A GENERALIZED HIERARCHICAL NONNEGATIVE TENSOR DECOMPOSITION

Joshua Vendrow[⋆]*, Jamie Haddock*[†]*, Deanna Needell*[⋆]

[⋆]University of California, Los Angeles
Department of Mathematics
520 Portola Plaza, Los Angeles, CA 90095

[†]Harvey Mudd College
Department of Mathematics
301 Platt Blvd, Claremont, CA 91711

ABSTRACT

Nonnegative matrix factorization (NMF) has found many applications including topic modeling and document analysis. Hierarchical NMF (HNMF) variants are able to learn topics at various levels of granularity and illustrate their hierarchical relationship. Recently, nonnegative tensor factorization (NTF) methods have been applied in a similar fashion in order to handle data sets with complex, multi-modal structure. Hierarchical NTF (HNTF) methods have been proposed, however these methods do not naturally generalize their matrix-based counterparts. Here, we propose a new HNTF model which directly generalizes a HNMF model special case, and provide a supervised extension. We also provide a multiplicative updates training method for this model. Our experimental results show that this model more naturally illuminates the topic hierarchy than previous HNMF and HNTF methods.

***Index Terms*—** hierarchical topic models, nonnegative matrix factorization, nonnegative tensor decomposition

1. INTRODUCTION

The complexity and size of available data continues to grow which in turn leads to an increasing demand for methods to interpret these large data sets. One important task is *topic modelling*, whose goal is to identify latent topics or trends within a set of data. One popular topic modeling approach is nonnegative matrix factorization (NMF), a dimensionality reduction technique that has had great success in the areas of document analysis, clustering, and classification [1, 2]. NMF is generalized to multi-modal data by the Nonnegative CANDECOMP/PARAFAC (CP) Decomposition (NCPD) [3, 4].

In topic modeling, one often wishes to additionally identify hierarchical relationships between topics learned at different granularities. Towards this goal, many hierarchical models have been developed that enforce an approximate linear relationship between subtopics and supertopics (topics collecting multiple subtopics). More specifically, Hierarchical NMF (HNMF) and Hierarchical nonnegative tensor factorization (HNTF) methods have been developed to factorize data sets simultaneously at multiple different granularities with factorizations learned at coarser granularities constrained by the factorizations learned at finer granularities; these sequential factorizations are often viewed as different layers of these models [5, 6, 7, 8, 9, 10, 11].

The most popular models for HNMF and HNTF share a few common issues: (1) these models only use partial information from the factorization at the previous layer in the subsequent layer's factorization; (2) the hierarchical relationships learned often only represent the latent structure in a subset of the modes of the data; and (3) there is no unified model for both matrices and tensors. In order to address these concerns, we develop Multi-HNTF, which provides a unified model for both matrices and tensors, and demonstrates significant improvement in learning latent hierarchical structures in multi-modal data.

1.1. Notation.

We follow the notational conventions of [12]; e.g., tensor $\mathbf{X}$, matrix $\boldsymbol{X}$, vector $\boldsymbol{x}$, and (integer or real) scalar x. We let $\otimes$ denote the vector outer product and adopt the CP decomposition notation

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

where $\boldsymbol{x}_j^{(i)}$ is the jth column of the ith factor matrix $\boldsymbol{X}_i$ [13].

1.2. Nonnegative matrix factorization

NMF is an approach typically applied in unsupervised tasks such as dimensionality-reduction, latent topic modeling, and clustering. Given nonnegative data matrix $\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{m \times n}$ and a user-defined target dimension $r \in \mathbb{N}$ with $r < \min\{m, n\}$, NMF seeks nonnegative factor matrices $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{m \times r}$, often referred to as the *dictionary* or *topic* matrix, and $\boldsymbol{S} \in \mathbb{R}_{\geq 0}^{r \times n}$, often referred to as the *representation* or *coefficient matrix*, such that $\boldsymbol{X} \approx \boldsymbol{AS}$. There are many formulations of this model (see e.g., [14, 1, 15]) but the most popular utilizes the Frobenius norm,

$$\underset{\boldsymbol{A} \geq 0, \boldsymbol{S} \geq 0}{\arg\min} \|\boldsymbol{X} - \boldsymbol{AS}\|_F^2. \quad (2)$$

The authors were partially supported by NSF DMS #2011140 and BIGDATA #1740325. JH is also partially supported by DMS #2211318.

