Provable Online CP/PARAFAC Decomposition of a Structured Tensor via Dictionary Learning

Sirisha Rambhatla
sirishar@usc.edu
USC University of Southern California

Xingguo Li
xingguol@cs.princeton.edu
Princeton University

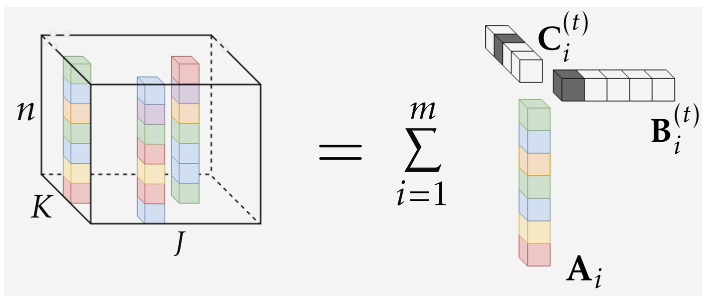
Jarvis Haupt
jdhaupt@umn.edu
University of Minnesota Driven to Discover

Neural Information Processing Systems

0. Overview

We consider the problem of factorizing a structured 3-way tensor into its constituent Canonical Polyadic (CP) factors. This decomposition, which can be viewed as a generalization of singular value decomposition (SVD) for tensors, reveals how the tensor dimensions (features) interact with each other.

Model



Goal: Recover $\mathbf{A}$, $\mathbf{B}^{(t)}$, $\mathbf{C}^{(t)}$ given $\underline{\mathbf{Z}}^{(t)}$

$\underline{\mathbf{Z}}^{(t)} \in \mathbb{R}^{n \times J \times K}$

$$\text{Tensor} \quad \underline{\mathbf{Z}}^{(t)} = \sum_{i=1}^{m} \mathbf{A}_i \circ \mathbf{B}_i^{(t)} \circ \mathbf{C}_i^{(t)}$$

Sparse

Incoherent columns

$\mathbf{A} \in \mathbb{R}^{n \times m}$

$\mathbf{B}^{(t)} \in \mathbb{R}^{J \times m}$

$\mathbf{C}^{(t)} \in \mathbb{R}^{K \times m}$

Since the factors are *a priori* unknown, the corresponding optimization problems are inherently non-convex. The existing guaranteed algorithms which handle this non-convexity **incur an irreducible error (bias)**, and only apply to cases where all factors have the same structure.

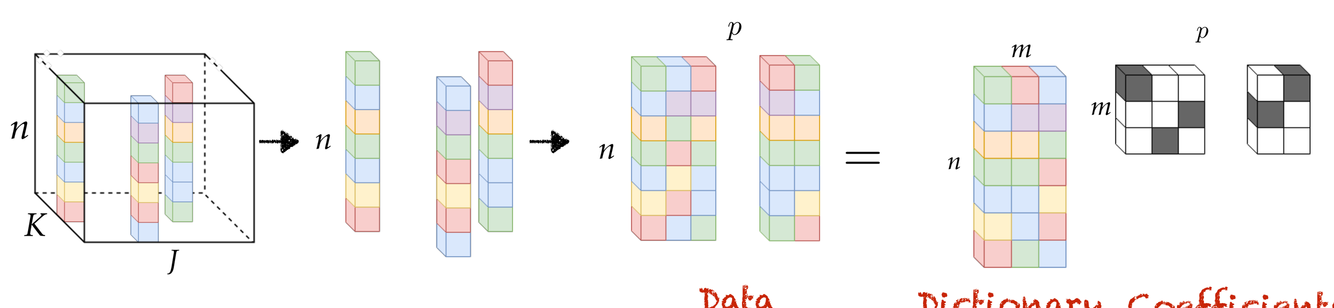
Under some relatively mild conditions on initialization, rank, and sparsity,

Contributions

TensorNOODL

- Recovers all factors
- Unbiased Estimation
- Linear Convergence
- Online Algorithm
- *"Neurally Plausible"*
- Scalable
- Easy to tune

1. How?



$\underline{\mathbf{Z}}^{(t)}$ $\mathbf{Y}$ $\mathbf{A}$ $\mathbf{X}$

A Dictionary Learning Problem!

