Separable Nonnegative CANDELINC: A Latent Anchor Approach to Identifiable Nonnegative Tucker Decomposition

John Hood [1] **Aaron Schein** [1]

Abstract

Nonnegative Tucker decomposition is a powerful tool for analyzing multiway data, but its parameters are generally not identifiable. Recent work addresses this problem, establishing identifiability guarantees under various sparsity assumptions on the factor matrices. This paper proposes a complementary approach inspired by the nonnegative canonical decomposition with linear constraints (nnCANDELINC). Introducing *latent anchors* to model Tucker's core tensor as a structured canonical polyadic (CP) decomposition, we combine results from CP and separable nonnegative matrix factorization to prove uniqueness through composition. We verify the uniqueness result numerically through a correspondence between uniqueness, statistical identifiability, and Fisher information. Synthetic experiments demonstrate that the proposed method bridges the gap between the identifiability of CP and expressivity of Tucker.

1. Introduction

Tensor methods have advanced scientific discovery in the biological, physical, and social sciences, with recent applications in signal processing (Sidiropoulos et al., 2017), genomics (Armingol et al., 2022), and network science (Aguiar et al., 2024; Hood et al., 2026). When using tensor methods for statistical modeling and scientific discovery, parameter identifiability, or a unique mapping from the set of all possible model parameters to the underlying scientific or statistical model is crucial for gaining a scientifically meaningful interpretation of these parameters (Wieland et al., 2021).

Nonnegative tensor decomposition methods, where the parameters and data are constrained to be nonnegative, have gained popularity due to their interpretable, 'parts-based' representation (Lee & Seung, 1999). There are two primary nonnegative tensor decomposition methods: the canonical polyadic (CP) decomposition (Hitchcock, 1927) and the Tucker decomposition (Tucker, 1966). While a tensor's CP decomposition is unique under mild conditions (Kruskal, 1977; Sidiropoulos & Bro, 2000), the Tucker decomposition is subject to rotational ambiguities and is generally not unique (Kolda & Bader, 2009). A special case of the nonnegative tensor decomposition is the nonnegative matrix factorization (NMF) (Lee & Seung, 2000). Even the NMF is not unique in general, and much work studies conditions under which it is (Gillis, 2020).

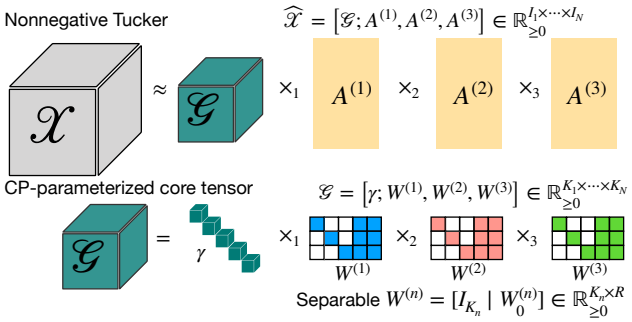

Figure 1. Reparameterizing the core tensor in Tucker. Separable nnCANDELINC parameterizes Tucker's core tensor $\mathcal{G}$ using a constrained CP decomposition $[\![\boldsymbol{\gamma}; W^{(1)}, \ldots, W^{(n)}]\!]$, in which each factor matrix takes the form $W^{(n)} = [I_{K_n} \mid W_0^{(n)}]$.

Similarly, a large body of recent work provides uniqueness results for the nonnegative Tucker decomposition under sparsity assumptions on some or all of the factor matrices (Zhou et al., 2015; Sun & Huang, 2023; Liu & Donnat, 2024; Agterberg & Zhang, 2025; Saha et al., 2025), leveraging results from the NMF literature. Each result assumes a level of sparsity on some or all of the factor matrices. In some settings these assumptions may be violated, complicating estimation and interpretation.

Rather than imposing sparsity on the factor matrices, we show that identifiability may be achieved under an assumption on the core tensor's rank. Our work extends CANDELINC (Carroll et al., 1980) and nnCANDELINC (Alexandrov et al., 2022) to the separable factor matrix setting. For estimation, we provide a block-coordinate descent algorithm which naturally extends the einsum-based

[1]Department of Statistics, University of Chicago, Illinois, U.S.A.. Correspondence to: John Hood <johnhood@uchicago.edu>.

Preprint. July 7, 2026.

framework of Hood & Schein (2026), requiring two trivial modifications to the existing algorithm.

