Neural Nonnegative CP Decomposition for Hierarchical Tensor Analysis

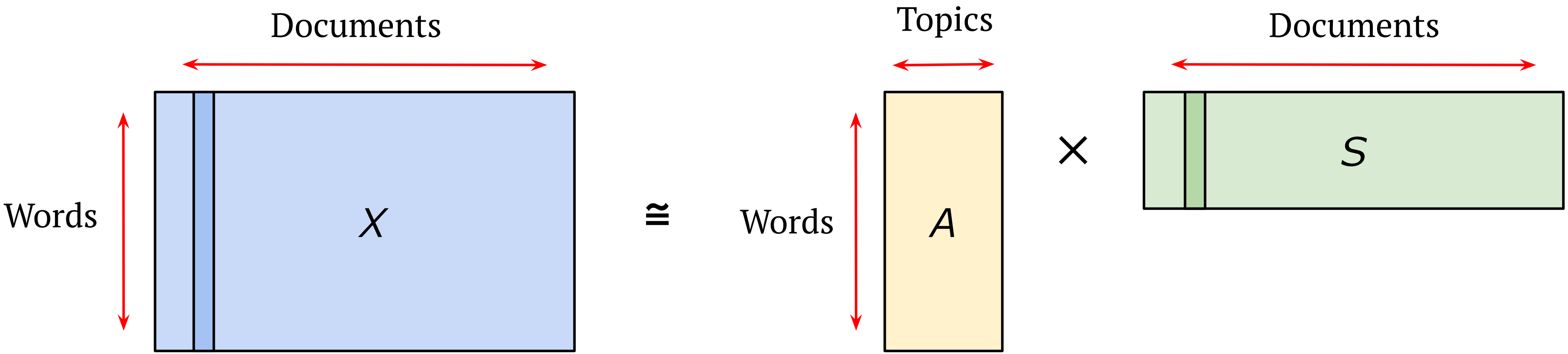
Joshua Vendrow, Jamie Haddock, Deanna Needell
UCLA Mathematics

Background

Nonnegative Matrix Factorization (NMF)

Given a nonnegative matrix X ($X \geq 0$), compute nonnegative A and S such that

$$X \approx AS$$

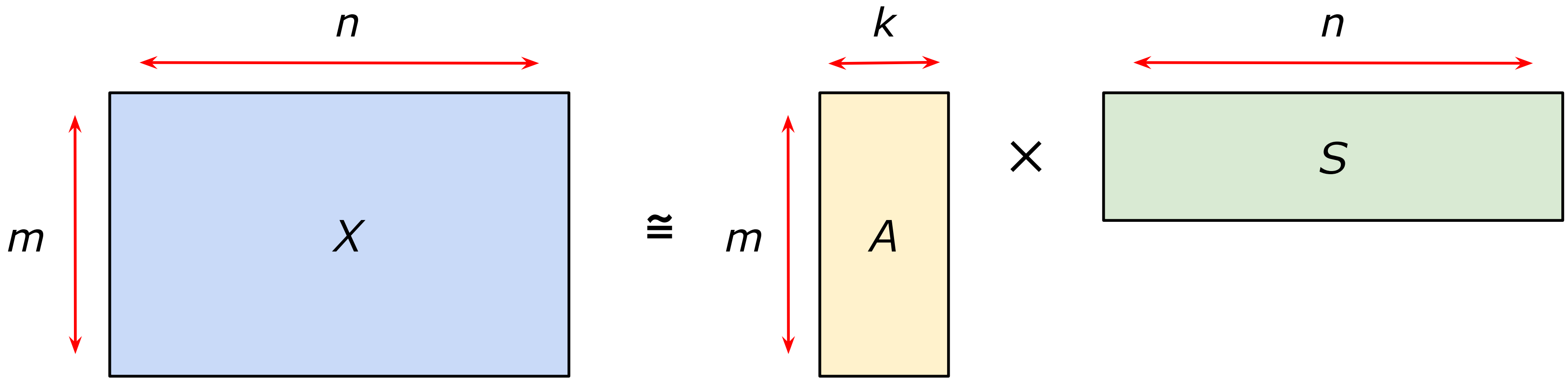


D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, vol. 401, no. 6755, pp. 788, 1999.

Nonnegative Matrix Factorization (NMF)

