxtimes \boldsymbol{b}_R \end{pmatrix}.$$

Proposition 3. *The Khatri-Rao product satisfies:*

- $\mathbf{A} \odot \mathbf{B} \neq \mathbf{B} \odot \mathbf{A}$ *in general;*
- $\mathbf{A} \odot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \odot \mathbf{B} + \mathbf{A} \odot \mathbf{C}$ *and* $(\mathbf{A} + \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot \mathbf{C} + \mathbf{B} \odot \mathbf{C}$*;*
- $\mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C}$*;*
- $(\mathbf{A} \odot \mathbf{B})^\mathsf{T}(\mathbf{A} \odot \mathbf{B}) = (\mathbf{A}^\mathsf{T}\mathbf{A}) \circledast (\mathbf{B}^\mathsf{T}\mathbf{B})$, *where* $\circledast$ *denotes the entry-wise (Hadamard) product;*
- $\mathbf{A} \odot \mathbf{B} = (\mathbf{A} \boxtimes \mathbf{B}) \begin{pmatrix} \boldsymbol{e}_1 \boxtimes \boldsymbol{e}_1 & \dots & \boldsymbol{e}_R \boxtimes \boldsymbol{e}_R \end{pmatrix}$*;*
- $\operatorname{vec}(\mathbf{AB}^\mathsf{T}) = (\mathbf{B} \boxtimes \mathbf{A}) \operatorname{vec}\mathbf{I}_R = (\mathbf{B} \odot \mathbf{A})\mathbf{1}_R$, *where R is the number of columns of* $\mathbf{A}$ *and of* $\mathbf{B}$*, and* $\mathbf{1}_R$ *is the R-dimensional vector whose every enty equals 1.*

2 Exercises

1. Show that $(\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} = (\mathbf{U}\boldsymbol{a}) \otimes (\mathbf{V}\boldsymbol{b}) \otimes (\mathbf{W}\boldsymbol{c})$.

2. The mode-1 unfolding of $\boldsymbol{\mathsf{X}} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, denoted $\mathbf{X}_{(1)} \in \mathbb{R}^{N_1 \times N_3 N_2}$, is defined so that

$$(\mathbf{X}_{(1)})_{i,\ell(j,k)} = x_{ijk},$$

with $\ell(j,k) := N_2(k-1) + j$. Similarly,

$$(\mathbf{X}_{(2)})_{j,p(i,k)} = x_{ijk},$$

with $p(i,k) := N_1(k-1) + i$ and

$$(\mathbf{X}_{(3)})_{k,q(i,j)} = x_{ijk},$$

with $q(i,j) := N_1(j-1) + i$.

   (a) Find generalized formulas for the unfoldings of tensors having any order and dimensions, that is, for $\boldsymbol{\mathsf{X}} \in \mathbb{R}^{N_1 \times \cdots \times N_d}$, $d > 2$.

   (b) Show that, if $\boldsymbol{\mathsf{X}} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$, then with the above definitions it follows that

$$\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^{\mathsf{T}},\ \mathbf{X}_{(2)} = \mathbf{B}(\mathbf{C} \odot \mathbf{A})^{\mathsf{T}},\ \mathbf{X}_{(3)} = \mathbf{C}(\mathbf{B} \odot \mathbf{A})^{\mathsf{T}}.$$

   (c) Find a general expression for $\mathbf{X}_{(i)}$ with $\boldsymbol{\mathsf{X}} = [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(d)}]\!]$.

   (d) Now, letting $\mathbf{X} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \boldsymbol{\mathsf{S}}$, show that

$$\mathbf{X}_{(1)} = \mathbf{U}\mathbf{S}_{(1)}(\mathbf{W} \boxtimes \mathbf{V})^{\mathsf{T}}.$$

3. Recall that $\operatorname{vec} \boldsymbol{\mathsf{X}} := \operatorname{vec} \mathbf{X}_{(1)}$. Show that

   (a) If $\mathbf{X} = (\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \boldsymbol{\mathsf{S}}$, then

$$\operatorname{vec} \boldsymbol{\mathsf{X}} = (\mathbf{W} \boxtimes \mathbf{V} \boxtimes \mathbf{U}) \operatorname{vec} \boldsymbol{\mathsf{S}}.$$

   (b) If $\mathbf{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$, then

$$\operatorname{vec} \boldsymbol{\mathsf{X}} = (\mathbf{C} \odot \mathbf{B} \odot \mathbf{A})\mathbf{1}_R.$$

   (c) Generalize the above formulas for any order d.