This paper presents a complementary approach: an identifiable nonnegative Tucker decomposition that assumes low-rank structure on the *core tensor*. Inspired by ideas from the separable NMF literature, we insert *latent anchors* into the core tensor to establish identifiability. We visualize our approach in Figure 1. The proposed model is a specific instance of "nonnegative einsum factorization" (Hood & Schein, 2026), enabling straightforward optimization which avoids the vertex-hunting procedure typically required for estimating such unique factorizations. Preliminary synthetic experiments show that the proposed method achieves a happy medium between CP and Tucker. Separable nnCANDELINC achieves lower reconstruction error than the identifiable-but-constrained CP decomposition and benefits from the connection between estimation and identifiability, showing improved stability over Tucker.

2. Model and Main Result

Notation. $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ denotes the order-N tensor. Scalars $x \in \mathbb{R}$, vectors $\mathbf{x} \in \mathbb{R}^{I_1}$, and matrices $X \in \mathbb{R}^{I_1 \times I_2}$ are order-0, order-1, and order-2 tensors, respectively. The product along the nth mode of $\mathcal{X}$ with $B \in \mathbb{R}^{I_n \times J}$ has size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$ and has elements $(\mathcal{X} \times_n B)_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} \mathcal{X}_{i_1,\ldots,i_N} B_{i_n,j}$.

Tucker decomposition. The Tucker decomposition approximates $\mathcal{X}$ with a low-rank approximation $\widehat{\mathcal{X}} \approx \mathcal{X}$, in which $\widehat{\mathcal{X}}$ is parametrized by a core tensor $\mathcal{G} \in \mathbb{R}^{K_1 \times \cdots \times K_N}$ and N factor matrices $(A^{(n)})_{n=1}^N$:

$$\widehat{\mathcal{X}} = [\![\mathcal{G}; A^{(1)}, \ldots, A^{(N)}]\!], \quad A^{(n)} \in \mathbb{R}^{I_n \times K_n}. \tag{1}$$

Element-wise, the Tucker decomposition is given by

$$\widehat{\mathcal{X}}_{i_1,\ldots,i_N} = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \cdots \sum_{k_N=1}^{K_N} G_{k_1,\ldots,k_N} \prod_{n=1}^{N} A^{(n)}_{i_n k_n}. \tag{2}$$

When all elements and parameters are nonnegative, such that $G_{k_1,\ldots,k_N}, A^{(n)}_{i_n k_n} \geq 0$, this decomposition is referred to as the *nonnegative Tucker decomposition*.

The Tucker decomposition is not unique, nonnegative or otherwise (Kolda & Bader, 2009). In particular, for invertible matrix[1] $U^{(n)}$ and the same $\widehat{\mathcal{X}}$ in (1),

$$\widehat{\mathcal{X}} = \left[\!\!\left[\mathcal{G} \times_n U^{(n)}; A^{(1)}, \ldots, A^{(n)} U^{(n)^{-1}}, \ldots, A^{(n)}\right]\!\!\right].$$

That is, multiple distinct Tucker decompositions construct the same tensor $\widehat{\mathcal{X}}$.

[1] In nonnegative settings, the transformed parameters must also preserve the nonnegativity constraints.

Unique factorizations. In the matrix setting, the same problem arises. For $X = WH$, we may write $X = (WU^{-1})(UH) = \tilde{W}\tilde{H}$. A large body of work establishes conditions on W or H to obtain a unique matrix factorization (Gillis, 2020). One such condition is *separability*, also referred to as the *anchor word assumption* in the topic modeling literature (Arora et al., 2012).

A matrix X is K-separable if it can be factored into $X = WH$, $H = [I_K \mid H_0]\Pi$, in which $\Pi \in \mathbb{R}^{I_2 \times I_2}$ is a permutation matrix, $W \in \mathbb{R}^{I_1 \times K}$ and $H_0 \in \mathbb{R}^{K \times I_2 - K}$. We call $X = WH$ a *separable NMF*. A separable NMF is unique: that is, if $X = WH$ and $X = \tilde{W}\tilde{H}$ are both separable matrix factorizations, then there exists a permutation matrix Π and diagonal matrix with positive entries D such that $\tilde{W} = WD\Pi$ and $\tilde{H} = \Pi^\top D^{-1} H$. Therefore, when discussing uniqueness, it is typical to consider a factorization unique up to permutation and scaling.

