Introduction

Hierarchical Models

Experiments

Conclusions

## **Hierarchical Nonnegative Tensor Decompositions**

by Jamie Haddock (Harvey Mudd College, Department of Mathematics)
on September 30, 2023,
AWM Research Symposium 2023 "Tensor Methods for data modeling"

> https://ieeexplore.ieee.org/document/9022678 (CAMSAP 2019) joint with M. Gao<sup>•</sup>, D. Molitor, E. Sadovnik, T. Will<sup>•</sup>, R. Zhang<sup>•</sup>, D. Needell

https://ieeexplore.ieee.org/document/9723126 (ACSSC 2021) joint with Joshua Vendrow<sup>•</sup>, Deanna Needell

https://ieeexplore.ieee.org/document/9747810 (ICASSP 2022) joint with Joshua Vendrow<sup>●</sup>, Deanna Needell

NSF\_DMS\_#2211318



Introduction

Hierarchical Models

Experiments

Conclusions

Consider applying for the Institute for Advanced Study <u>Women and Mathematics</u> summer program!

Deadline is in February each year.

Motivation ●○○○

Introduction

Hierarchical Models

Experiment:

Conclusions

## Motivation

[3/20]

Motivation ○●○○	1	Introd	<b>uction Hierarchics</b>		E <b>xperin</b> 00	ients			<b>Conclusions</b> 000
ſ	my migraines. issues too, but the	Of course I migraines	are my main	h-dimen		o <b>na</b> Patier		lat	a
		ctor was gr	eat, realized it was a	heart	3	3	0	1	
	to recog pain but on and weaknes	just st	quick. I didn't quite	weakness	2	0	0	0	
	and dar expect for pain is high cho to doct week bef	realized Sometim	it was exactly what she had.	chest	0	2	0	0	
a	any pre had hear want m few hour took an l	vision.	definitely there but really I felt more a	migraine	0	0	2	3	
	from tha attack ar team of d	m tha ack ar m of extend o driving. drink a l lighthead	tightness in my chest than anything. It left me short of breath, which was probably making me lightheaded. The EKG indicated that my heart had several blockages that would need a stent. My	lightheaded	0	2	2	1	
				pain		2	2		
		most deb	After my heart attack, I completely changed my lifestyle. I quit smoking, started an exercise regimen and diet						

Patient Surveys

**Term-Document Matrix** 

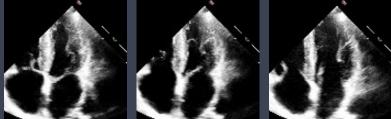
Motivation ○●○○		Introd 0000			<b>xperin</b>	nents			Conclusions
n issu	ny migraines. C aes too, but the :	)f course I   migraines ;	are my main	h-dimens		na Patier		lat	a
tha ligh	t pai ithe: My doo	ow. My priority is getting doctor was great, realized it was a		heart		3		1	
con	trol know wa ecos pain but	ow wa in but just str aknes I told her pect fc realized gh cho Sometim	quick. I didn't quite	weakness	2	0	0	0	
and pai	l dai expect fc n is : high cho locto week bet		it was exactly what she had.	chest	0			0	
any	any pre want m from tha attack ar team of c	vision. I the pain aggravate	chest pain. I had been feeling lightheaded and nauseous. The pain was definitely there but really I felt more a	migraine	0	0			
		om tha of use of the stand of t	tightness in my chest than anything. It left me short of breath, which was probably making me lightheaded. The EKG	lightheaded	0	2	2	1	
			blockages that would need a stent. My cardiologists were able to clear the	pain	3	2		4	
	p	ationt	Sumouc	Т	arm_	Docu	mont	t Matr	iv

Patient Surveys

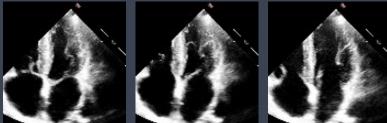
**Term-Document Matrix** 

Understand symptom trends and shared patient experiences automatically.





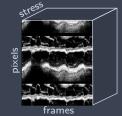








Learn cohesive parts and separate noise in medical image studies.



Introduction

Hierarchical Models

Experiments

Conclusions

#### Can we tell how the resulting parts/topics are related?

#### Can we tell how the resulting parts/topics are related?

#### How do we choose the number of topics or parts to learn?

#### Can we tell how the resulting parts/topics are related?

How do we choose the number of topics or parts to learn?