4. Show that, if $\mathbf{U}, \mathbf{V}, \mathbf{W}$ have full column rank, then

$$\operatorname{mrank}(\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \boldsymbol{\mathsf{X}} = \operatorname{mrank} \boldsymbol{\mathsf{X}}$$

and

$$\operatorname{rank}(\mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \boldsymbol{\mathsf{X}} = \operatorname{rank} \boldsymbol{\mathsf{X}}.$$

5. Show that $(\mathbf{A} \odot \mathbf{B})^{\mathsf{T}}(\mathbf{A} \odot \mathbf{B}) = (\mathbf{A}^{\mathsf{T}}\mathbf{A}) \circledast (\mathbf{B}^{\mathsf{T}}\mathbf{B})$.

6. Show that the family $\{\boldsymbol{u}_{r_1} \otimes \boldsymbol{v}_{r_2} \otimes \boldsymbol{w}_{r_3}\}_{r_1 \in [R_1], r_2 \in [R_2], r_3 \in [R_3]}$ is linearly independent iff the families $\{\boldsymbol{u}_{r_1}\}_{r_1 \in [R_1]}$, $\{\boldsymbol{v}_{r_2}\}_{r_2 \in [R_2]}$ and $\{\boldsymbol{w}_{r_3}\}_{r_3 \in [R_3]}$ are linearly independent, where $[R] := \{1, \ldots, R\}$.

7. Show that $\boldsymbol{a} \otimes \boldsymbol{b} \otimes \boldsymbol{c} + \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w}$ has rank one iff these two terms are collinear over at least two modes, that is, there exist $\alpha, \beta \in \mathbb{R}$ such that one of the following items is true:

    - $\boldsymbol{a} = \alpha \boldsymbol{u}$ *and* $\boldsymbol{b} = \beta \boldsymbol{v}$;
    - $\boldsymbol{a} = \alpha \boldsymbol{u}$ *and* $\boldsymbol{c} = \beta \boldsymbol{w}$;
    - $\boldsymbol{b} = \alpha \boldsymbol{v}$ *and* $\boldsymbol{c} = \beta \boldsymbol{w}$.

8. Show that if $\mathbf{D} = \sum_{r=1}^{R} \boldsymbol{e}_r^{\otimes 3}$ is essentially unique, then any PD $\mathbf{X} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]_R$ with full-column-rank factors $\mathbf{A}, \mathbf{B}, \mathbf{C}$ is also essentially unique.

9. Show that $\frac{\partial}{\partial \boldsymbol{u}} \mathbf{X} \cdot \boldsymbol{u}^3 = 3\, \mathbf{X} \cdot \boldsymbol{u}^2$.

3 References, organized per topic

Books [Hackbusch, 2012], [Favier, 2021]

Applications [Hsu and Kakade, 2013],[Anandksumar et al., 2014], [Vasilescu and Terzopoulos, 2002], [Vasilescu and Terzopoulos, 2002], [Lebedev et al., 2015], [Hong et al., 2020], [Favier et al., 2012], [Goulart and Burt, 2021], [Novikov et al., 2021], [Dolgov et al., 2020]

Block-term decomposition [De Lathauwer, 2008], [De Lathauwer, 2011], [Goulart and Comon, 2019], [Rontogiannis et al., 2021], [Giampouras et al., 2022], [Domanov and Lathauwer, 2020], [Khamidullina et al., 2023], [Goulart et al., 2020], [Sorber et al., 2013]

Tensor networks [Cichocki et al., 2016], [Cichocki et al., 2017]

Multilinear rank and Tucker model [Hitchcock, 1928], [Tucker, 1966]

High-order singular value decomposition & variants [De Lathauwer et al., 2000a], [Vannieuwenhoven et al., 2012], [Goulart and Comon, 2017], [da Silva et al., 2016]

Best rank-one approximation and tensor eigenpairs [Kofidis and Regalia, 2002], [Kolda and Mayo, 2011],[Lim, 2005], [Cartwright and Sturmfels, 2013], [Huang et al., 2022]