Our representation. We mirror the nnCANDELINC model, imposing low-rank structure in the core tensor $\mathcal{G}$ to parameterize it as a rank-R CP decomposition:

$$\mathcal{G}_{k_1,\ldots,k_N} = \lambda \sum_{r=1}^{R} \gamma_r \prod_{n=1}^{N} W^{(n)}_{k_n,r}. \tag{3}$$

Letting $R \geq \max_n K_n$, we impose the separability condition directly on the core-factor matrices $W^{(n)} \in \mathbb{R}^{K_n \times R}$, such that the first K_n columns are the identity matrix I_{K_n}:

$$W^{(n)} = \left[I_{K_n} \mid W_0^{(n)}\right] \tag{4}$$

and refer to the first K_n columns as *latent anchors*, which anchor each of the K_n latent dimensions of the nth mode.

Under the CP decomposition of $\mathcal{G}$ (3), the Tucker decomposition of $\widehat{\mathcal{X}}$ (2) simplifies to

$$\begin{aligned}\widehat{\mathcal{X}}_{i_1,\ldots,i_N} &= \sum_{k_1,\ldots,k_N} \left[\lambda \sum_{r=1}^{R} \gamma_r \prod_{n=1}^{N} W^{(n)}_{k_n,r}\right] \prod_{n=1}^{N} A^{(n)}_{i_n,k_n} \qquad (5)\\ &= \lambda \sum_{r=1}^{R} \gamma_r \prod_{n=1}^{N} \left[\sum_{k_n=1}^{K_n} W^{(n)}_{k_n,r} A^{(n)}_{i_n,k_n}\right] \\ &= \lambda \sum_{r=1}^{R} \gamma_r \prod_{n=1}^{N} M^{(n)}_{i_n,r}, \qquad (6)\end{aligned}$$

a rank-R CP decomposition, or nonnegative CANDELINC, with separable factor matrices $M^{(n)} = A^{(n)} W^{(n)} \in \mathbb{R}^{I_n \times R}$, where each factor matrix has nonnegative rank of *at most* K_n. We call this model separable nnCANDELINC.

Remark 1. Expression (5) permits a spectrum of structure. When $K_n = R$ for all n, then $W^{(n)} = I_{K_n}$ and $M^{(n)} = A^{(n)}$, matching the rank-R CP decomposition exactly. When $R > \max_n K_n$ and $W_0^{(n)} > 0$, $\mathcal{G}$ is dense.

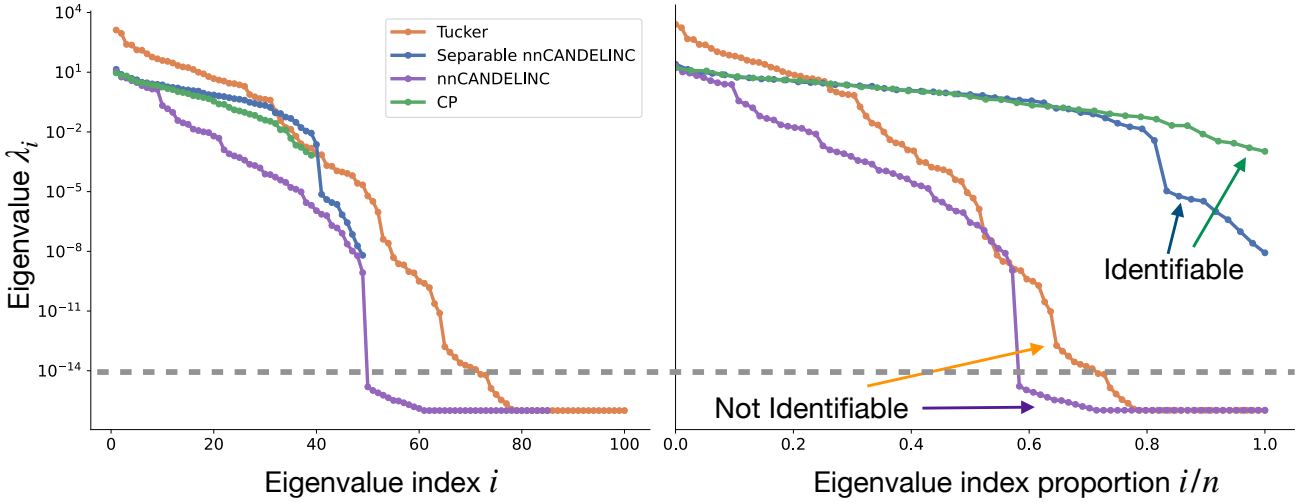

Figure 2. Spectra of the Fisher information matrices for Tucker, CP, nnCANDELINC, and separable nnCANDELINC under the Poisson model. Example is shown for $R = 5$, $N = 3$, $K_n = 4$. Tucker and the nnCANDELINC are not identifiable and have singular Fisher information matrices. Alternatively, CP and separable nnCANDELINC are both identifiable, with nonsingular Fisher information matrices.