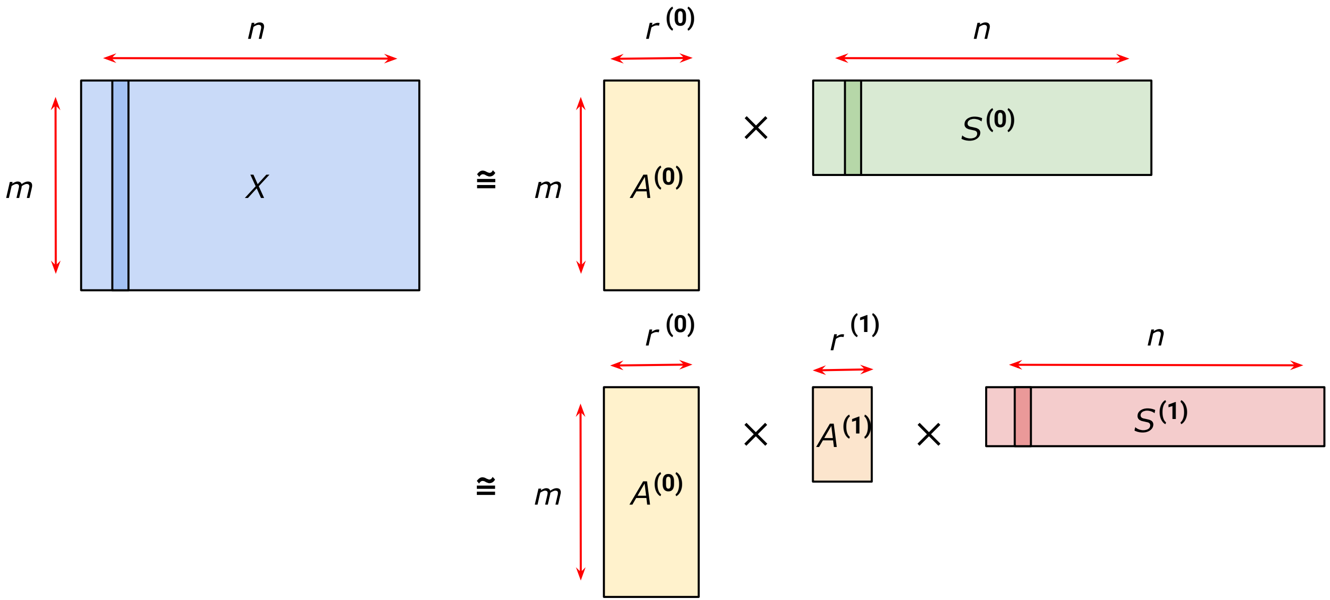
Formulated as the optimization task:

$$\underset{\boldsymbol{A}\geq 0, \boldsymbol{S}\geq 0}{\arg\min} \|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{S}\|_F^2$$



D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, vol. 401, no. 6755, pp. 788, 1999.

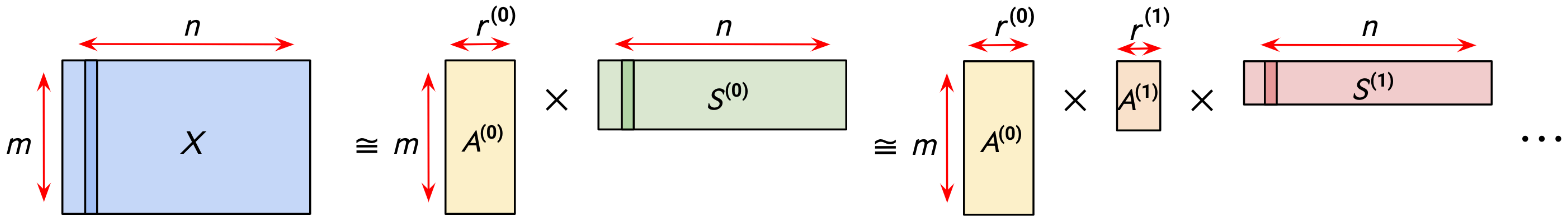
Hierarchical NMF

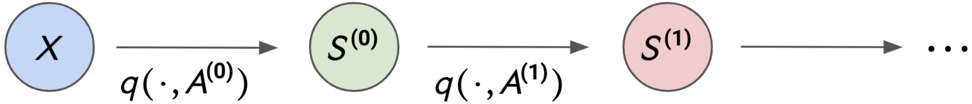


Neural NMF

Regard the A matrices as weights, and determine S matrices from A matrices, and define

$$q(X, A) := \operatorname{argmin}_{S \geq 0} \| X - AS \|_F^2$$



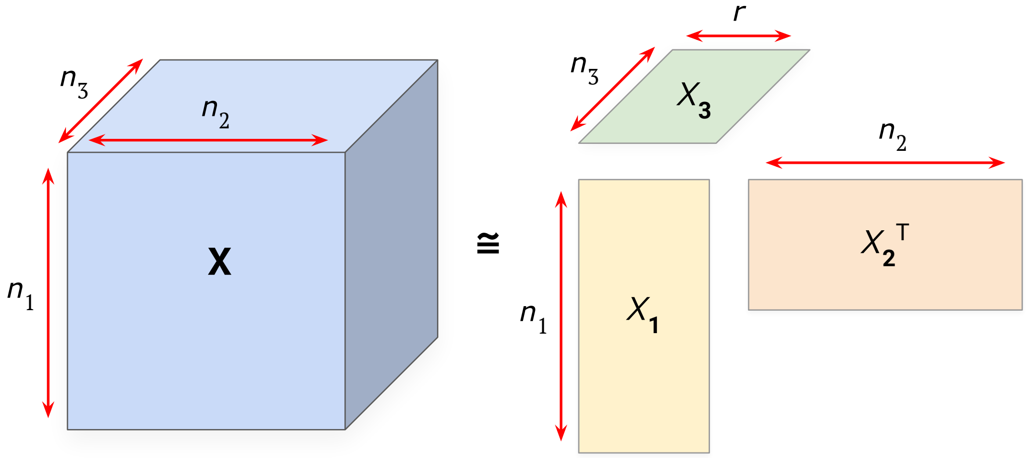


M. Gao et al. “Neural nonnegative matrix factorization for hierarchical multilayer topic modeling,” in Proc. Int. Workshop on Comp. Adv. in Multi-Sensor Adaptive Process., 2019.