Best low-mrank approximation [De Lathauwer et al., 2000b], [Xu, 2018], [Lebeau et al., 2024a]

Polyadic decomposition [Hitchcock, 1927], [Carroll and Chang, 1970], [Harshmans, 1970], [Rajih et al., 2008], [Uschmajew, 2012], [Yang, 2023], [Sanchez and Kowalski, 1990], [Leurgans et al., 1993], [Beltrán et al., 2019], [Bhaskara et al., 2014], [Cattell, 1944], [Vannieuwenhoven, 2015], [Sorber et al., 2013], [Domanov and Lathauwer, 2014], [Kolda and Hong, 2020], [Yang et al., 2018], [Hopkins et al., 2019]

CPD uniqueness [Kruskal, 1977]

Existence and conditioning of best low-rank approximation [de Silva and Lim, 2008], [Qi et al., 2020], [Evert and De Lathauwer, 2022], [Paatero, 2000], [Breiding and Vannieuwenhoven, 2018]

Orthogonal tensor decompositions [Kolda, 2001], [Robeva, 2016]

Tensor PCA, spiked models and related topics [Ben Arous et al., 2019], [Montanari and Richard, 2014], [Ros et al., 2019], [Goulart et al., 2022], [Seddik et al., 2024], [Ben Arous et al., 2022], [Ben Arous et al., 2023], [Perry et al., 2020], [Montanari et al., 2017], [**?**], [Lebeau et al., 2024b]

Tensor train decomposition [Oseledets, 2011]

Generic rank [Abo et al., 2009], [Comon et al., 2009], [Comon and ten Berge, 2008]

Maximal rank [Sumi et al., 2010], [Bremner and Hu, 2013], [Kruskal, 1989]

References

[Abo et al., 2009] Abo, H., Ottaviani, G., and Peterson, C. (2009). Induction for secant varieties of segre varieties. *Transactions of the American Mathematical Society*, 361(2):767–792.

[Anandksumar et al., 2014] Anandksumar, A., Ge, R., Hsu, D. J., Kakade, S. M., Telgarsky, M., et al. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.*, 15(1):2773–2832.

[Beltrán et al., 2019] Beltrán, C., Breiding, P., and Vannieuwenhoven, N. (2019). Pencil-based algorithms for tensor rank decomposition are not stable. 40(2):739–773.

[Ben Arous et al., 2022] Ben Arous, G., Gheissari, R., and Jagannath, A. (2022). High-dimensional limit theorems for sgd: Effective dynamics and critical scaling. *Advances in Neural Information Processing Systems*, 35:25349–25362.

[Ben Arous et al., 2023] Ben Arous, G., Huang, D. Z., and Huang, J. (2023). Long random matrices and tensor unfolding. *The Annals of Applied Probability*, 33(6B):5753–5780.

[Ben Arous et al., 2019] Ben Arous, G., Mei, S., Montanari, A., and Nica, M. (2019). The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330.

[Bhaskara et al., 2014] Bhaskara, A., Charikar, M., Moitra, A., and Vijayaraghavan, A. (2014). Smoothed analysis of tensor decompositions. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 594–603, New York, NY.

[Breiding and Vannieuwenhoven, 2018] Breiding, P. and Vannieuwenhoven, N. (2018). The condition number of join decompositions. *SIAM Journal on Matrix Analysis and Applications*, 39(1):287–309.

[Bremner and Hu, 2013] Bremner, M. R. and Hu, J. (2013). On kruskal's theorem that every $3\times 3\times 3$ array has rank at most 5. *Linear algebra and its applications*, 439(2):401–421.

[Carroll and Chang, 1970] Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319.

[Cartwright and Sturmfels, 2013] Cartwright, D. and Sturmfels, B. (2013). The number of eigenvalues of a tensor. *Linear algebra and its applications*, 438(2):942–952.

[Cattell, 1944] Cattell, R. B. (1944). "Parallel proportional profile" and other principles for determining the choice of factors by rotation. *Psychometrika*, 9(4):267–283.

[Cichocki et al., 2016] Cichocki, A., Lee, N., Oseledets, I., Phan, A.-H., Zhao, Q., Mandic, D. P., et al. (2016). Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends in Machine Learning*, 9(4-5):249–429.