To avoid scaling ambiguities, we require the columns of $W^{(n)}$ to lie on the simplex $\Delta^{K_n-1} := \{x \in \mathbb{R}^{K_n} \mid x_i \geq 0, \sum_{i=1}^{K_n} x_i = 1\}$. Similarly, the columns of $A^{(n)}$ lie in Δ^{I_n-1} and $\boldsymbol{\gamma} \in \Delta^{R-1}$ so that all scaling is absorbed by $\lambda \in \mathbb{R}_{>0}$.

Lemma 2.1. *Under the constraint $W^{(n)} = [I_{K_n} \mid W_0^{(n)}]$, the map $(A^{(n)}, W^{(n)}) \to A^{(n)}W^{(n)} = M^{(n)}$ is injective.*

We provide a proof in Appendix A. The representation of $\widehat{\mathcal{X}}$ under (6) highlights an equivalence between low-rank Tucker and the CP decomposition with low-rank factor matrices. Noting this, we pull from the uniqueness results established for the CP decomposition and tailor them to Tucker. We focus on the work of Kruskal (1977), which establishes conditions for uniqueness in the $N=3$ setting, as well as the $N>3$ extension of Sidiropoulos & Bro (2000).

Kruskal's main result involves a matrix's *Kruskal rank*. For a matrix A, its *Kruskal rank* $\kappa(A)$ is the maximum integer κ such that any $\kappa(A)$ columns of A are linearly independent. While, $\kappa(A) \leq \text{Rank}(A)$, they are equal almost everywhere (a.e.) (Allman et al., 2009).

Lemma 2.2. *(Kruskal). Let $\mathcal{X} = [\![\boldsymbol{\gamma}; M^{(1)}, \ldots, M^{(n)}]\!]$ be a rank-R CP decomposition of $\mathcal{X}$ where $M^{(n)}$ has Kruskal rank $\kappa(M^{(n)}) = \kappa_n$. If $2R \leq \sum_{n=1}^{N} \kappa_n - N + 1$, the decomposition is unique up to permutation and scaling.*

Intuitively, Kruskal's theorem states that if the CP rank R is bounded by a sum involving the factor matrices' Kruskal ranks, then the CP decomposition is unique. Operating under this assumption, Lemmas 2.1 and 2.2 provide the basis for our uniqueness result, which we state next.

Theorem 2.3. *Let $2R \leq \sum_{n=1}^{N} K_n - N + 1$. Then the decomposition (5) is unique up to permutation a.e.*

The proof is given in Appendix A. We apply a compositional argument to prove injectivity. By Lemma 2.1, $A^{(n)}W^{(n)}$ uniquely reconstructs $M^{(n)}$. The result of Allman et al. (2009) implies $M^{(n)}$ has Kruskal rank $\kappa_n = K_n$ almost everywhere, with respect to the Lebesgue measure on the defined parameter space. Kruskal's theorem implies that the rank-R CP decomposition $[\![\boldsymbol{\gamma}; M^{(1)}, \ldots, M^{(n)}]\!]$ is unique, and each factor matrix has a unique separable NMF $M^{(n)} = A^{(n)}W^{(n)}$. Overall, each distinct configuration $(\lambda, \boldsymbol{\gamma}, (A^{(n)}, W^{(n)})_{n=1}^{N})$ yields a distinct $\widehat{\mathcal{X}}$.

3. Empirical Results

First, we validate Theorem 2.3 numerically. Then, through synthetic experiments, we evaluate our method's empirical expressivity and identifiability relative to CP and Tucker.