Nonnegative CP Decomposition (NCPD)

Given a nonnegative tensor **X** (**X** $\geq$ 0), compute nonnegative X_1, X_2 , … X_k such that

$$\mathbf{X} \approx [\![\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_k]\!] \equiv \sum_{j=1}^{r} \boldsymbol{x}_j^{(1)} \otimes \boldsymbol{x}_j^{(2)} \otimes \cdots \otimes \boldsymbol{x}_j^{(k)}$$

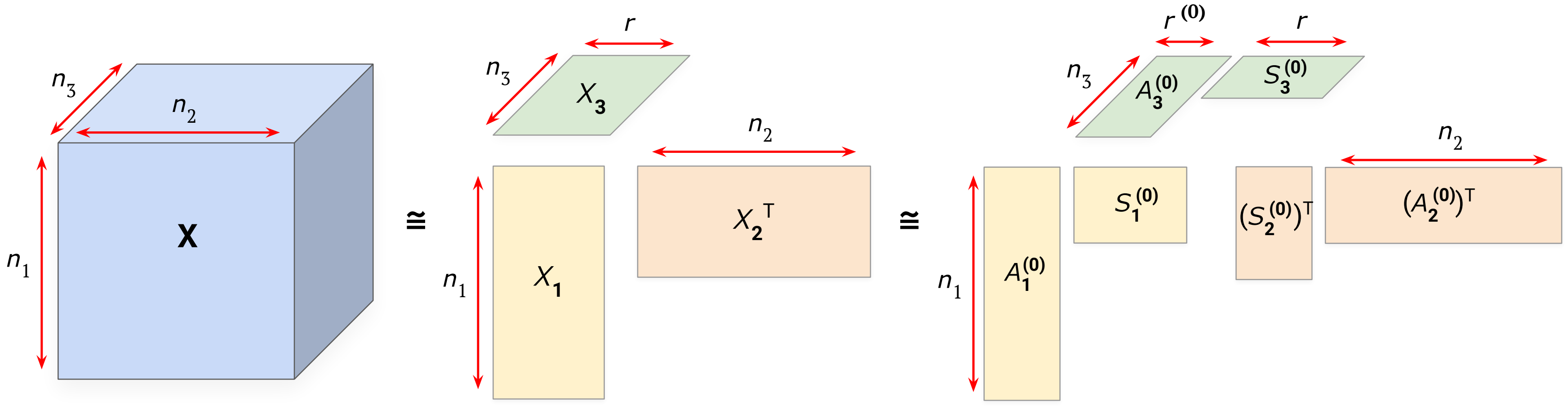


Our Contributions

Hierarchical Nonnegative CP Decomposition (HNCPD)

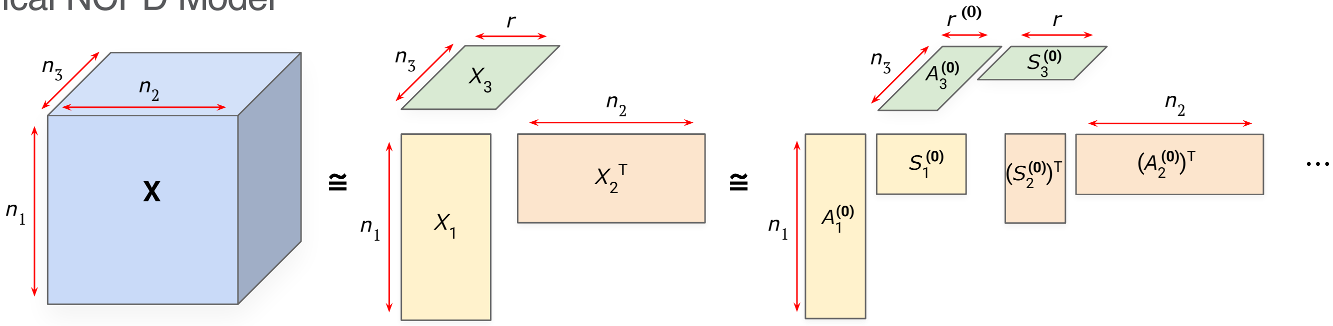
Apply a hierarchical NMF onto each factor matrix:

$$\boldsymbol{X}_i \approx \widetilde{\boldsymbol{X}}_i \equiv \boldsymbol{A}_i^{(0)} \boldsymbol{A}_i^{(1)} \ldots \boldsymbol{A}_i^{(\ell-2)} \boldsymbol{S}_i^{(\ell-2)}$$

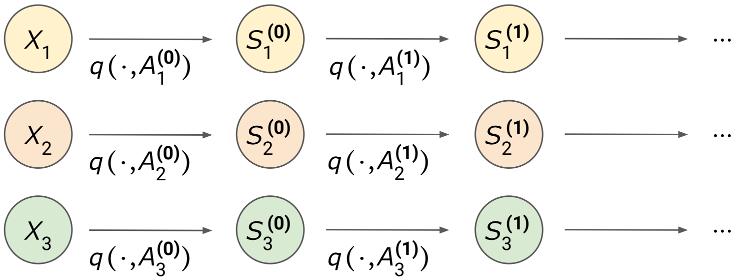


Neural Nonnegative CP Decomposition (Neural NCPD)