[Cichocki et al., 2017] Cichocki, A., Phan, A.-H., Zhao, Q., Lee, N., Oseledets, I., Sugiyama, M., Mandic, D. P., et al. (2017). Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives. *Foundations and Trends® in Machine Learning*, 9(6):431–673.

[Comon and ten Berge, 2008] Comon, P. and ten Berge, J. M. F. (2008). Generic and typical ranks of three-way arrays. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3313–3316. IEEE.

[Comon et al., 2009] Comon, P., Ten Berge, J. M. F., De Lathauwer, L., and Castaing, J. (2009). Generic and typical ranks of multi-way arrays. *Linear Algebra and its Applications*, 430(11-12):2997–3007.

[da Silva et al., 2016] da Silva, A. P., Comon, P., and de Almeida, A. L. (2016). A finite algorithm to compute rank-1 tensor approximations. *IEEE Signal Processing Letters*, 23(7):959–963.

[De Lathauwer, 2008] De Lathauwer, L. (2008). Decompositions of a higher-order tensor in block terms–Part II: Definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1033–1066.

[De Lathauwer, 2011] De Lathauwer, L. (2011). Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms. *SIAM J. Matrix Anal. Appl.*, 32(4):1451–1474.

[De Lathauwer et al., 2000a] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000a). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278.

[De Lathauwer et al., 2000b] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000b). On the best rank-1 and rank-$(r_1, r_2, \ldots, r_n)$ approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1324–1342.

[de Silva and Lim, 2008] de Silva, V. and Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22.

[Dolgov et al., 2020] Dolgov, S., Anaya-Izquierdo, K., Fox, C., and Scheichl, R. (2020). Approximation and sampling of multivariate probability distributions in the tensor train decomposition. *Statistics and Computing*, 30:603–625.

[Domanov and Lathauwer, 2014] Domanov, I. and Lathauwer, L. D. (2014). Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition. *SIAM Journal on Matrix Analysis and Applications*, 35(2):636–660.

[Domanov and Lathauwer, 2020] Domanov, I. and Lathauwer, L. D. (2020). On uniqueness and computation of the decomposition of a tensor into multilinear rank-$(1, L_r, L_r)$ terms. *SIAM Journal on Matrix Analysis and Applications*, 41(2):747–803.

[Eckart and Young, 1936] Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

[Evert and De Lathauwer, 2022] Evert, E. and De Lathauwer, L. (2022). Guarantees for existence of a best canonical polyadic approximation of a noisy low-rank tensor. *SIAM Journal On Matrix Analysis And Applications*, 43(1):328–369.

[Favier, 2021] Favier, G. (2021). *Matrix and Tensor Decompositions in Signal Processing, Volume 2*. John Wiley & Sons.

[Favier et al., 2012] Favier, G., Kibangou, A. Y., and Bouilloc, T. (2012). Nonlinear system modeling and identification using volterra-parafac models. *International Journal of Adaptive Control and Signal Processing*, 26(1):30–53.

[Giampouras et al., 2022] Giampouras, P. V., Rontogiannis, A. A., and Kofidis, E. (2022). Block-term tensor decomposition model selection and computation: The bayesian way. *IEEE Transactions on Signal Processing*, 70:1704–1717.

[Goulart and Burt, 2021] Goulart, J. H. de M. and Burt, P. M. S. (2021). Volterra kernels of bilinear systems have tensor train structure. In *Proceedings of the 29th European Signal Processing Conference (EUSIPCO)*, pages 1070–1074.

[Goulart and Comon, 2017] Goulart, J. H. de M. and Comon, P. (2017). A novel non-iterative algorithm for low-multilinear-rank tensor approximation. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 653–657. IEEE.

[Goulart and Comon, 2019] Goulart, J. H. de M. and Comon, P. (2019). On the minimal ranks of matrix pencils and the existence of a best approximate block-term tensor decomposition. *Linear Algebra Appl.*, 561:161–186.

[Goulart et al., 2022] Goulart, J. H. de M., Couillet, R., and Comon, P. (2022). A random matrix perspective on random tensors. *Journal of Machine Learning Research*, 23(264):1–36.

[Goulart et al., 2020] Goulart, J. H. de M., de Oliveira, P. M. R., Farias, R. C., Zarzoso, V., and Comon, P. (2020). Alternating group lasso for block-term tensor decomposition and application to ECG source separation. *IEEE Transactions on Signal Processing*, 68:2682–2696.