3.1. Numerical Validation

A parametric statistical model P_θ is identifiable if the mapping $\theta \to P_\theta$ is injective for $\theta \in \Theta$. We draw on Rothenberg (1971), who relates P_θ's identifiability to its Fisher information matrix $\mathcal{I}_\theta = \mathbb{E}_{\mathcal{X}\sim P_\theta}[\nabla_\theta \log P_\theta(\mathcal{X})][\nabla_\theta \log P_\theta(\mathcal{X})]^\top]$. Rothenberg (1971) proved that P_θ is locally identifiable at θ iff $\mathcal{I}_\theta$ is nonsingular. We consider the Poisson tensor model:

$$\mathcal{X}_{i_1,\ldots,i_N} \mid \mathcal{X}^* \overset{\text{ind.}}{\sim} \text{Poisson}(\mathcal{X}^*_{i_1,\ldots,i_N}). \tag{7}$$

In the Poisson Tucker model, the rate tensor is defined by $\mathcal{X}^* = [\![\mathcal{G}; A^{(1)}, \ldots, A^{(n)}]\!]$. Poisson Tucker is not locally identifiable: the Fisher information is singular due to the rotational invariances described in Section 2. However, the unique parameterization $(\lambda, \boldsymbol{\gamma}, A^{(n)}, W^{(n)})$ implies local identifiability. We verify this implication, using a Monte Carlo approximation to estimate the Fisher information

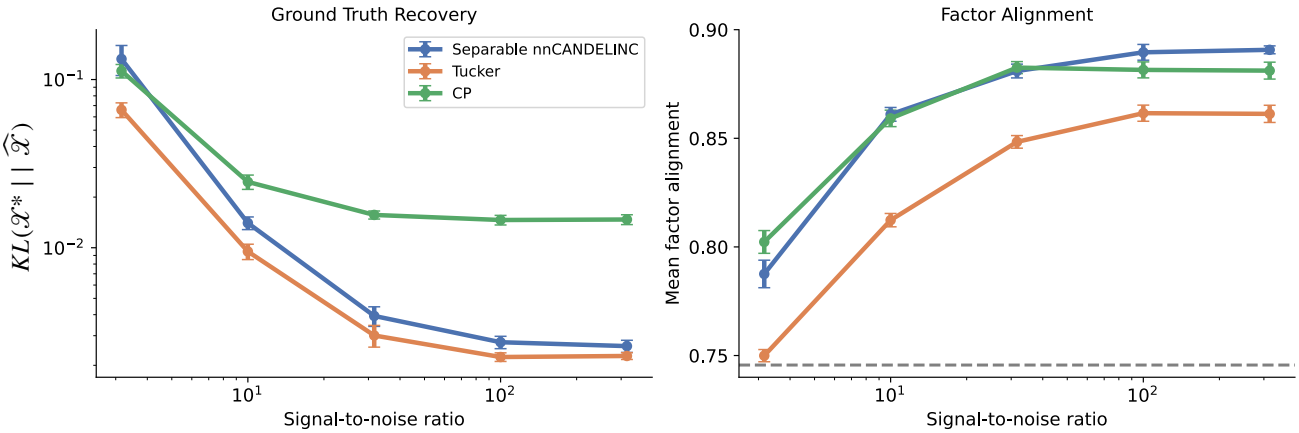

Figure 3. Separable nnCANDELINC achieves the identifiability of CP and expressivity of Tucker. Left: ground truth recovery of the true underlying rate tensor $\mathcal{X}^*$. Right: Mean factor alignment across re-runs. Example is shown for $R = 11$, $N = 3$, $K_n = 8$, $I_n = 10$. The gray line denotes the factor alignment for two matrices of size $I_n \times K_n$ with columns uniformly drawn from the simplex $\Delta^{I_n - 1}$.

matrices of the CP, Tucker, nnCANDELINC and separable nnCANDELINC under the simplex constraints from Section 2. For each model, we sample parameters uniformly and estimate $\mathcal{I}_\theta$ and its eigendecomposition $\mathcal{I}_\theta = V \Lambda V^\top$. Figure 2 shows the eigenvalues λ_i in descending order. We note that for Tucker, due to its rotational invariances, there are many zero eigenvalues, implying singular $\mathcal{I}_\theta$. However, for CP and separable nnCANDELINC, the smallest eigenvalues are approximately 10^{-4} and 10^{-8}, respectively, suggesting nonsingular $\mathcal{I}_\theta$.

3.2. Estimation, Recovery, and Alignment

We modify the block coordinate descent method of Hood & Schein (2026) to minimize the KL divergence from $\mathcal{X}$ to $\widehat{\mathcal{X}}$, initializing the $W^{(n)}$ to be separable and normalizing parameters post-training to satisfy the simplex constraints specified in Section 2. Due to the reduction in core tensor parameters, the computational and memory costs reduce from $\mathcal{O}(\prod_{n=1}^N K_n)$ of standard Tucker to $\mathcal{O}(\sum_{n=1}^N (R - K_n) K_n)$ of separable nnCANDELINC: a reduction from exponential in N to linear.