# Hierarchical matrix factorization and tensor decomposition topic models!

Introduction ●○○○○ Hierarchical Models

Experiment:

Conclusions

## Introduction



## » Nonnegative Matrix Factorization (NMF)

 $\boldsymbol{\mathsf{Model}}:$  Given nonnegative data  $\boldsymbol{\mathsf{X}},$  compute nonnegative  $\boldsymbol{\mathsf{A}}$  and  $\boldsymbol{\mathsf{S}}$  of lower rank so that

 $X \approx AS$ .



Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.



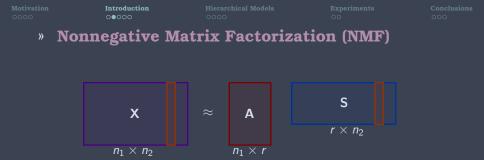
## » Nonnegative Matrix Factorization (NMF)

 $\boldsymbol{\mathsf{Model}}:$  Given nonnegative data  $\boldsymbol{\mathsf{X}},$  compute nonnegative  $\boldsymbol{\mathsf{A}}$  and  $\boldsymbol{\mathsf{S}}$  of lower rank so that

 $X \approx AS$ .



Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.



Employed for dimensionality-reduction and topic modeling

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.





- Employed for dimensionality-reduction and topic modeling
- ▷ Often formulated as

$$\min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0}, \mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} \|\mathbf{X}-\mathbf{AS}\|_F^2 \quad \text{or} \quad \min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0}, \mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} D(\mathbf{X}\|\mathbf{AS}).^1$$

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.





 $n_1 \times n_2$ 

Employed for dimensionality-reduction and topic modeling

▷ Often formulated as

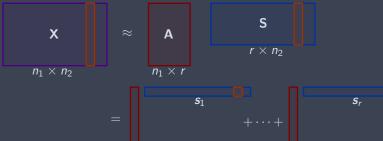
$$\min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0}, \mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} \|\mathbf{X}-\mathbf{AS}\|_F^2 \quad \text{or} \quad \min_{\mathbf{A}\in\mathbb{R}^{n_1\times r}_{\geq 0}, \mathbf{S}\in\mathbb{R}^{r\times n_2}_{\geq 0}} D(\mathbf{X}\|\mathbf{AS}).^1$$

 $n_1 \times r$ 

#### non-convex optimization problems

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.





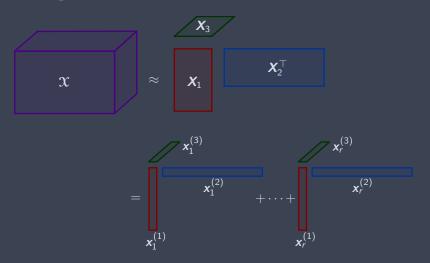
 $a_1$ 

 $a_r$ 

Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.

Introduction		
00000		

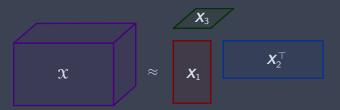
#### » Nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)



Carroll, J. Douglas, and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." Psychometrika 35.3 (1970): 283-319. Harshman, Richard A. "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis." (1970): 1-84.

Introduction		
0000		

#### » Nonnegative CANDECOMP/PARAFAC (CP) decomposition (NCPD)



 $\begin{tabular}{ll} \begin{tabular}{ll} & \begin{tabular}{ll} \begin{tabular}{ll} $ & \begin{tabula$ 

Carroll, J. Douglas, and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." Psychometrika 35.3 (1970): 283-319. Harshman, Richard A. "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis." (1970): 1-84.

Introduction

Hierarchical Models

Experiments

Conclusions

# » Hierarchical NMF

Model: Sequentially factorize

Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006). 947.

Introduction

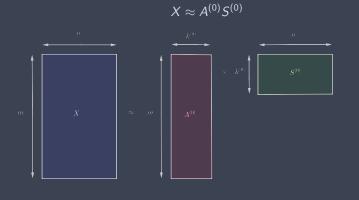
Hierarchical Models

Experiments

Conclusions

# » Hierarchical NMF

### Model: Sequentially factorize



Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006). 947.

