Nonnegative Bilinear Matrix Factorization*

Dorian Vandenthoren[1]

University of Mons, Rue de Houdain 9, 7000 Mons, Belgium

Abstract. This Master's thesis studies a bilinear nonnegative matrix factorization model. The main contribution of this work is a new highly efficient block coordinate descent algorithm, and several acceleration schemes to make it faster. The algorithm is then used on two applications: unsupervised hypersperal unmixing and topic modeling.

1 Motivation

Nonnegative matrix factorization (NMF) is a feature extraction method which aims to get the best possible approximation of a given nonnegative matrix, X, by the product of two smaller nonnegative matrices, W and H, so that $X \approx WH$. NMF has many applications, including hyperspectral image unmixing and topic modeling; see [1]. In hyperspecral unmixing, each pixel is modeled as a nonnegative linear combination of feature vectors that correspond to pure materials. However, in some images, some rays of light may bounce off two or more materials before reaching the eye of the camera, which is not modeled by NMF. Similarly, in most topic models (such as LDA), the interaction between topics is not taken into account (that is, the probability of using a word is a linear combination of the probabilities of the different topics to use that word).

2 Nonnegative Bilinear Matrix Factorization

In order to model rays of light that have bounced back on two materials before being measured, or to model the interactions between two topics (for example, a word that is used by two topics will be more likely to appear if a document discussed these two topics), we consider in this Master's thesis the following nonnegative bilinear matrix factorization (NBMF) model [2]:

$$\min_{W \geq 0, H \geq 0, Z \geq 0} \quad \left\| X - WH - (W \odot W)Z \right\|_F^2, \tag{1}$$

where $\|.\|_F$ is the Frobenius norm (least squares), and the matrix $(W \odot W)$ is constructed as $[W_{:1} \odot W_{:2}, W_{:1} \odot W_{:3}, \ldots, W_{:2} \odot W_{:3}, \ldots, W_{:(r-1)} W_{:r}]$, with $W_{:j}$ being the jth column of W and $\odot$ being the component-wise product. The term

* This work is a synthesis of the Master's thesis defended at the University of Mons in June 2022, and supervised by Nicolas Gillis and Arnaud Vandaele.

$(W \odot W)Z$ allows one to take into account interactions between feature vectors (such as materials in hyperspectral images, or topics in document analysis).

The main contribution of this thesis is a bloc coordinate descent method to solve NBMF (1). The idea follows the so-called hierarchial alternating least squares algorithm (HALS) for NMF [3] that is one of the most efficient algorithm for NMF [1]. It updates the columns of W and rows of H in a cyclic manner with a closed-form solution. In summary, we adapted the updates of the HALS algorithm for NMF to NBMF; see Table 1 for the formulas. The algorithm is guaranteed to decrease the objective function at each iteration, and to converge to a stationary point under mild assumptions.

Table 1. Update formulas: comparison between NMF and NBMF.

	$X \approx WH$ HALS for NMF	$X \approx WH + (W \odot W)Z$ Proposed NBMF updates
$H_{j:} +=$	$\dfrac{W_{:j}{}^\top X - W_{:j}{}^\top WH}{W_{:j}{}^\top W_{:j}}$	$\dfrac{W_{:j}{}^\top X - W_{:j}{}^\top WH - W_{:j}{}^\top (W \odot W)Z}{W_{:j}{}^\top W_{:j}}$
$Z_{j:} +=$	/	$\dfrac{(W \odot W)_{:j}{}^\top X - (W \odot W)_{:j}{}^\top WH - (W \odot W)_{:j}{}^\top (W \odot W)Z}{(W \odot W)_{:j}{}^\top (W \odot W)_{:j}}$
$W_{:j} +=$	$\dfrac{XH_{j:}{}^\top - WHH_{j:}{}^\top}{H_{j:}H_{j:}{}^\top}$	$\dfrac{\left[\begin{array}{l} XH_{j:}{}^\top - WHH_{j:}{}^\top - (W \odot W)ZH_{j:}{}^\top + \left[W_{:\bar{j}} \circ (Z_{\kappa(\bar{j}):} X^\top)^\top\right] \mathbb{1}_{r-1} \\ - \left[W_{:\bar{j}} \circ (WHZ_{\kappa(\bar{j}):}{}^\top)^\top\right] \mathbb{1}_{r-1} - \left[W_{:\bar{j}} \circ \left((W \odot W)ZZ_{\kappa(\bar{j}):}{}^\top\right)^\top\right] \mathbb{1}_{r-1} \end{array}\right]}{\left[H_{j:}H_{j:}{}^\top .+ 2W_{:\bar{j}} Z_{\kappa(\bar{j}):} H_{j:}{}^\top + \left[W_{:\bar{j}} \circ \left(W_{:\bar{j}} Z_{\kappa(\bar{j}):} Z_{\kappa(\bar{j}):}{}^\top\right)\right] \mathbb{1}_{r-1}\right]}$

3 Applications: hyperspectral unmixing, topic modeling

On hyperspectral images, the algorithm successfully identifies materials and their interactions; see Fig. 1 for an example where 'Mat$i$' stand for Material number i.

Applied to the TDT-2 dataset (American news documents of 1998), the algorithm was able to establish links between different extracted topics, e.g., the Pope's trip to Cuba was linked with aviation (because the Pope suggested his followers to take the plane) and to the tobacco industry (because of Cuba).

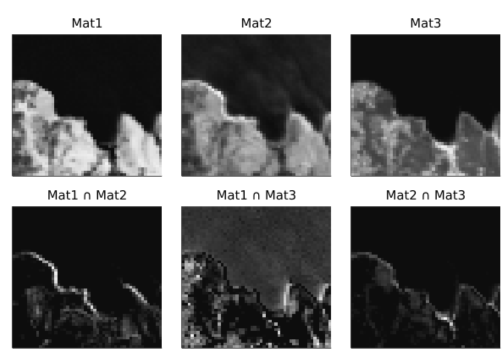

Fig. 1. Result on the Moffet field dataset.

References

1. Gillis, N.: Nonnegative Matrix Factorization, SIAM, Philadelphia, 2020.
2. Dobigeon, N., Tourneret, J. Y., Richard, C., Bermudez, J. C. M., McLaughlin, S., Hero, A. O.: Nonlinear unmixing of hyperspectral images: Models and algorithms. IEEE Signal Processing Magazine, 31(1), 82-94, 2014.
3. Cichocki, A., Zdunek, R., Amari, S. I.: Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In International Conference on Independent Component Analysis and Signal Separation (pp. 169-176). Springer, Berlin, Heidelberg, 2007.