Here, we show that even when the core tensor is full rank, the sacrifice in recovering the true rate tensor is small, relative to that of CP. We generate synthetic tensors $\mathcal{X}^* = [\![\mathcal{G}; A^{(1)}, A^{(2)}, A^{(3)}]\!]$ in which the ground-truth is full Tucker. We generate $\mathcal{G}, A^{(n)}$ randomly: each column of each factor is drawn uniformly from the simplex, as is the core tensor. As such, the core tensor is full-rank and $||\mathcal{X}^*||_1 = 1$. We sample each element $\mathcal{X}_{i_1,\ldots,i_N} \overset{\text{ind.}}{\sim} \text{Poisson}(c \cdot \mathcal{X}^*_{i_1,\ldots,i_N})$. The scalar $c \in \mathbb{R}_{>0}$ dictates the signal-to-noise ratio. We vary $c \in \{10^4, 10^5, \ldots, 10^8\}$ for a signal-to-noise ratio of $\sqrt{c}/(\prod_{n=1}^N I_n)$. We then fit each model to $\mathcal{X}$ ten times to yield a reconstruction $\widehat{\mathcal{X}}$ for each.

Evaluation. We evaluate each method's ability to recover the true underlying tensor, $\mathcal{X}^*$, by measuring the KL divergence between $\mathcal{X}^*$ and the estimate $\widehat{\mathcal{X}}$. We also evaluate each method's stability over random initializations, defined next.

For two factor matrices $U^{(n)}, \tilde{U}^{(n)} \in \mathbb{R}^{I_n \times K_n}$, we report the mean cosine similarity of their columns, after matching via the Hungarian matching algorithm. Across the ten runs, we compute mean cosine similarity, averaging over modes, for each model. We call this metric the *mean factor alignment*, which lies in $[0, 1]$; higher values imply greater similarity.

Results. In Figure 3, separable nnCANDELINC and Tucker nearly recover $\mathcal{X}^*$ as the signal-to-noise ratio increases. The CP decomposition, due to its model misspecification, achieves the worst fit. In the lower plot, we observe that CP and separable nnCANDELINC achieve similar alignment scores, while Tucker's is significantly lower. This result is consistent with our expectations. Taken together, separable nnCANDELINC achieves both comparable alignment to CP and comparable recovery of $\mathcal{X}^*$ to full-rank Tucker.

4. Conclusion

This paper introduces separable nnCANDELINC, a unique parameterization of the nonnegative Tucker decomposition obtained through a compositional approach to identification. In preliminary experiments, the proposed identifiable model achieves fit competitive with unconstrained, unidentifiable Tucker. Since the proposed method parameterizes a nonnegative Tucker decomposition, it may also be interpreted as such, permitting widespread use in many scientific settings. Future directions include integration into Bayesian methods, additional experimentation, and understanding the method's scalability and limitations under model misspecification.

Software and Data

Code is available at https://github.com/jhood3/einfact.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Agterberg, J. and Zhang, A. R. Estimating higher-order mixed memberships via the $\ell_{2,\infty}$ tensor perturbation bound. *Journal of the American Statistical Association*, 120(550):1214–1224, 2025.

Aguiar, I., Taylor, D., and Ugander, J. A tensor factorization model of multilayer network interdependence. *Journal of Machine Learning Research*, 25(282):1–54, 2024.

Alexandrov, B., DeSantis, D. F., Manzini, G., and Skau, E. W. Nonnegative canonical tensor decomposition with linear constraints: nncandelinc. *Numerical Linear Algebra with Applications*, 29(6):e2443, 2022.

Allman, E. S., Matias, C., and Rhodes, J. A. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37, 2009.

Armingol, E., Baghdassarian, H. M., Martino, C., Perez-Lopez, A., Aamodt, C., Knight, R., and Lewis, N. E. Context-aware deconvolution of cell–cell communication with tensor-cell2cell. *Nature Communications*, 13(1): 3665, 2022.

Arora, S., Ge, R., and Moitra, A. Learning topic models–going beyond svd. In *2012 IEEE 53rd annual symposium on foundations of computer science*, pp. 1–10. IEEE, 2012.

Carroll, J. D., Pruzansky, S., and Kruskal, J. B. Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45(1):3–24, 1980.