Introduction

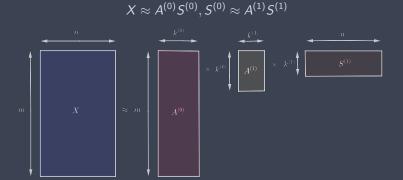
Hierarchical Models

Experiments

Conclusions

# » Hierarchical NMF

Model: Sequentially factorize



Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

Introduction

Hierarchical Models

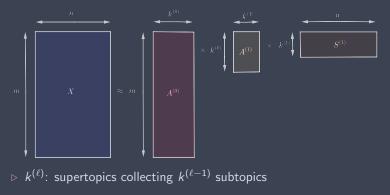
Experiments

Conclusions

# » Hierarchical NMF

Model: Sequentially factorize

 $X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, ..., S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})} \approx A^{(\mathcal{L}$ 



Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

Introduction ○○○●○ Hierarchical Models

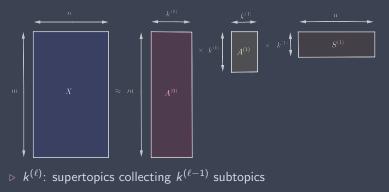
Experiments

Conclusions

# » Hierarchical NMF

Model: Sequentially factorize

 $X \approx A^{(0)}S^{(0)}, S^{(0)} \approx A^{(1)}S^{(1)}, S^{(1)} \approx A^{(2)}S^{(2)}, ..., S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L}-1)} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}S^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})}, S^{(\mathcal{L})} \approx A^{(\mathcal{L})} \approx A^{(\mathcal{L}$ 



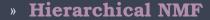
 $\,\,\triangleright\,\,$  provides relationship between data matrix modes and  $k^{(\ell)}$  topics

Cichocki, Andrzej, and Rafal Zdunek. "Multilayer nonnegative matrix factorisation." ELECTRONICS LETTERS-IEE 42.16 (2006): 947.

Introduction ○○○○● Hierarchical Models

Experiments

Conclusions







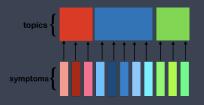
**Iotivation** 

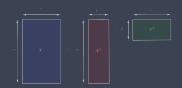
Introduction ○○○○● Hierarchical Models

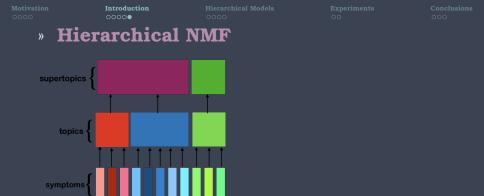
Experiments

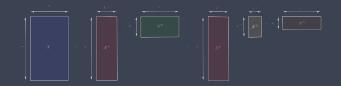
Conclusions

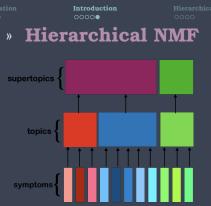
# » Hierarchical NMF







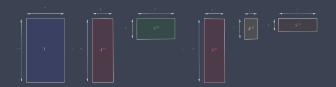


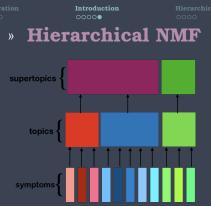


**Experiments** 

Conclusions

## elucidates the hierarchical relationships of learned topics



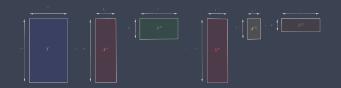




Conclusions

## elucidates the hierarchical relationships of learned topics

 no need to choose a fixed model rank (number of topics)



Introduction

Hierarchical Models

Experiment:

Conclusions

## **Hierarchical Models**

		Hierarchical Models		
		0000		
» F	<b>lierarchical</b>	Tensor Dec	ompositio	ns

How do we generalize HNMF to a higher-order tensor model?

	Hierarchical Models	
	0000	

#### How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.

	Hierarchical Models	
	0000	

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.
- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

	Hierarchical Models	
	0000	

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.
- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

	Hierarchical Models	
	0000	

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.
- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

\* Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

	Hierarchical Models	
	0000	

#### » Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.
- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

\* Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021. Not a single hierarchical relationship, good training method.

	Hierarchical Models	
	0000	

#### » Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.
- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

\* Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

Not a single hierarchical relationship, good training method. \* Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

	Hierarchical Models	
	0000	

#### » Hierarchical Tensor Decompositions

How do we generalize HNMF to a higher-order tensor model?

- Vasilescu, M. Alex O., and Eric Kim. "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors." arXiv preprint arXiv:1911.04180 (2019).
- Song, Le, et al. "Hierarchical tensor decomposition of latent tree graphical models." International Conference on Machine Learning. PMLR, 2013.
- Grasedyck, Lars. "Hierarchical singular value decomposition of tensors." SIAM Journal on Matrix Analysis and Applications 31.4 (2010): 2029-2054.
- Cichocki, Andrzej, Rafal Zdunek, and Shun-ichi Amari. "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization." International Conference on Independent Component Analysis and Signal Separation. Springer, Berlin, Heidelberg, 2007.