2. Main Result

Algorithm 1: TensorNOODL: Neurally plausible alternating Optimization-based Online Dictionary Learning for Tensor decompositions.

- Input: Tensor $\underline{Z}^{(t)}$, an appropriate initial estimate $A^{(0)}$ of A, and parameters
- Output: Estimate of A, $B^{(t)}$, and $C^{(t)}$ at each iteration t
- Estimate the sparse matrix X for the current iterate — Iterative Hard Thresholding
- Untangle the Khatri-Rao Product Structure: Recover $B^{(t)}$ and $C^{(t)}$ given X — Singular Value Decomposition-based Algorithm
- Update the incoherent factor A based on X — Approximate Gradient Descent

TensorNOODL

Main Result Under some conditions on initialization and incoherence of $\mathbf{A}^{(t)}$ and sparsity of $\mathbf{B}^{(t)}$ and $\mathbf{C}^{(t)}$, with probability at least $(1-\delta_{\text{alg}})$ for some small constant δ_{alg} the estimate $\hat{\mathbf{X}}^{(t)}$ at t-th iteration has the correct signed-support and satisfies

$$(\hat{\mathbf{X}}_{i,j}^{(t)} - \mathbf{X}_{i,j}^{*(t)})^2 \leq \zeta^2 := \mathcal{O}(s(1-\omega)^{t/2}\|\mathbf{A}_i^{(0)} - \mathbf{A}_i^*\|), \forall (i,j) \in \text{supp}(\mathbf{X}^{*(t)}).$$

Consequently, `UntangleKRP` recovers the supports of the sparse factors $\mathbf{B}^{*(t)}$ and $\mathbf{C}^{*(t)}$ correctly, and $\|\hat{\mathbf{B}}_i^{(t)} - \mathbf{B}_i^{*(t)}\|_2 \leq \epsilon_B$ and $\|\hat{\mathbf{C}}_i^{(t)} - \mathbf{C}_i^{*(t)}\|_2 \leq \epsilon_C$, where $\epsilon_B = \epsilon_C = \mathcal{O}(\frac{\zeta^2}{\alpha\beta})$.

Furthermore, the estimate $\mathbf{A}^{(t)}$ at t-th iteration satisfies

$$\|\mathbf{A}_i^{(t)} - \mathbf{A}_i^*\|^2 \leq (1-\omega)^t\|\mathbf{A}_i^{(0)} - \mathbf{A}_i^*\|^2, \ \forall\ t = 1, 2, \ldots$$

for some $0 < \omega < 1/2$.

Assumptions

Incoherence of A*	Properties of Sparse Factors	Appropriate Initialization	Sparsity	Parameter Choice
Columns of $\mathbf{A}^*$ are sufficiently "spread-out"	Two sources of randomness: **Location** of the non-zero entries (Support) and, **Values** taken by the non-zero entries	$\|\mathbf{A}^{(0)} - \mathbf{A}^*\| \leq 2\|\mathbf{A}^*\|$ $\|\mathbf{A}_i^{(0)} - \mathbf{A}_i^*\| \leq \epsilon_0$	$\alpha\beta = \mathcal{O}(\sqrt{n}/m\mu\ \log(n))$ for $m = \Omega(\log(\min(J,K))/\alpha\beta)$	Appropriately chosen step size and threshold parameters

3. Dealing with the Kronecker Dependence Structure: Details

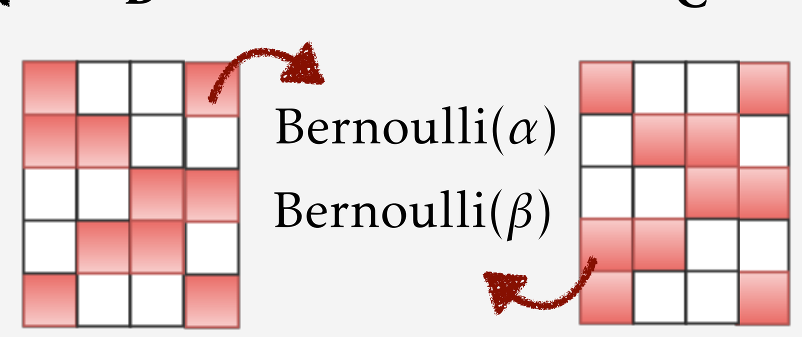
a.