[Hackbusch, 2012] Hackbusch, W. (2012). *Tensor spaces and numerical tensor calculus*, volume 42 of *Springer series in computational mathematics*. Springer-Verlag.

[Harshmans, 1970] Harshmans, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA working papers in phonetics*, 16(1):84.

[Hitchcock, 1927] Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189.

[Hitchcock, 1928] Hitchcock, F. L. (1928). Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, 7(1-4):39–79.

[Hong et al., 2020] Hong, D., Kolda, T. G., and Duersch, J. A. (2020). Generalized canonical polyadic tensor decomposition. *SIAM Review*, 62(1):133–163.

[Hopkins et al., 2019] Hopkins, S. B., Schramm, T., and Shi, J. (2019). A robust spectral algorithm for overcomplete tensor decomposition. In *Conference on Learning Theory*, pages 1683–1722. PMLR.

[Hsu and Kakade, 2013] Hsu, D. and Kakade, S. M. (2013). Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science (ITCS)*, pages 11–20, New York, NY, USA.

[Huang et al., 2022] Huang, J., Huang, D. Z., Yang, Q., and Cheng, G. (2022). Power iteration for tensor PCA. *Journal of Machine Learning Research*, 23(128):1–47.

[Jagannath et al., 2020] Jagannath, A., Lopatto, P., and Miolane, L. (2020). Statistical thresholds for Tensor PCA. *The Annals of Applied Probability*, 30(4):1910–1933.

[Khamidullina et al., 2023] Khamidullina, L., Seidl, G., Podkurkov, I., Korobkov, A., and Haardt, M. (2023). Enhanced solutions for the block-term decomposition in rank-$(L_r, L_r, 1)$ terms. *IEEE Transactions on Signal Processing.*

[Kofidis and Regalia, 2002] Kofidis, E. and Regalia, P. A. (2002). On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 23(3):863–884.

[Kolda, 2001] Kolda, T. G. (2001). Orthogonal tensor decompositions. *SIAM Journal on Matrix Analysis and Applications*, 23(1):243–255.

[Kolda and Hong, 2020] Kolda, T. G. and Hong, D. (2020). Stochastic gradients for large-scale tensor decomposition. *SIAM Journal on Mathematics of Data Science*, 2(4):1066–1095.

[Kolda and Mayo, 2011] Kolda, T. G. and Mayo, J. R. (2011). Shifted power method for computing tensor eigenpairs. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1095–1124.

[Kruskal, 1977] Kruskal, J. B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95–138.

[Kruskal, 1989] Kruskal, J. B. (1989). Rank, decomposition, and uniqueness for 3-way and N-way arrays. In *Multiway data analysis*, pages 7–18.

[Lebeau et al., 2024a] Lebeau, H., Chatelain, F., and Couillet, R. (2024a). A random matrix approach to low-multilinear-rank tensor approximation. *arXiv preprint arXiv:2402.03169*.

[Lebeau et al., 2024b] Lebeau, H., Seddik, M. E. A., and Goulart, J. H. de M. (2024b). Performance gaps in multi-view clustering under the nested matrix-tensor model. In *12th International Conference on Learning Representations (ICLR 2024)*, pages 1–29, Wien, Austria.

[Lebedev et al., 2015] Lebedev, V., Ganin, Y., Rakhuba, M., Oseledets, I., and Lempitsky, V. (2015). Speeding-up convolutional neural networks using fine-tuned CP-decomposition. In *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, CA, USA.

[Leurgans et al., 1993] Leurgans, S. E., Ross, R. T., and Abel, R. B. (1993). A decomposition for three-way arrays. *SIAM Journal on Matrix Analysis and Applications*, 14(4):1064–1083.

[Lim, 2005] Lim, L.-H. (2005). Singular values and eigenvalues of tensors: a variational approach. In *IEEE CAMSAP*, pages 129–132, Puerto Vallarta, Mexico.

[Montanari et al., 2017] Montanari, A., Reichman, D., and Zeitouni, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Transactions on Information Theory*, 63(3):1572–1579.

[Montanari and Richard, 2014] Montanari, A. and Richard, E. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2897–2905, Montréal, Canada.

[Novikov et al., 2021] Novikov, G. S., Panov, M. E., and Oseledets, I. V. (2021). Tensor-train density estimation. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 161 of *Proceedings of Machine Learning Research*, pages 1321–1331. PMLR.