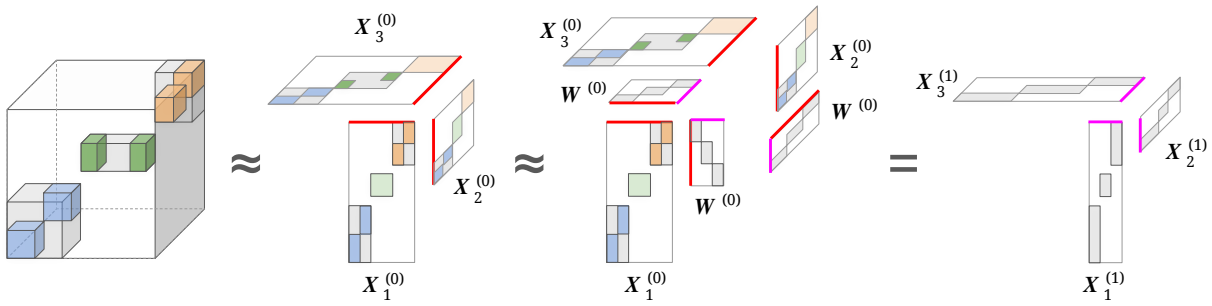

Fig. 1. Visualization of a two layer Multi-HNTF on a tensor with three modes.

Here and throughout, $\boldsymbol{A} \geq 0$ denotes the constraint that $\boldsymbol{A}$ is entry-wise nonnegative. The columns of $\boldsymbol{A}$ are often referred to as *topics*; the NMF approximations to the data (columns of $\boldsymbol{X}$) are additive nonnegative combinations of these topic vectors. This property of NMF approximations yields interpretability since the strength of relationship between a given data point (column of $\boldsymbol{X}$) and the topics of $\boldsymbol{A}$ is given in the coefficient vector (corresponding column of $\boldsymbol{S}$). For this reason, NMF has found popularity in applications such as document clustering [16], and image and audio processing [17], Supervised variants of NMF that jointly factorize both the data matrix $\boldsymbol{X}$ and a matrix of supervision information (e.g., class labels) $\boldsymbol{Y}$ have been proposed [18, 19].

1.3. Hierarchical nonnegative matrix factorization

The most popular hierarchical NMF model decomposes a data matrix $\boldsymbol{X}$ by repeatedly applying NMF to the $\boldsymbol{S}$ matrix output by the previous decomposition, so that each $\boldsymbol{S}^{(i)} \approx \boldsymbol{A}^{(i+1)}\boldsymbol{S}^{(i+1)}$. Given desired ranks $r_0, r_1, ..., r_{\mathcal{L}}$, this process recursively produces the set of factorizations

$$\begin{aligned} \boldsymbol{X} &\approx \boldsymbol{A}^{(0)}\boldsymbol{S}^{(0)}, \\ \boldsymbol{X} &\approx \boldsymbol{A}^{(0)}\boldsymbol{A}^{(1)}\boldsymbol{S}^{(1)}, \\ &\vdots \\ \boldsymbol{X} &\approx \boldsymbol{A}^{(0)}\boldsymbol{A}^{(1)}\cdots\boldsymbol{A}^{(\mathcal{L})}\boldsymbol{S}^{(\mathcal{L})}. \end{aligned} \tag{3}$$

Many complex optimization algorithms have been proposed to improve the quality of the factorization and minimize cascading errors, and are mostly based upon this model for HNMF [5, 6, 20, 21, 7].

1.4. Nonnegative CP Decomposition (NCPD).

The NCPD generalizes NMF to higher-order tensors; specifically, given an order-k tensor $\mathbf{X} \in \mathbb{R}_{\geq 0}^{n_1 \times n_2 \times \cdots \times n_k}$ and a fixed integer r, the approximate NCPD of $\mathbf{X}$ seeks $\boldsymbol{X}_1 \in \mathbb{R}_{\geq 0}^{n_1 \times r}, \boldsymbol{X}_2 \in \mathbb{R}_{\geq 0}^{n_2 \times r}, \cdots, \boldsymbol{X}_k \in \mathbb{R}_{\geq 0}^{n_k \times r}$ so that

$$\mathbf{X} \approx [\![\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_k]\!]. \tag{4}$$

The $\boldsymbol{X}_i$ matrices will be referred to as the NCPD factor matrices. This decomposition has found numerous applications in the area of *dynamic topic modeling* where one seeks to discover topic emergence and evolution [22, 23, 24].