Khatri-Rao Product (KRP)

$$\mathbf{Z}_1^\top = \mathbf{A}\underbrace{(\mathbf{C} \odot \mathbf{B})^\top}_{\mathbf{S}} \in \mathbb{R}^{n \times JK}$$

For Dictionary Learning $\mathbf{Y} =$ ~~non-zero columns~~ non-zero and independent of $\mathbf{Z}_1^T$

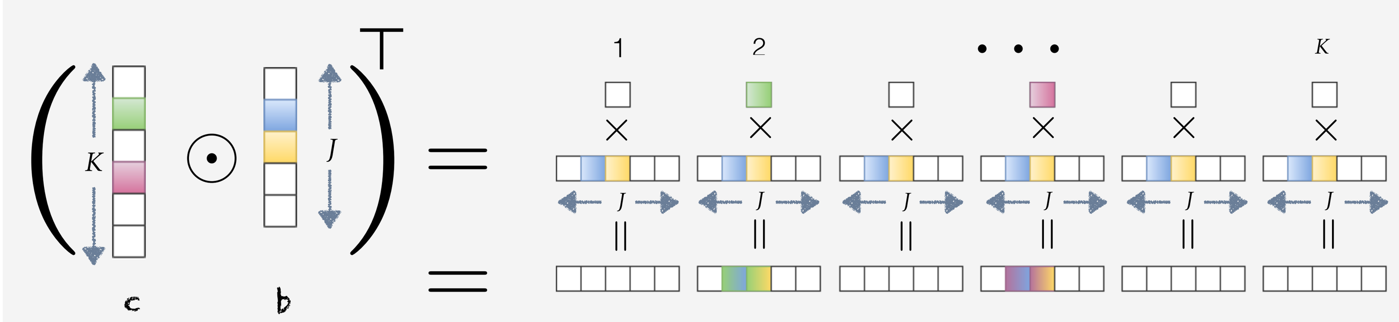
b.