Gillis, N. *Nonnegative Matrix Factorization*. SIAM, 2020.

Hitchcock, F. L. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927. doi: 10.1002/sapm192761164.

Hood, J. and Schein, A. Near-universal multiplicative updates for nonnegative einsum factorization. *Forty-Third International Conference of Machine Learning*, 2026.

Hood, J., De Bacco, C., and Schein, A. Broad spectrum structure discovery in large-scale higher-order networks. *Nature Communications*, 2026.

Kolda, T. G. and Bader, B. W. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Kruskal, J. B. 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, 1977.

Lee, D. and Seung, H. S. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 13, 2000.

Lee, D. D. and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788–791, October 1999. ISSN 0028-0836, 1476-4687. doi: 10.1038/44565. URL https://www.nature.com/articles/44565.

Liu, Y. and Donnat, C. Tensor topic modeling via hosvd. *arXiv preprint arXiv:2501.00535*, 2024.

Rothenberg, T. J. Identification in parametric models. *Econometrica: Journal of the Econometric Society*, pp. 577–591, 1971.

Saha, S., Barbarino, G., and Gillis, N. Identifiability of nonnegative tucker decompositions–part i: Theory. *arXiv preprint arXiv:2505.12713*, 2025.

Sidiropoulos, N. D. and Bro, R. On the uniqueness of multilinear decomposition of n-way arrays. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3):229–239, 2000.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

Sun, Y. and Huang, K. Volume-regularized nonnegative tucker decomposition with identifiability guarantees. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

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

Wieland, F.-G., Hauber, A. L., Rosenblatt, M., Tönsing, C., and Timmer, J. On structural and practical identifiability. *Current Opinion in Systems Biology*, 25:60–69, 2021.

Zhou, G., Cichocki, A., Zhao, Q., and Xie, S. Efficient nonnegative Tucker decompositions: Algorithms and uniqueness. *IEEE Transactions on Image Processing*, 24(12): 4990–5003, 2015.

A. Proofs.

Proof of Lemma 2.1.

Proof. WLOG, we drop the dependence on (n) in our notation. Suppose that $AW = M = \tilde{A}\tilde{W}$. Then

$$M = AW = A[I_K \mid W_0] = [A \mid AW_0]. \tag{8}$$

Also,

$$M = \tilde{A}\tilde{W} = \tilde{A}[I_K \mid \tilde{W}_0] = [\tilde{A} \mid \tilde{A}\tilde{W}_0]. \tag{9}$$

Setting the first blocks equal to each other yields $A = \tilde{A}$. Then, since $A \in \mathbb{R}^{I_n \times K_n}$, $K_n \leq I_n$, we may take the left inverse of A, which yields

$$AW = \tilde{A}\tilde{W} = A\tilde{W} \implies W = \tilde{W}. \tag{10}$$

As such, if $(A, W) \to M, (\tilde{A}, \tilde{W}) \to M$, then $(A, W) = (\tilde{A}, \tilde{W})$. We have shown injectivity. □

Proof of Theorem 2.3.

Proof. We must show that the mapping $(\lambda, \boldsymbol{\gamma}, A^{(n)}, W^{(n)}) \to \widehat{\mathcal{X}}$ is injective up to permutation. In particular, we will show that if $(\lambda, \boldsymbol{\gamma}, A^{(n)}, W^{(n)}), (\tilde{\lambda}, \tilde{\boldsymbol{\gamma}}, \tilde{A}^{(n)}, \tilde{W}^{(n)}) \to \widehat{\mathcal{X}}$, then there exists a permutation matrix $\Pi \in \{0,1\}^{R \times R}$ and permutation matrices $\tilde{\Pi}^{(n)} \in \{0,1\}^{K_n \times K_n}$ such that

i. $\lambda = \tilde{\lambda}$

ii. $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}\Pi$

iii. $\tilde{A}^{(n)} = A\tilde{\Pi}^{(n)\top}$

iv. $\tilde{W}^{(n)} = \tilde{\Pi}^{(n)} W^{(n)} \Pi$.

i. Due to the simplex constraints of Section 2, $\tilde{\lambda} = ||\widehat{\mathcal{X}}||_1 = \lambda$.

ii-iv. Suppose that

$$[\![\boldsymbol{\gamma}; M^{(1)}, \ldots, M^{(n)}]\!] = \widehat{\mathcal{X}}/\lambda = [\![\tilde{\boldsymbol{\gamma}}; \tilde{M}^{(1)}, \ldots, \tilde{M}^{(n)}]\!] \tag{11}$$

where $M^{(n)} = A^{(n)} W^{(n)}$ and $\tilde{M}^{(n)} = \tilde{A}^{(n)} \tilde{W}^{(n)}$.