[Oseledets, 2011] Oseledets, I. V. (2011). Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317.

[Paatero, 2000] Paatero, P. (2000). Construction and analysis of degenerate parafac models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3):285–299.

[Park and Ozeki, 2009] Park, H. and Ozeki, T. (2009). Singularity and Slow Convergence of the EM algorithm for Gaussian Mixtures. *Neural Processing Letters*, 29(1):45–59.

[Perry et al., 2020] Perry, A., Wein, A. S., and Bandeira, A. S. (2020). Statistical limits of spiked tensor models. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 230–264. Institut Henri Poincaré.

[Qi et al., 2020] Qi, Y., Michałek, M., and Lim, L.-H. (2020). Complex best r-term approximations almost always exist in finite dimensions. *Applied and Computational Harmonic Analysis*, 49(1):180–207.

[Rajih et al., 2008] Rajih, M., Comon, P., and Harshman, R. A. (2008). Enhanced line search: A novel method to accelerate PARAFAC. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1128–1147.

[Razaviyayn et al., 2013] Razaviyayn, M., Hong, M., and Luo, Z.-Q. (2013). A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153.

[Robeva, 2016] Robeva, E. (2016). Orthogonal decomposition of symmetric tensors. *SIAM Journal on Matrix Analysis and Applications*, 37(1):86–102.

[Rontogiannis et al., 2021] Rontogiannis, A. A., Kofidis, E., and Giampouras, P. V. (2021). Block-term tensor decomposition: Model selection and computation. *IEEE Journal of Selected Topics in Signal Processing*, 15(3):464–475.

[Ros et al., 2019] Ros, V., Ben Arous, G., Biroli, G., and Cammarota, C. (2019). Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions. *Physical Review X*, 9(1):11003.

[Sanchez and Kowalski, 1990] Sanchez, E. and Kowalski, B. R. (1990). Tensorial resolution: A direct trilinear decomposition. *Journal of Chemometrics*, 4(1):29–45.

[Seddik et al., 2024] Seddik, M. E. A., Guillaud, M., and Couillet, R. (2024). When random tensors meet random matrices. *The Annals of Applied Probability*, 34(1A):203–248.

[Sirovich and Kirby, 1987] Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519–524.

[Sorber et al., 2013] Sorber, L., Van Barel, M., and De Lathauwer, L. (2013). Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(l_r,l_r,1) terms, and a new generalization. *SIAM Journal on Optimization*, 23(2):695–720.

[Sumi et al., 2010] Sumi, T., Miyazaki, M., and Sakata, T. (2010). About the maximal rank of 3-tensors over the real and the complex number field. *Annals of the Institute of Statistical Mathematics*, 62(4):807–822.

[Tucker, 1966] Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311.

[Uschmajew, 2012] Uschmajew, A. (2012). Local convergence of the alternating least squares algorithm for canonical tensor approximation. *SIAM Journal on Matrix Analysis and Applications*, 33(2):639–652.

[Vannieuwenhoven, 2015] Vannieuwenhoven, N. (2015). The tensor rank decomposition. PhD thesis.

[Vannieuwenhoven et al., 2012] Vannieuwenhoven, N., Vandebril, R., and Meerbergen, K. (2012). A new truncation strategy for the higher-order singular value decomposition. *SIAM Journal on Scientific Computing*, 34(2):A1027–A1052.

[Vasilescu and Terzopoulos, 2002] Vasilescu, M. A. O. and Terzopoulos, D. (2002). Multilinear analysis of image ensembles: Tensorfaces. In *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, pages 447–460, Copenhagen, Denmark. Springer.

[Vidal, 2003] Vidal, G. (2003). Efficient classical simulation of slightly entangled quantum computations. *Physical review letters*, 91(14):147902.

[Xu, 2018] Xu, Y. (2018). On the convergence of higher-order orthogonal iteration. *Linear and Multilinear Algebra*, 66(11):2247–2265.

[Yang et al., 2018] Yang, B., Zamzam, A., and Sidiropoulos, N. D. (2018). Parasketch: Parallel tensor factorization via sketching. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 396–404. SIAM.

[Yang, 2023] Yang, Y. (2023). On global convergence of alternating least squares for tensor approximation. 84(2):509–529.