Results depend upon hyperparameter choice (mode).

\* Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "Neural nonnegative CP decomposition for hierarchical tensor analysis." 2021 55th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2021.

Not a single hierarchical relationship, good training method. \* Vendrow, Joshua, Jamie Haddock, and Deanna Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

\*\* Single hierarchical relationship, naive training method. \*\*

Introduction

Hierarchical Models

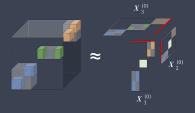
Experiments

Conclusions

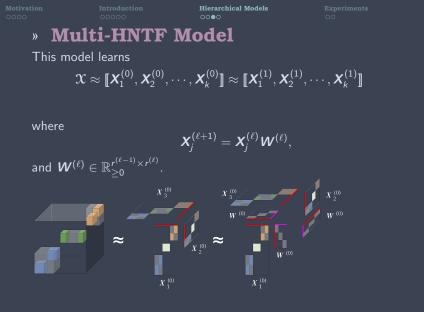
#### » Multi-HNTF Model

#### This model learns

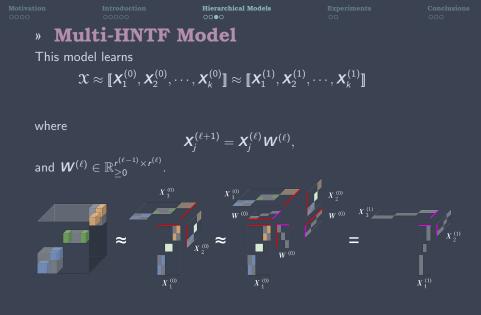
 $\mathfrak{X} \approx \llbracket \boldsymbol{X}_1^{(0)}, \boldsymbol{X}_2^{(0)}, \cdots, \boldsymbol{X}_k^{(0)} 
rbracket$ 



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

Introduction

Hierarchical Models

Experiments

Conclusions

#### » Multi-HNTF Model

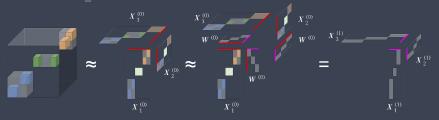
This model learns

$$\begin{aligned} \boldsymbol{\mathfrak{X}} &\approx \llbracket \boldsymbol{X}_1^{(0)}, \boldsymbol{X}_2^{(0)}, \cdots, \boldsymbol{X}_k^{(0)} \rrbracket \approx \llbracket \boldsymbol{X}_1^{(1)}, \boldsymbol{X}_2^{(1)}, \cdots, \boldsymbol{X}_k^{(1)} \rrbracket \approx \cdots \\ &\approx \llbracket \boldsymbol{X}_1^{(\mathcal{L}-1)}, \boldsymbol{X}_2^{(\mathcal{L}-1)}, \cdots, \boldsymbol{X}_k^{(\mathcal{L}-1)} \rrbracket \end{aligned}$$

where

$$oldsymbol{X}_j^{(\ell+1)} = oldsymbol{X}_j^{(\ell)} oldsymbol{W}^{(\ell)},$$

and 
$$\boldsymbol{W}^{(\ell)} \in \mathbb{R}_{>0}^{r^{(\ell-1)} \times r^{(\ell)}}$$



Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

**Iotivation** 

**ntroduction** 

Hierarchical Models

Experiments

Conclusions

### » Multi-HNTF Model

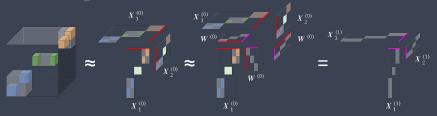
This model learns

$$\begin{aligned} \boldsymbol{\mathfrak{X}} &\approx \llbracket \boldsymbol{X}_1^{(0)}, \boldsymbol{X}_2^{(0)}, \cdots, \boldsymbol{X}_k^{(0)} \rrbracket \approx \llbracket \boldsymbol{X}_1^{(1)}, \boldsymbol{X}_2^{(1)}, \cdots, \boldsymbol{X}_k^{(1)} \rrbracket \approx \cdots \\ &\approx \llbracket \boldsymbol{X}_1^{(\mathcal{L}-1)}, \boldsymbol{X}_2^{(\mathcal{L}-1)}, \cdots, \boldsymbol{X}_k^{(\mathcal{L}-1)} \rrbracket \end{aligned}$$

where

$$oldsymbol{X}_j^{(\ell+1)} = oldsymbol{X}_j^{(\ell)} oldsymbol{W}^{(\ell)},$$

and 
$$\boldsymbol{W}^{(\ell)} \in \mathbb{R}_{>0}^{r^{(\ell-1)} \times r^{(\ell)}}$$



#### A single hierarchical relationship for all modes!

Vendrow, H., Needell. "A Generalized Hierarchical Nonnegative Tensor Decomposition." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.