Hierarchical NCPD Model



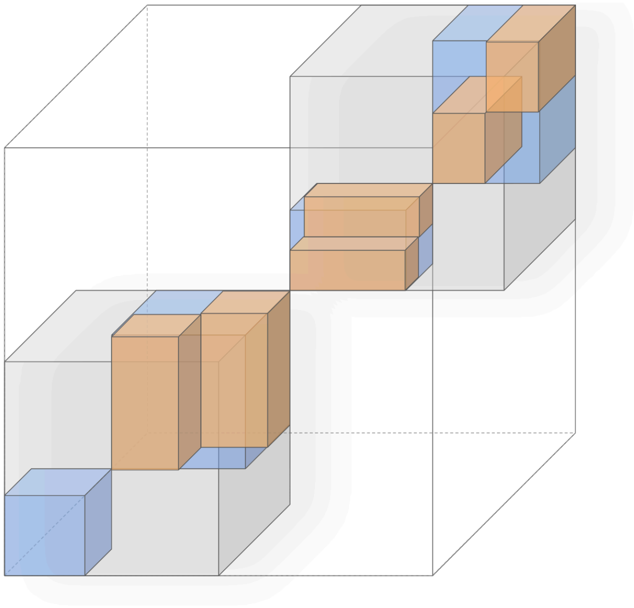
Neural NMF Branches



Experiments

Synthetic Tensor Data Set Experiments

- CP rank seven tensor of size 40 × 40 × 40
- Overlapping and non-overlapping blocks of varying size and intensity to form a hierarchical structure
- Positive random noise added to each entry



Synthetic Tensor Data Set Experiments

TABLE I: Topic modeling loss and relative reconstruction loss, C_{rel}, on the synthetic dataset for Neural NCPD, Standard HNCPD, HNTF, Neural NMF, and Standard HNMF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor, and for matrix methods we report results for flattening along each mode of the tensor.

		Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
Method	Mode	$7-2$	$7-4$	$4-2$	$7-2$	$7-4$	$4-2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$
Neural HNCPD		**0.043**	0.042	**0.042**	**0.087**	**0.087**	**0.081**	0.119	0.252	0.563	**0.454**	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	**0.454**	0.612	0.892
HNTF [12]	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	**0.454**	0.576	0.781
	2	0.087	**0.040**	0.101	0.090	0.116	0.142	0.119	0.309	0.665	**0.454**	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.164	0.119	0.417	0.713	**0.454**	0.560	0.747
Neural NMF [9]	1	0.154	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	**0.693**
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	**0.118**	0.235	**0.558**	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	**0.118**	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	**0.118**	**0.233**	0.563	0.472	**0.503**	0.717

Synthetic Tensor Data Set Experiments

TABLE I: Topic modeling loss and relative reconstruction loss, C_{rel}, on the synthetic dataset for Neural NCPD, Standard HNCPD, HNTF, Neural NMF, and Standard HNMF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor, and for matrix methods we report results for flattening along each mode of the tensor.

		Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
Method	Mode	$7-2$	$7-4$	$4-2$	$7-2$	$7-4$	$4-2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$
Neural HNCPD		**0.043**	0.042	**0.042**	**0.087**	**0.087**	**0.081**	0.119	0.252	0.563	**0.454**	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	**0.454**	0.612	0.892
HNTF [12]	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	**0.454**	0.576	0.781
	2	0.087	**0.040**	0.101	0.090	0.116	0.142	0.119	0.309	0.665	**0.454**	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.164	0.119	0.417	0.713	**0.454**	0.560	0.747
Neural NMF [9]	1	0.154	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	**0.693**
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	**0.118**	0.235	**0.558**	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	**0.118**	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	**0.118**	**0.233**	0.563	0.472	**0.503**	0.717

Synthetic Tensor Data Set Experiments

TABLE I: Topic modeling loss and relative reconstruction loss, C_{rel}, on the synthetic dataset for Neural NCPD, Standard HNCPD, HNTF, Neural NMF, and Standard HNMF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor, and for matrix methods we report results for flattening along each mode of the tensor.

		Topic Modeling Loss						Relative Reconstruction Loss					
		$\sigma^2 = 0.1$			$\sigma^2 = 0.4$			$\sigma^2 = 0.1$			$\sigma^2 = 0.4$		
Method	Mode	$7-2$	$7-4$	$4-2$	$7-2$	$7-4$	$4-2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$	$r=7$	$r^{(0)}=4$	$r^{(1)}=2$
Neural HNCPD		**0.043**	0.042	**0.042**	**0.087**	**0.087**	**0.081**	0.119	0.252	0.563	**0.454**	0.508	0.714
Standard HNCPD		0.106	0.101	0.189	0.145	0.193	0.204	0.119	0.494	0.828	**0.454**	0.612	0.892
HNTF [12]	1	0.163	0.236	0.182	0.171	0.144	0.170	0.119	0.502	0.795	**0.454**	0.576	0.781
	2	0.087	**0.040**	0.101	0.090	0.116	0.142	0.119	0.309	0.665	**0.454**	0.587	0.765
	3	0.078	0.122	0.106	0.084	0.111	0.164	0.119	0.417	0.713	**0.454**	0.560	0.747
Neural NMF [9]	1	0.154	0.192	0.105	0.169	0.219	0.127	0.146	0.268	0.593	0.478	0.521	0.705
	2	0.075	0.244	0.146	0.153	0.190	0.160	0.141	0.289	0.585	0.475	0.513	0.710
	3	0.119	0.164	0.110	0.158	0.197	0.140	0.151	0.236	0.576	0.477	0.512	**0.693**
Standard HNMF	1	0.098	0.182	0.052	0.164	0.219	0.139	**0.118**	0.235	**0.558**	0.472	0.524	0.707
	2	0.080	0.199	0.090	0.151	0.213	0.088	**0.118**	0.245	0.566	0.472	0.505	0.709
	3	0.060	0.165	0.085	0.137	0.193	0.114	**0.118**	**0.233**	0.563	0.472	**0.503**	0.717