Under this parameterization, we claim that the Kruskal rank $\kappa(M^{(n)})$ is maximal. That is, $\kappa(M^{(n)}) = \text{Rank}(M^{(n)}) = K_n$ almost surely under any continuous distribution over the entries $A^{(n)}$ and $W_0^{(n)}$. By definition,

$$\kappa(M^{(n)}) \leq \text{Rank}(M^{(n)}) = K_n. \tag{12}$$

Consider any subset $\mathcal{J}$ of K_n columns of $M^{(n)}$, denoted by $M_{\mathcal{J}}^{(n)} = A^{(n)} W_{\mathcal{J}}^{(n)}$. Since $W^{(n)} = [I_{K_n} \mid W_0^{(n)}]$ the entries of W_0 are strictly positive and continuously generated, $W_{\mathcal{J}}^{(n)}$ has maximal rank K_n, and so its columns are linearly independent. Similarly, $A^{(n)}$ with continuously-generated entries is injective for $I_n \geq K_n$. Therefore, linear independence is preserved, and so the columns of $A^{(n)} W_{\mathcal{J}}^{(n)} = M_{\mathcal{J}}^{(n)}$ are linearly independent. This holds for any subset of K_n columns indexed by $\mathcal{J} \subseteq [R]$, therefore $\kappa(M^{(n)}) \geq K_n$. We conclude that $K_n = \kappa(M^{(n)})$.

Thus, letting $2R \leq \sum_{n=1}^{N} K_n - N + 1$, we apply Kruskal's theorem. By Kruskal's theorem, there exist diagonal matrices $D^{(n)} \in \mathbb{R}_{\geq 0}^{R \times R}$ and permutation matrix $\Pi \in \{0,1\}^{R \times R}$ such that

$$\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}\Pi, \quad \tilde{M}^{(n)} = M^{(n)} D^{(n)} \Pi. \tag{13}$$

The scaling ambiguities of $D^{(n)}$ are deleted by the simplex constraint. Since

$$1 = \sum_{i_n=1}^{I_n} \tilde{M}_{i_n,r} = d^{(n)}_{r,r} \sum_{i_n=1}^{I_n} M^{(n)}_{i_n,r} \tag{14}$$

$$= d^{(n)}_{r,r} \sum_{i_n=1}^{I_n} \sum_{k_n=1}^{K_n} A^{(n)}_{i_n,k_n} W^{(n)}_{k_n,r} \tag{15}$$

$$= d_{r,r} \sum_{k_n=1}^{K_n} \left(\sum_{i_n=1}^{I_n} A^{(n)}_{i_n,k_n} \right) W^{(n)}_{k_n,r} = d_{r,r} \sum_{k_n=1}^{K_n} W^{(n)}_{k_n,r} \tag{16}$$

$$= d_{r,r} \tag{17}$$

for each column $r \in [R]$, $D^{(n)} = I_R$.

So $\tilde{M}^{(n)} = M^{(n)}\Pi$, i.e. $\tilde{A}^{(n)}\tilde{W}^{(n)} = A^{(n)}W^{(n)}\Pi = A^{(n)}\bar{W}^{(n)}$, letting $\bar{W}^{(n)} = W^{(n)}\Pi$. Since $\bar{W}^{(n)}\Pi^\top = W^{(n)}\Pi\Pi^\top = W^{(n)} = [I_{K_n} \mid W_0^{(n)}]$, $\bar{W}^{(n)}$ is separable. So $\tilde{A}^{(n)}\tilde{W}^{(n)}$ and $A^{(n)}\bar{W}^{(n)}$ are separable NMFs of the same matrix, implying that, by Gillis (2020), there exists a permutation matrix $\tilde{\Pi}^{(n)} \in \{0,1\}^{K_n \times K_n}$ such that

$$\tilde{A}^{(n)} = A^{(n)}\tilde{\Pi}^{(n)\top}, \quad \tilde{W}^{(n)} = \tilde{\Pi}^{(n)}\bar{W}^{(n)} = \tilde{\Pi}^{(n)}W^{(n)}\Pi. \tag{18}$$

□