**tivation Introduction** 

Hierarchical Models

Experiments

Conclusions

# » Training Process

1: procedure MULTI-HNTF(X)

2: 
$$\{\boldsymbol{X}_{i}^{(0)}\}_{i=1}^{k} \leftarrow \mathsf{NCPD}(\boldsymbol{\mathcal{X}}, r_{0})$$
3: 
$$\mathbf{for} \ \ell = 0 \dots \mathcal{L} \ \mathbf{do}$$
4: 
$$\boldsymbol{W}^{(\ell)} \leftarrow \operatorname{argmin}_{\boldsymbol{W} \in \mathbb{R}_{+}^{r_{\ell} \times r_{\ell+1}}} \|\boldsymbol{\mathcal{X}} - [\![\boldsymbol{X}_{1}^{(\ell)} \boldsymbol{W}, \dots, \boldsymbol{X}_{k}^{(\ell)} \boldsymbol{W}]\!] \|$$
5: 
$$\mathbf{for} \ i = 0 \dots k \ \mathbf{do}$$
6: 
$$\boldsymbol{X}_{i}^{(\ell+1)} = \boldsymbol{X}_{i}^{(\ell)} \boldsymbol{W}^{(\ell)}$$

Motivation Introduction Hierarchical Models

Experiments

Conclusions

## » Training Process

## 1: procedure MULTI-HNTF( $\mathfrak{X}$ )

2: 
$$\{\boldsymbol{X}_{i}^{(0)}\}_{i=1}^{k} \leftarrow \mathsf{NCPD}(\boldsymbol{\mathcal{X}}, r_{0})$$
  
3: for  $\ell = 0 \dots \mathcal{L}$  do  
4:  $\boldsymbol{W}^{(\ell)} \leftarrow \operatorname{argmin}_{\boldsymbol{W} \in \mathbb{R}_{+}^{\ell_{\ell} \times r_{\ell+1}}} \|\boldsymbol{\mathcal{X}} - [\![\boldsymbol{X}_{1}^{(\ell)}\boldsymbol{W}, \dots, \boldsymbol{X}_{k}^{(\ell)}\boldsymbol{W}]\!]\|$   
5: for  $i = 0, \dots k$  do  
6:  $\boldsymbol{X}_{i}^{(\ell+1)} = \boldsymbol{X}_{i}^{(\ell)}\boldsymbol{W}^{(\ell)}$ 

▷ Can be approximated via NMF method on each mode with averaging of learned *W* matrix across modes. 
 Introduction
 Hierarchical Models

 0000
 0000
 0000

**Exp** 00 Conclusions

## » Training Process

#### 1: procedure MULTI-HNTF( $\mathfrak{X}$ ) 2: $\{\mathbf{X}_{i}^{(0)}\}_{i=1}^{k} \leftarrow \mathsf{NCPD}(\mathfrak{X}, r_{0})$

3: for  $\ell = 0 \dots \mathcal{L}$  do

4: 
$$\boldsymbol{W}^{(\ell)} \leftarrow \operatorname{argmin}_{\boldsymbol{W} \in \mathbb{R}^{\ell \ell imes \ell_{\ell+1}}_+} \| \mathfrak{X} - [\![\boldsymbol{X}_1^{(\ell)} \boldsymbol{W}, \dots, \boldsymbol{X}_k^{(\ell)} \boldsymbol{W}]\!] \|$$

5: **for** 
$$i = 0...k$$
 **do**  
6:  $X^{(\ell+1)} = X^{(\ell)} W^{(\ell)}$ 

- Can be approximated via NMF method on each mode with averaging of learned *W* matrix across modes.
- $\,\triangleright\,$  Could/should also be trained in a neural network framework.