Synthetic Tensor Data Set Experiments

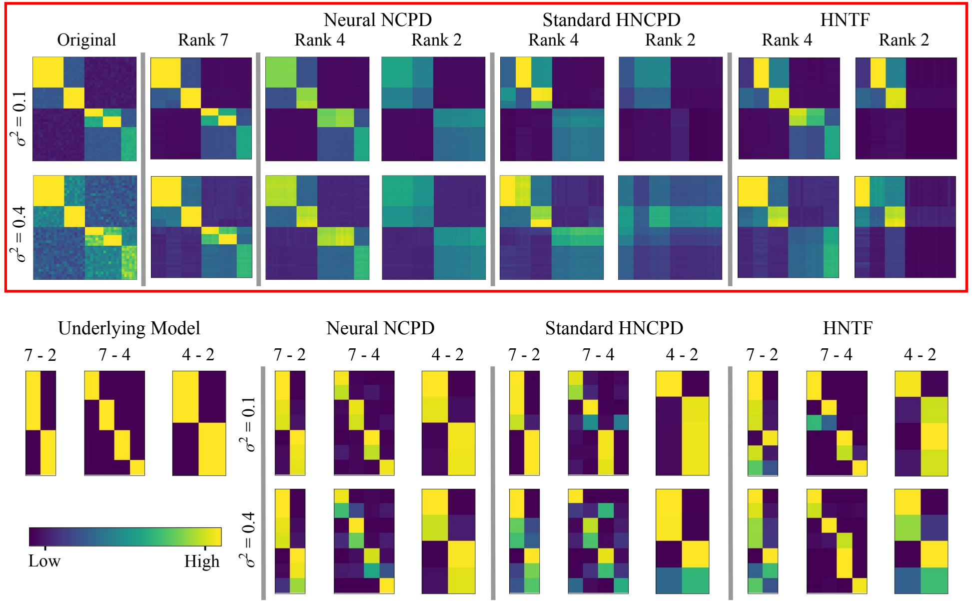


Fig. 2: (Top left) Data tensor $\mathbf{X}$ with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPD, and HNTF approximations of $\mathbf{X}$. (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPD, and HNTF, with rows and columns permuted to minimize topic error.

Synthetic Tensor Data Set Experiments

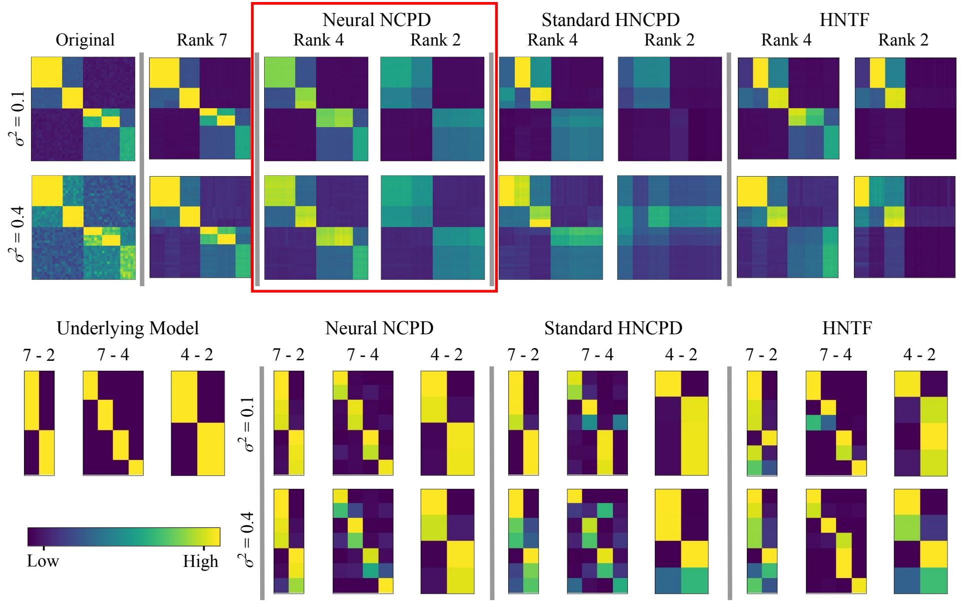


Fig. 2: (Top left) Data tensor $\mathbf{X}$ with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPD, and HNTF approximations of $\mathbf{X}$. (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPD, and HNTF, with rows and columns permuted to minimize topic error.