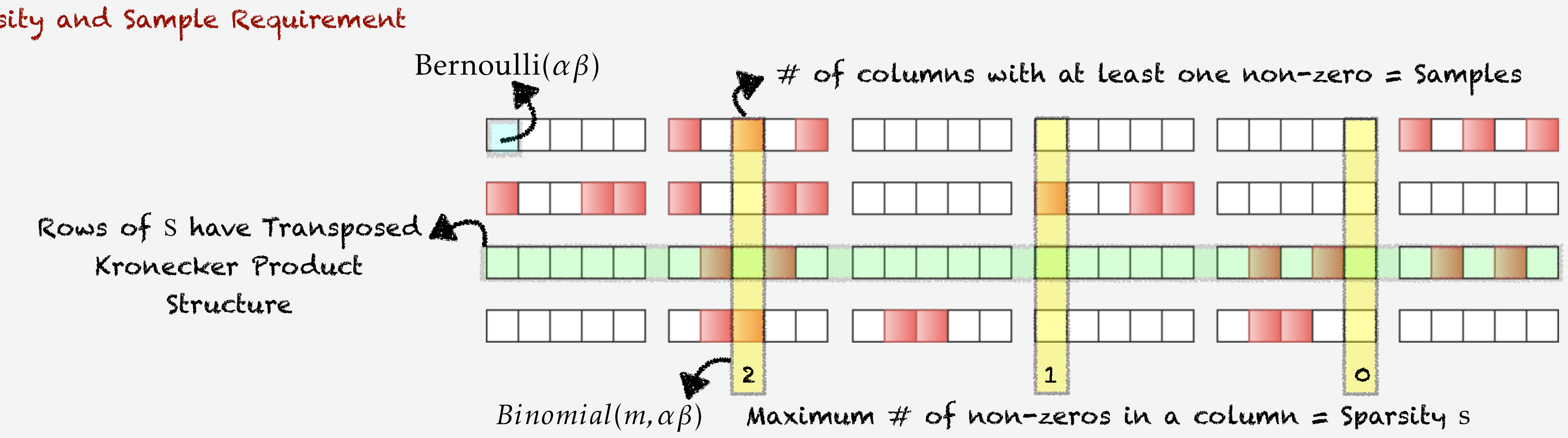
Entries from sub-Gaussian Distribution

Entries from Rademacher Distribution

c. Origin of Kronecker Product Dependence Structure



d. Sparsity and Sample Requirement



4. Experiments

Synthetic Data

Fig. 1: Convergence Performance. We compare the performance of **TensorNOODL** with other tensor-structure agnostic DL algorithms [3,4]

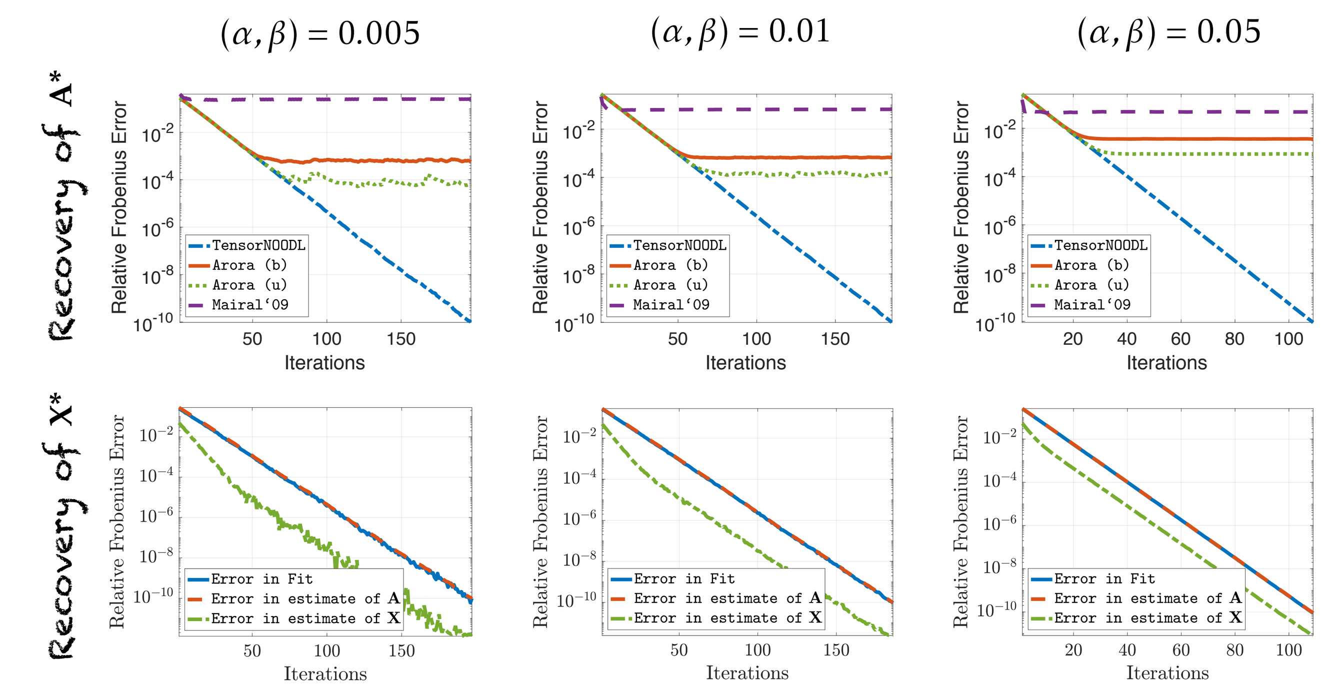
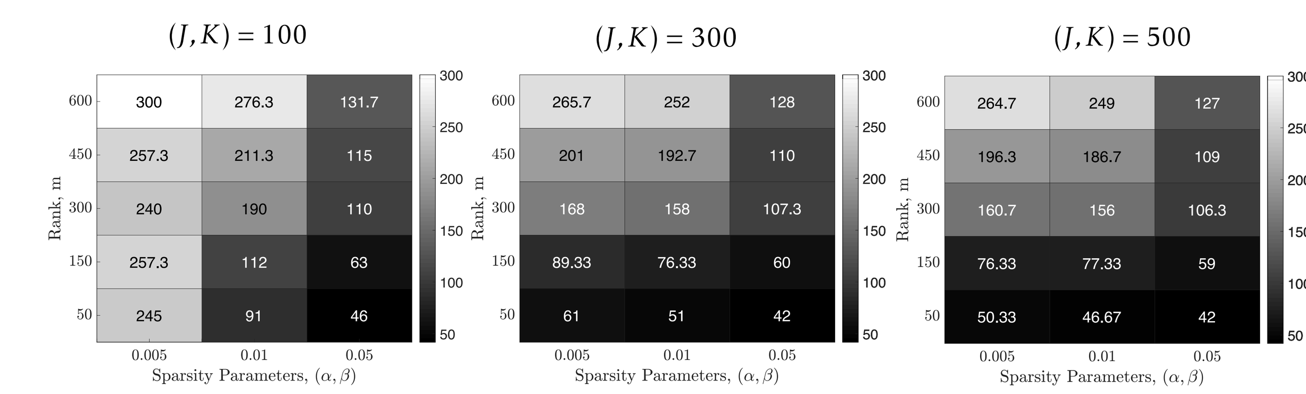