Introduction

Hierarchical Models

Experiment:

Conclusions

#### Experiments

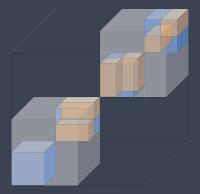
Introduction

Hierarchical Models

Experiments

Conclusions

#### » Synthetic Tensor



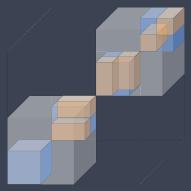
The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure. ation

ntroduction

Hierarchical Models

Experiments ○● Conclusions

#### » Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

#### Relative reconstruction error.

Method	$r_0 = 7$	<i>r</i> <sub>1</sub> = 4	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747

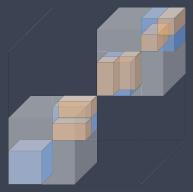
Introduction

Hierarchical Models

Experiments

Conclusions

#### » Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.



Projections of tensor approximation at each layer of Multi-HNTF.

#### Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747
Neural HNCPD [Vendrow, et. al.]	0.454	0.508	0.714

Introduction

Hierarchical Models

Experiment:

Conclusions ●○○

#### Conclusions

Introduction

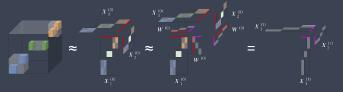
Hierarchical Models

Experiments

Conclusions ○●○

#### » Conclusions

▷ Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.



introduction

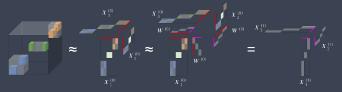
Hierarchical Models

Experiments

Conclusions ○●○

#### » Conclusions

Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.



b Model can be trained by your favorite NMF method with an additional projection step.

ntroduction

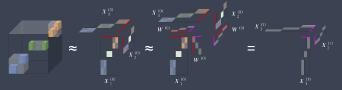
Hierarchical Models

Experiments

Conclusions ○●○

#### » Conclusions

 Multi-HNTF is a hierarchical tensor decomposition model that generalizes hierarchical NMF.



- b Model can be trained by your favorite NMF method with an additional projection step.
- Develop backpropagation framework for Multi-HNTF and first layer NCPD.

Introduction

Hierarchical Models

Experiments

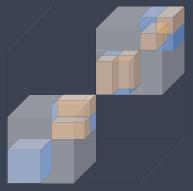
Conclusions ○○●

#### » Thanks for listening!

Questions?

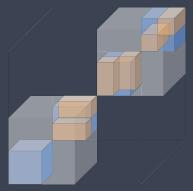
- M. Gao, J. Haddock, D. Molitor, D. Needell, E. Sadovnik, T. Will, and R. Zhang. Neural nonnegative matrix factorization for hierarchical multilayer topic modeling. In <u>Proc.</u> <u>Interational Workshop on Computational Advances in Multi-Sensor Adaptive Processing</u>, 2019.
- [2] J. Vendrow, J. Haddock, and D. Needell. Neural nonnegative CP decomposition for hierarchical tensor analysis. In <u>Asilomar Conf. on Signals, Systems, Computers (ACSSC)</u>, 2021.
- J. Vendrow, J. Haddock, and D. Needell. A generalized hierarchical nonnegative tensor decomposition. In <u>Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)</u>, 2022.
- [4] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. <u>Nature</u>, 401(6755):788–791, 1999.
- J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. <u>Psychometrika</u>, 35(3):283–319, 1970.
- [6] Richard A Harshman et al. Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970.
- [7] Andrzej Cichocki and Rafal Zdunek. Multilayer nonnegative matrix factorisation. <u>Electronics</u> Letters, 42(16):947–948, 2006.

## » Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

### » Synthetic Tensor

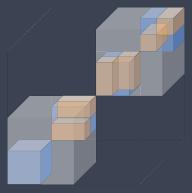


The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.

#### Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Neural HNCPD [Vendrow, et. al.]	0.454	0.508	0.714
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747

### » Synthetic Tensor



The table lists relative reconstruction errors on the tensor on the left for models learned with 7-4-2 topic structure. Below, we visualize the Multi-HNTF learned approximations for a synthetic tensor with 7-3 topic structure.



Projections of tensor approximation at each layer of Multi-HNTF.

#### Relative reconstruction error.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Neural HNCPD [Vendrow, et. al.]	0.454	0.508	0.714
Standard HNCPD [Vendrow, et. al.]	0.454	0.612	0.892
HNTF-1 [Cichocki, et. al.]	0.454	0.576	0.781
HNTF-2 [Cichocki, et. al.]	0.454	0.587	0.765
HNTF-3 [Cichocki, et. al.]	0.454	0.560	0.747