Synthetic Tensor Data Set Experiments

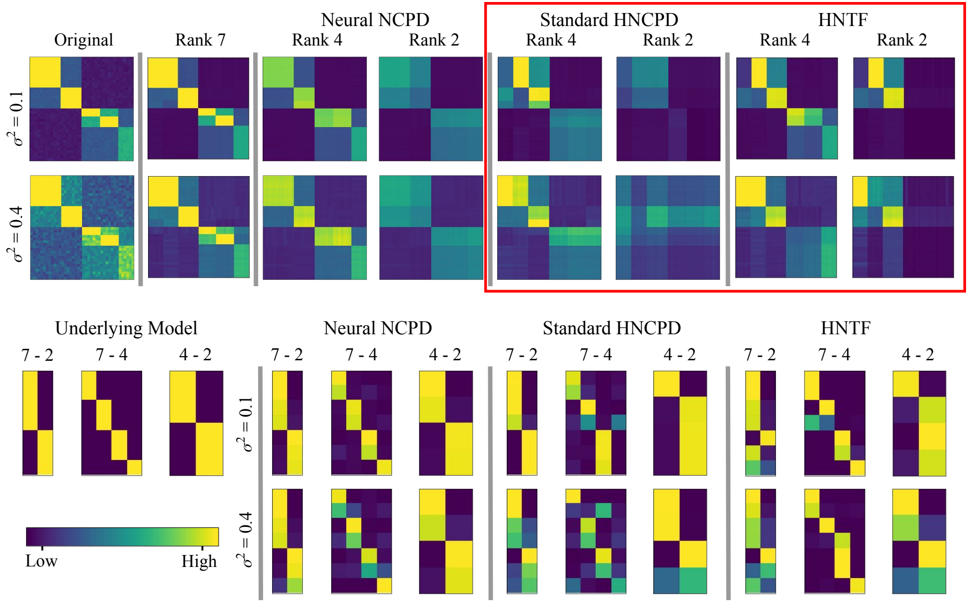


Fig. 2: (Top left) Data tensor $\mathbf{X}$ with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPD, and HNTF approximations of $\mathbf{X}$. (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPD, and HNTF, with rows and columns permuted to minimize topic error.

Synthetic Tensor Data Set Experiments

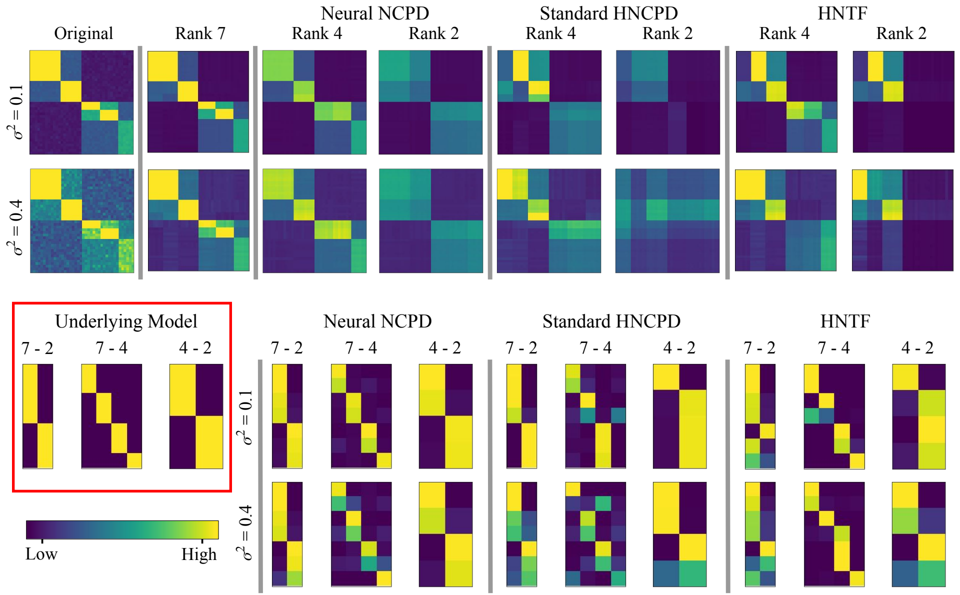


Fig. 2: (Top left) Data tensor $\mathbf{X}$ with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPD, and HNTF approximations of $\mathbf{X}$. (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPD, and HNTF, with rows and columns permuted to minimize topic error.

Synthetic Tensor Data Set Experiments

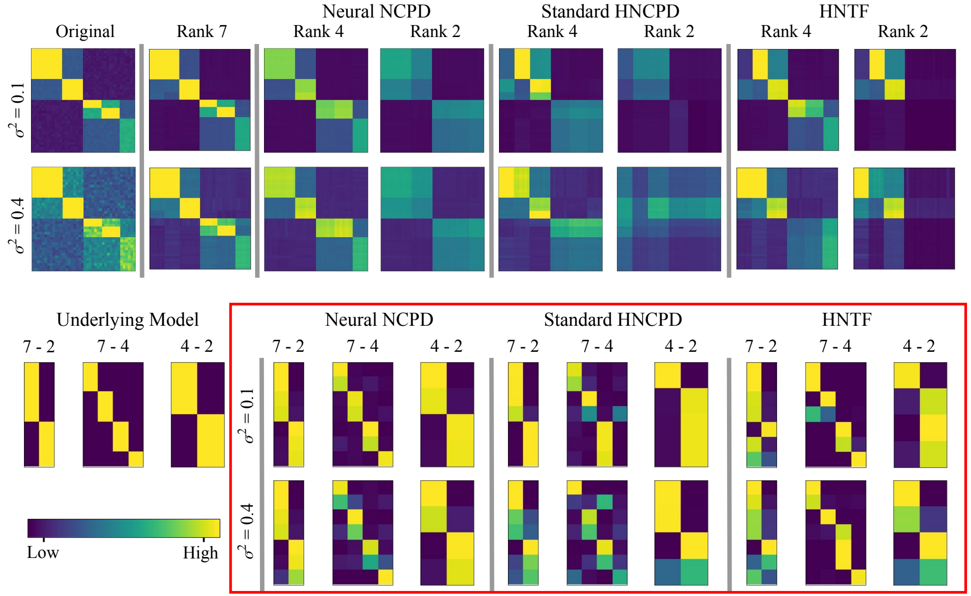
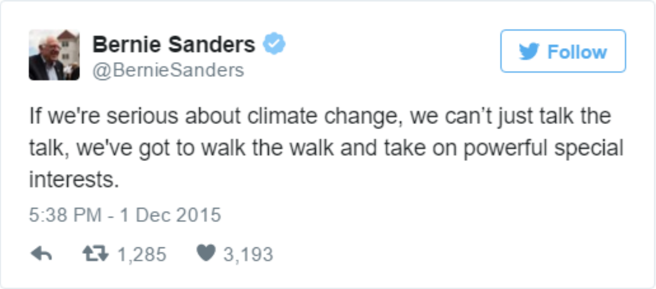