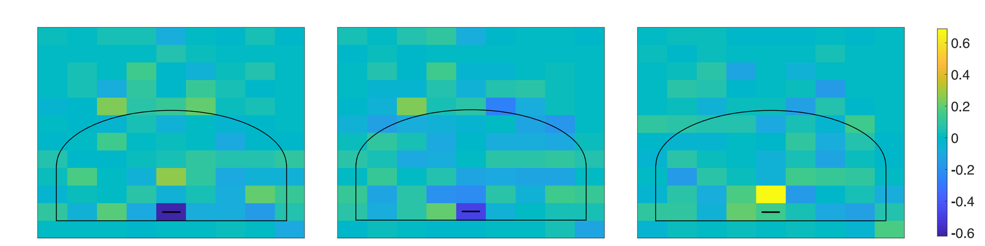
Fig. 2: Number of Iterations as a Surrogate for Sample Requirement to Reach a Target Tolerance of 10^{-10}



Take-away: TensorNOODL achieves orders of magnitude better performance, recovering the factors at a linear rate, while also providing guarantees on recovery of ALL factors!

Real-World Data

Fig. 3: NBA Shot-Pattern Analysis
100 Players, 27 Weeks from 2018-19



(a) Element 4   (b) Element 5   (c) Element 6

Corresponding Sparse factor (Players) Coefficients

Player	Element 4	Element 5	Element 6
James Harden	0.1992	0.0678	0.2834
Devin Booker	0.0114	0.0104	0.4668

TensorNOODL

https://github.com/srambhatla/TensorNOODL
https://arxiv.org/abs/2006.16442

References

[1] S. Rambhatla, X. Li, and J. Haupt. Provable Online CP/PARAFAC Decomposition of a Structured Tensor via Dictionary Learning. To appear in the proceedings of the *Neural Information Processing Systems (NeurIPS)*, 2020.

[2] S. Rambhatla, X. Li, and J. Haupt. NOODL: Provable Online Learning for Dictionary Learning and Sparse Coding. *International Conference on Learning Representations (ICLR)*, 2019.

[3] S. Arora, R. Ge, T. Ma and A. Moitra. Simple, efficient, and neural algorithms for sparse coding. *In Conference on Learning Theory (COLT)*, 2015.

[4] J. Mairal, F. Bach, J. Ponce and G. Sapiro. Online dictionary learning for sparse coding. In Proceedings of the *International Conference on Machine Learning (ICML)*. 2009.

The authors graciously acknowledge the support from the DARPA YFA, Grant N66001-14-1-4047. The authors also express their gratitude to Prof. Nikos Sidiropoulos and Di Xiao for helpful discussions.