1.5. Related work

Previous works have developed hierarchical tensor decomposition models and methods [8, 9, 10, 11]. The models most similar to ours are that of [25], which we refer to as hierarchical nonnegative tensor factorization (HNTF), and [11], which we refer to as HNCPD. HNTF consists of a sequence of NCPDs, in which one of the factor matrix is held constant at each layer while the remaining factor matrices produce the tensor that is decomposed at the next layer, and as a result the performance of HNTF varies significantly based on which data mode appears first in the representation of the tensor. We refer to 'HNTF-i' as HNTF applied to the representation of the tensor where the modes are reordered with mode i first. HNCPD consists of an initial NCPD followed by an HNMF applied to each of the resulting factor matrices. In what follows, Neural HNCPD denotes the HNCPD model trained with a neural network architecture, while Standard HNCPD denotes the HNCPD model trained with a multiplicative updates method [11].

1.6. Contribution and Organization

In Section 2, we introduce our proposed hierarchical tensor decomposition model, Multi-HNTF, first for the special case of matrix data, and then in general for tensor data. In Section 3, we perform topic modeling experiments on two document analysis data sets and one synthetic data set, and compare our model to other HNMF and HNTF models. Finally, in Section 4 we summarize our findings and discuss future work.

2. MODEL

Here we describe our proposed model for hierarchical matrix and tensor decomposition, Multi-HNTF. Like previous hierarchical nonnegative matrix factorization models, we seek a se-

ries of factorizations in which sequential factor matrices are approximate factors of their predecessors (thus leading to a clear linear relationship between learned topics). However, unlike previous matrix factorization models, we propose a hierarchical decomposition model that generalizes naturally to higher-order tensors. As the matrix model is a special case of the tensor decomposition model, we give pseudo-code only for the tensor model but give intuition for the matrix case first.

Matrix model: Our model consists of a sequence of nonnegative matrix factorizations that decompose input data matrix $\boldsymbol{X} \in \mathbb{R}^{m \times n}$. Given desired ranks $r_0, r_1, \cdots, r_{\mathcal{L}}$, this process produces $\boldsymbol{A}^{(\ell)} \in \mathbb{R}^{m \times r_\ell}$ and $\boldsymbol{S}^{(\ell)} \in \mathbb{R}^{r_\ell \times n}$ such that $\boldsymbol{X} \approx \boldsymbol{A}^{(\ell)}\boldsymbol{S}^{(\ell)}$, while constraining a linear relationship between successive factor matrices,

$$\boldsymbol{A}^{(\ell+1)} = \boldsymbol{A}^{(\ell)}\boldsymbol{W}^{(\ell)}, \;\; \boldsymbol{S}^{(\ell+1)} = (\boldsymbol{W}^{(\ell)})^T\boldsymbol{S}^{(\ell)}$$

where $\boldsymbol{W}^{(\ell)} \in \mathbb{R}^{r_\ell \times r_{\ell+1}}$ for each $\ell = 0 \ldots \mathcal{L}$. We note that the matrix $\boldsymbol{W}^{(\ell)}$ collects the r_ℓ subtopics into $r_{\ell+1}$ supertopics at the ℓth layer.

Tensor model: We generalize this hierarchical model to tensor data by applying the same linear relationship between successive factor matrices in a NCPD model (the tensor generalization of NMF). Given desired ranks $r_0, r_1, \cdots, r_{\mathcal{L}}$ and input tensor $\mathbf{X} \in \mathbb{R}^{n_1, n_2, \cdots, n_k}$, this process produces $\boldsymbol{X}_1^{(\ell)} \in \mathbb{R}^{n_1 \times r_\ell}, \boldsymbol{X}_2^{(\ell)} \in \mathbb{R}^{n_2 \times r_\ell}, \cdots, \boldsymbol{X}_k(\ell) \in \mathbb{R}^{n_k \times r_\ell}$ such that $\mathbf{X} \approx [\![\boldsymbol{X}_1^{(\ell)}, \boldsymbol{X}_2^{(\ell)}, \ldots, \boldsymbol{X}_k^{(\ell)}]\!]$. As in the matrix model, we begin with an initial rank r_0 decomposition; in this case, it is an initial rank r_0 NCPD $\mathbf{X} \approx [\![\boldsymbol{X}_1^{(0)}, \boldsymbol{X}_2^{(0)}, \ldots, \boldsymbol{X}_k^{(0)}]\!]$. We then constrain the linear relationship between successive factor matrices,