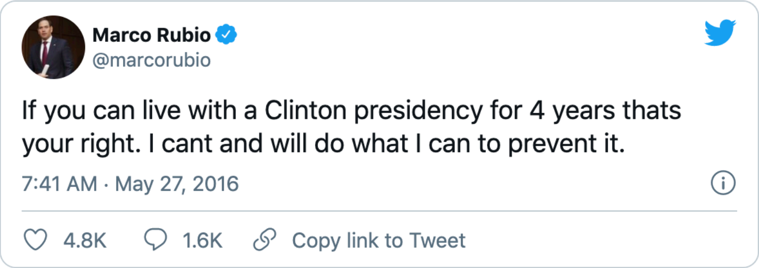
Fig. 2: (Top left) Data tensor $\mathbf{X}$ with two levels of noise. (Top right) ranks 7, 5, and 3 Neural NCPD, Standard HNCPD, and HNTF approximations of $\mathbf{X}$. (Bottom left) Underlying topic modelling matrix. (Bottom right) topic modelling matrices discovered by Neural NCPD, Standard HNCPD, and HNTF, with rows and columns permuted to minimize topic error.

Twitter Political Data Set

- A data set of tweets sent by political candidates during the 2016 election season
- We subset the tweets from eight politicians, four Republicans and four Democrats:

  Hillary Clinton, Tim Kaine, Martin O'Malley, Bernie Sanders, Ted Cruz, John Kasich, Marco Rubio, and Donald Trump.





J. Littman, L. Wrubel, and D. Kerchner, “2016 United States Presidential Election Tweet Ids,” 2016.

Twitter Political Data Set

- We use a bag-of-words (12,721 words in corpus) representation of all tweets made by a candidate
- We bin all tweets made by a candidate each 30 days (from Feb to Dec 2016)
- Resulting tensor is size

$$\underset{\#\text{ Candidates}}{8} \times \underset{\#\text{ Months}}{10} \times \underset{\#\text{ Words}}{12721}$$

J. Littman, L. Wrubel, and D. Kerchner, “2016 United States Presidential Election Tweet Ids,” 2016.

Neural NCPD in Twitter Political Data Set

Rank 8 Topics

Topic 1	Topic 2	Topic 3	Topic 4
trump	senate	martinomalley	berniesanders
hillary	florida	hillaryclinton	people
donald	zika	realdonaldtrump	bernie
president	venezuela	campaigning	must
timkaine	nicolasmaduro	maryland	change
Topic 5	**Topic 6**	**Topic 7**	**Topic 8**
tedcruz	johnkasich	marcorubio	crooked
cruz	kasich	teammarco	hillary
ted	ohio	vote	thank
internet	john	flsen	great
choosecruz	gov	click	clinton

Rank 4 Topics

Topic 1	Topic 2
marcorubio	trump
teammarco	hillary
vote	donald
flsen	people
click	vote
Topic 3	**Topic 4**
tedcruz	senate
cruz	florida
ted	zika
johnkasich	venezuela
kasich	nicolasmaduro

Rank 2 Topics

Topic 1
trump
hillary
vote
people
donald
Topic 2
tedcruz
cruz
ted
johnkasich
kasich

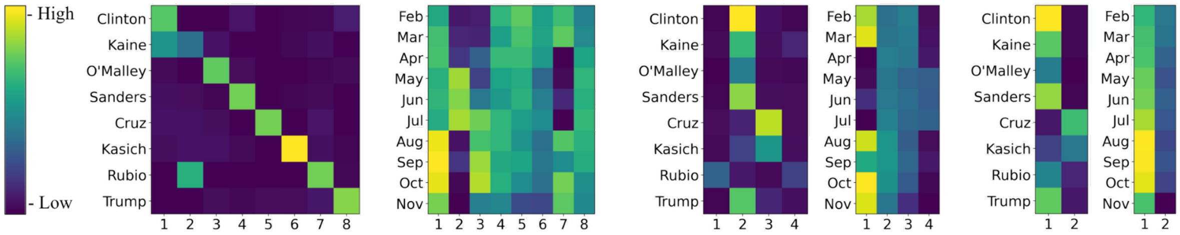


Fig. 3: A three-layer Neural NCPD on the Twitter dataset at ranks $r = 8$, $r^{(0)} = 4$ and $r^{(1)} = 2$. At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

Neural NCPD in Twitter Political Data Set

Rank 8 Topics

Topic 1	Topic 2	Topic 3	Topic 4
trump	senate	martinomalley	berniesanders
hillary	florida	hillaryclinton	people
donald	zika	realdonaldtrump	bernie
president	venezuela	campaigning	must
timkaine	nicolasmaduro	maryland	change
Topic 5	**Topic 6**	**Topic 7**	**Topic 8**
tedcruz	johnkasich	marcorubio	crooked
cruz	kasich	teammarco	hillary
ted	ohio	vote	thank
internet	john	flsen	great
choosecruz	gov	click	clinton

Rank 4 Topics

Topic 1	Topic 2
marcorubio	trump
teammarco	hillary
vote	donald
flsen	people
click	vote
Topic 3	**Topic 4**
tedcruz	senate
cruz	florida
ted	zika
johnkasich	venezuela
kasich	nicolasmaduro

Rank 2 Topics

Topic 1
trump
hillary
vote
people
donald
Topic 2
tedcruz
cruz
ted
johnkasich
kasich

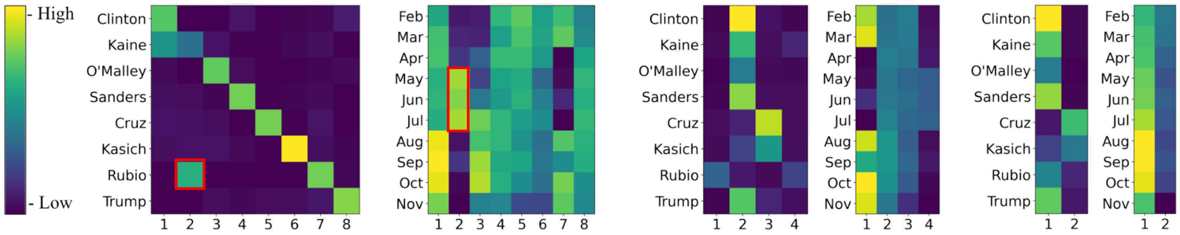


Fig. 3: A three-layer Neural NCPD on the Twitter dataset at ranks $r = 8$, $r^{(0)} = 4$ and $r^{(1)} = 2$. At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

Neural NCPD in Twitter Political Data Set

Rank 8 Topics

Topic 1	Topic 2	Topic 3	Topic 4
trump	senate	martinomalley	berniesanders
hillary	florida	hillaryclinton	people
donald	zika	realdonaldtrump	bernie
president	venezuela	campaigning	must
timkaine	nicolasmaduro	maryland	change
Topic 5	**Topic 6**	**Topic 7**	**Topic 8**
tedcruz	johnkasich	marcorubio	crooked
cruz	kasich	teammarco	hillary
ted	ohio	vote	thank
internet	john	flsen	great
choosecruz	gov	click	clinton

Rank 4 Topics

Topic 1	Topic 2
marcorubio	trump
teammarco	hillary
vote	donald
flsen	people
click	vote
Topic 3	**Topic 4**
tedcruz	senate
cruz	florida
ted	zika
johnkasich	venezuela
kasich	nicolasmaduro

Rank 2 Topics

Topic 1
trump
hillary
vote
people
donald
Topic 2
tedcruz
cruz
ted
johnkasich
kasich

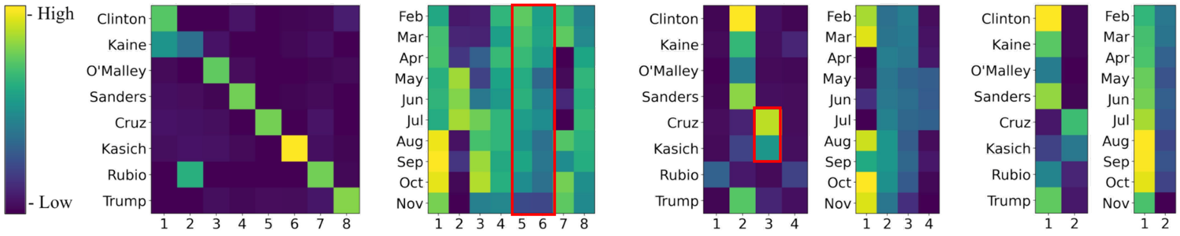


Fig. 3: A three-layer Neural NCPD on the Twitter dataset at ranks $r = 8$, $r^{(0)} = 4$ and $r^{(1)} = 2$. At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

Standard NCPD in Twitter Political Data Set

Clinton and Kasich in the same topics

Rank 4 Topics

Topic 1	Topic 2
trump	berniesanders
hillary	people
johnkasich	bernie
ohio	mustt
kasich	vote

Topic 3	Topic 4
crooked	tedcruz
hillary	cruz
thank	ted
marcorubio	internet
great	choosecruz

Rank 2 Topics

Topic 1
trump
hillary
vote
people
donald

Topic 2
tedcruz
cruz
ted
johnkasich
kasich

Clinton and Kasich in separate topics

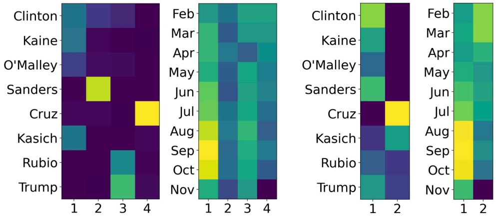


Fig. 5: Ranks 4 and 2 Standard NCPD of the Twitter dataset. At each rank, we display the top five keywords and candidate and temporal mode heatmaps.

Twitter Political Data Set Numerical Results

TABLE II: Relative reconstruction loss, C_{rel}, on the Twitter political dataset for Neural NCPD, Standard NCPD, Standard HNCPD, and HNTF at ranks $r = 8$, $r^{(0)} = 4$, and $r^{(1)} = 8$. For HNTF we display the loss given the three possible arrangements of the tensor.

Method	$r = 8$	$r^{(0)} = 4$	$r^{(1)} = 2$
Neural NCPD	0.834	**0.883**	**0.918**
Standard NCPD	0.834	0.889	0.919
Standard HNCPD	0.834	0.931	0.950
HNTF-1 [12]	0.834	0.890	0.927
HNTF-2 [12]	0.834	0.909	0.956
HNTF-3 [12]	0.834	0.895	0.942

Thank You

J. Vendrow, J. Haddock. D. Needell. "Neural Nonnegative CP Decomposition for Hierarchical Tensor Analysis.” Proc. 53rd Asilomar Conf. on Signals, Systems and Computers, 2021.