$$\boldsymbol{X}_i^{(\ell+1)} = \boldsymbol{X}_i^{(\ell)}\boldsymbol{W}^{(\ell)}$$

for each $\ell = 0 \ldots \mathcal{L}$ and $i = 1 \ldots k$, and compute $\boldsymbol{W}^{(\ell)}$ such that the new factor matrices $\boldsymbol{X}_i^{(\ell+1)}$ form a rank $r_{\ell+1}$ NCPD,

$$\mathbf{X} \approx [\![\boldsymbol{X}_1^{(\ell)}\boldsymbol{W}^{(\ell)}, \boldsymbol{X}_2^{(\ell)}\boldsymbol{W}^{(\ell)}, \cdots, \boldsymbol{X}_k^{(\ell)}\boldsymbol{W}^{(\ell)}]\!].$$

See Figure 1 for a schematic of this model. In Algorithm 1 we display pseudocode for the Multi-HNTF process. We note that Line 4 of Algorithm 1 is only approximate minimization. One could apply any approximate minimization scheme; we apply a multiplicative updates [15] and averaging scheme (we omit details here due to space constraints).

Algorithm 1 Multi-HNTF

1: **procedure** MULTI-HNTF($\mathbf{X}$)
2: $\quad\{\boldsymbol{X}_i^{(0)}\}_{i=1}^k \leftarrow \text{NCPD}(\mathbf{X}, r_0)$
3: $\quad$**for** $\ell = 0 \ldots \mathcal{L}$ **do**
4: $\quad\quad\boldsymbol{W}^{(\ell)} \leftarrow \underset{\boldsymbol{W} \in \mathbb{R}_+^{r_\ell \times r_{\ell+1}}}{\arg\min} \|\mathbf{X} - [\![\boldsymbol{X}_1^{(\ell)}\boldsymbol{W}, \ldots, \boldsymbol{X}_k^{(\ell)}\boldsymbol{W}]\!]\|$
5: $\quad\quad$**for** $i = 0 \ldots k$ **do**
6: $\quad\quad\quad\boldsymbol{X}_i^{(\ell+1)} = \boldsymbol{X}_i^{(\ell)}\boldsymbol{W}^{(\ell)}$

3. EXPERIMENTS

Here, we present the results of applying Multi-HNTF to the 20 Newsgroups data set [26], a synthetic tensor data set [11], and a Twitter political data set [27], along with comparisions to Hierarchical NMF, Neural HNCPD, HNCPD, and HNTF. Our reconstruction loss is the Frobenius norm of difference of the original and reconstructed tensors. Code can be found in `https://github.com/jvendrow/MultiHNTF`.

3.1. 20 Newsgroups dataset

The 20 Newsgroups dataset is a collection of text documents containing messages from newsgroups on the distributed discussion system Usenet [26]. We use a subset of 1000 documents split evenly amongst ten newsgroups (graphics, hardware, forsale, motorcycles, baseball, medicine, space, guns, mideast, and religion) which naturally combine into six supertopics (computer, forsale, recreation, science, politics, religion). We run a two layer Multi-HNTF and HNMF with no supervision, and supervision at both layers, at ranks $r_0 = 10$ and $r_1 = 6$, and report the results in Table 1. We see that with and without supervision, Multi-HNTF outperforms HNMF in reconstruction loss and classification accuracy.

Table 1. Reconstruction loss and classification accuracy at the second layer of two layer Multi-HNTF and HNMF on the 20 newsgroup data set.

	Recon Loss		Accuracy	
Method	Unsup.	Sup.	Unsup.	Sup.
Multi-HNTF	**30.81**	**30.91**	**0.516**	**0.737**
HNMF	30.82	31.45	0.507	0.636

3.2. Synthetic tensor dataset

In order to measure the capacity of Multi-HNTF to identify hierarchical relationships on multi-modal tensor data sets, we run Multi-HNTF on a synthetic tensor data set introduced in [11]. This dataset is a rank seven tensor of size $40 \times 40 \times 40$ comprised of blocks overlayed to form a hierarchical structure, with positive Gaussian noise added to each entry. In Table 2 we display the relative reconstruction loss on the synthetic dataset for Multi-HNTF and comparable models. We see that Multi-HNTF outperforms Standard HNCPD and every ordering of HNTF at each rank $r_1 = 4$ and $r_2 = 2$. Neural HNCPD is able to outperform Multi-HNTF, however due to the repeated forward and backward propagation process, Neural HNCPD utilizes a more complex training method.

3.3. Twitter political dataset

The Twitter political data set [27] is a multi-modal data set of tweets sent by eight political candidates during the 2016

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
trump	tedcruz
hillary	cruz
vote	ted
people	internet
berniesanders	choosecruz
Topic 3	**Topic 4**
johnkasich	crooked
kasich	hillary
ohio	thank
john	great
gov	realdonaldtrump

Rank 2 Topics

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

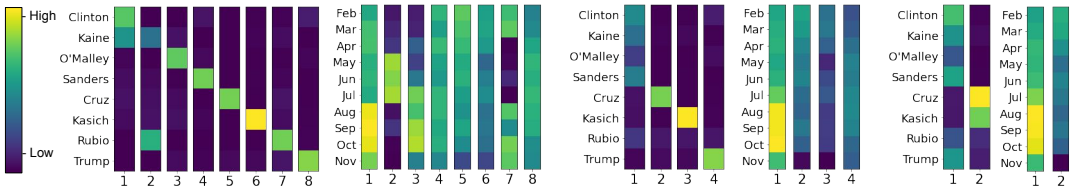

Fig. 2. A three-layer Multi-HNTF on the Twitter dataset at ranks $r_0 = 8$, $r_1 = 4$ and $r_2 = 2$. At each rank, we display the top keywords and topic heatmaps for candidate and temporal modes.

Table 2. Relative reconstruction loss on the synthetic dataset for Multi-HNTF, Neural HNCPD, Standard HNCPD, and HNTF with two levels of noise over 10 trials. For HNTF we report runs on three re-orderings of the modes the tensor.

Method	$r_0 = 7$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.454	0.548	0.721
Neural HNCPD [11]	0.454	**0.508**	**0.714**
Standard HNCPD [11]	0.454	0.612	0.892
HNTF-1 [25]	0.454	0.576	0.781
HNTF-2 [25]	0.454	0.587	0.765
HNTF-3 [25]	0.454	0.560	0.747

Table 3. Relative reconstruction loss on the Twitter political dataset for Multi-HNTF, Neural HNCPD, Standard NCPD, Standard HNCPD, and HNTF (for each of the possible arrangements of the tensor) at ranks $r_0 = 8$, $r_1 = 4$, and $r_2 = 2$.

Method	$r_0 = 8$	$r_1 = 4$	$r_2 = 2$
Multi-HNTF	0.834	0.887	0.920
Neural HNCPD [11]	0.834	**0.883**	**0.918**
Standard NCPD [11]	0.834	0.889	0.919
Standard HNCPD	0.834	0.931	0.950
HNTF-1 [25]	0.834	0.890	0.927
HNTF-2 [25]	0.834	0.909	0.956
HNTF-3 [25]	0.834	0.895	0.942

election season, four Democratic candidates (Hillary Clinton, Tim Kaine, Martin O'Malley, and Bernie Sanders) and four Republican candidates (Ted Cruz, John Kasich, Marco Rubio, and Donald Trump). Following the procedure in [11], we collect all the tweets sent by one politician within each bin of 30 days, from February to December 2016, and combine them into a bag-of-words representation summarizing that politician's twitter activity for the 30 day period. We cap each 30-day bins per politician to 100 tweets to avoid over-fitting to a single month. This forms a tensor of size $8 \times 10 \times 12721$ with 8 politicians, 10 time periods of 30 days, and 12721 words; multi-modal data with candidate, temporal, and text modes.

In Table 3, we list relative reconstruction loss for Multi-HNTF and comparable methods. We see that Multi-HNTF outperforms every method other than Neural HNCPD. In Figure 2 we display a visualization of the topics and keywords learned by Multi-HNTF. At rank 8, we see that there is nearly a one-to-one relationship between topics and candidates, at rank 4 many of the democratic candidates combine into a single topic, and at rank 2 Cruz and Kasich are separated from the other candidates. This makes sense because both candidates were Republicans who left the race at similar times. Note that at rank 2, the first topic, which includes the two final presidential candidates, remained strong until the November election, while the topic corresponding to Cruz and Kasich, who dropped out earlier, has weaker presence in later months.

4. CONCLUSION

We propose Multi-HNTF, a novel HNTF model that naturally generalizes from a special-case HNMF model. Our initial experiments suggest this model provides improvements both for matrix data over a standard HNMF model, and for multi-modal tensor data sets over other HNTF models. We expect that by optimizing our model with a more involved method akin to those of [6, 20, 7, 11] we could further improve the performance of Multi-HNTF over other models.

5. REFERENCES

[1] 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.

[2] D. da Kuang, J. Choo, and H. Park, "Nonnegative matrix factorization for interactive topic modeling and document clustering," pp. 215–243, 10 2015.

[3] J. Douglas Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[4] R. A. Harshman et al., "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," 1970.

[5] J. Flenner and B. Hunter, "A deep non-negative matrix factorization neural network," 2018, Unpublished.

[6] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE T. Pattern Anal.*, vol. 39, no. 3, pp. 417–429, 2016.

[7] 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 *Proc. Int. Workshop on Comp. Adv. in Multi-Sensor Adaptive Process.*, 2019.

[8] M. A. O. Vasilescu and E. Kim, "Compositional hierarchical tensor factorization: Representing hierarchical intrinsic and extrinsic causal factors," *arXiv preprint arXiv:1911.04180*, 2019.

[9] L. Song, M. Ishteva, A. Parikh, E. Xing, and H. Park, "Hierarchical tensor decomposition of latent tree graphical models," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 334–342.

[10] L. Grasedyck, "Hierarchical singular value decomposition of tensors," *SIAM J. Matrix Anal. A.*, vol. 31, no. 4, pp. 2029–2054, 2010.

[11] J. Vendrow, J. Haddock, and D. Needell, "Neural nonnegative CP decomposition for hierarchical tensor analysis," in *Proc. 54th Asilomar Conf. Sig. Syst. Comp.*, 2021.

[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 1, MIT Press, 2016.

[13] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[14] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[15] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neur. In.*, 2001, pp. 556–562.

[16] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. SIGIR Conf. Inform. Retriev.*, 2003, pp. 267–273.

[17] A. Cichocki, R. Zdunek, and S. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," in *Proc. Int. Conf. Acoust. Speech Sig. Process.* IEEE, 2006, vol. 5, pp. V–V.

[18] H. Lee, J. Yoo, and S. Choi, "Semi-supervised nonnegative matrix factorization," *IEEE Signal Proc. Let.*, vol. 17, no. 1, pp. 4–7, 2009.

[19] J. Haddock, L. Kassab, S. Li, A. Kryshchenko, R. Grotheer, E. Sizikova, C. Wang, T. Merkh, R. W. M. A. Madushani, M. Ahn, D. Needell, and K. Leonard, "Semi-supervised nonnegative matrix factorization models for document classification," in *Proc. 54th Asilomar Conf. Sig. Syst. Comp.*, 2021.

[20] J. Le Roux, J. R. Hershey, and F. Weninger, "Deep NMF for speech separation," in *Proc. Int. Conf. Acoust. Spee.* IEEE, 2015, pp. 66–70.

[21] X. Sun, N. M. Nasrabadi, and T. D. Tran, "Supervised multilayer sparse coding networks for image classification," *CoRR*, vol. abs/1701.08349, 2017.

[22] A. Cichocki, R. Zdunek, and S. Amari, "Nonnegative matrix and tensor factorization [lecture notes]," *IEEE Signal Proc. Mag.*, vol. 25, no. 1, pp. 142–145, 2007.

[23] A. Traoré, M. Berar, and A. Rakotomamonjy, "Non-negative tensor dictionary learning," 2018.

[24] A. Saha and V. Sindhwani, "Learning evolving and emerging topics in social media: a dynamic NMF approach with temporal regularization," in *Proc. ACM Int. Conf. Web Search Data Min.*, 2012, pp. 693–702.

[25] A. Cichocki, R. Zdunek, and S. Amari, "Hierarchical als algorithms for nonnegative matrix and 3d tensor factorization," in *Int. Conf. on Indep. Component Anal. and Sig. Separ.* Springer, 2007, pp. 169–176.

[26] K. Lang, "20 newsgroups," Jan 2008.

[27] J. Littman, L. Wrubel, and D. Kerchner, "2016 United States Presidential Election Tweet Ids," 2016.