Minimum-Volume Weighted Symmetric Nonnegative Matrix Factorization for Clustering

Tianxiang Gao[†], Sigurdur Olafsson[†], Songtao Lu[‡]

[†]Department of IMSE, Iowa State University
[‡]Department of ECE, Iowa State University

{gaotx,olafsson,songtao}@iastate.edu

Abstract

In recent years, nonnegative matrix factorization (NMF) attracts much attention in machine learning and signal processing fields due to its interpretability of data in a low dimensional subspace. For clustering problems, symmetric nonnegative matrix factorization (SNMF) as an extension of NMF factorizes the similarity matrix of data points directly and outperforms NMF when dealing with nonlinear data structure. However, the clustering results of SNMF is very sensitive to noisy data. In this paper, we propose a *minimum-volume weighted SNMF* (MV-WSNMF) based on the relationship between robust NMF and SNMF. The proposed MV-WSNMF can approximate the similarity matrices flexibly such that the resulting performance is more robust against noise. A computationally efficient algorithm is also proposed with convergence guarantee. The numerical simulation results show the improvement of the proposed algorithm with respective to clustering accuracy in comparison with the state-of-the-art algorithms.

Introduction

Unsupervised learning problems can be understood as factorizing a data matrix X via factors U and V subject to different constrains

$$\begin{aligned} &\underset{\mathbf{U},\mathbf{V}}{\text{minimize}} \ \|\mathbf{X}-\mathbf{U}\mathbf{V}^T\|_F^2 \\ &\text{subject to } \ \mathbf{V}\in\{0,1\}^{N\times K}, \mathbf{V}\mathbf{1}_K=\mathbf{1}_N, \end{aligned} \tag{1}$$

NMF and dimension reduction have the same formulation but with various constrains. For example, NMF relaxes discreteness but enforces nonnegativity for a parts-based representation.

Main Contributions

- A minimum-volume WSNMF method is proposed, which shows strong robustness against noise and provides accurate clustering results.
- An efficient four-block cyclic optimization algorithm is derived for solving the MV-WSNF problem, where every limit point converges to a stationary point.

Minimum-Volume Enclosing Simplex

In *hyperspectral unmixing* (HU) problem, MVES [3] is to circumscribe the data as compact as possible, which is formulated as follows

$$\begin{aligned} &\underset{\mathbf{U},\mathbf{Z}}{\text{minimize}} \ \|\mathbf{X}-\mathbf{U}\mathbf{Z}\|_F^2+\beta\text{vol}(\mathbf{U}) \\ &\text{subject to } \ \mathbf{Z}\geq 0, \mathbf{Z}^T\mathbf{1}_K=\mathbf{1}_N, \end{aligned} \tag{2}$$

where $\mathbf{U}$ represents a material signature matrix, and $\beta>0$ is a parameter for taking the tradeoff between the approximation error and volume minimization, and $\text{vol}(\mathbf{U})\triangleq\log\det(\mathbf{U}^T\mathbf{U})$ [1].

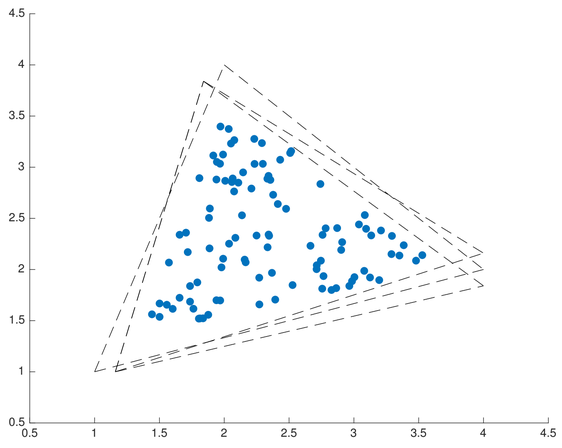

Figure 1: Geometric interpretation of the nonuniqueness of the solution.

Minimum-Volume WSNMF

To conquer the nonlinear data structure with noisy measurements, we propose *minimum-volume WSNMF* (MV-WSNMF) for clustering problems as follows

$$\begin{aligned} &\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \ \|\mathbf{A}-\mathbf{H}\mathbf{S}\mathbf{H}^T\|_F^2+\beta\log\det(\mathbf{S}+\epsilon\mathbf{I}_K) \\ &\text{subject to } \ \mathbf{H}\geq 0, \end{aligned} \tag{3}$$

where $\mathbf{A}$ is similarity matrix, and $\epsilon>0$ is a number in order to prevent the second term unbounded and also ensures $\mathbf{S}+\epsilon\mathbf{I}_K$ is *positive definite*.

Variable Splitting for MV-WSNMF

To solve the problem (3), we propose a *variable splitting* algorithm for MV-WSNMF that can converge to a stationary point. The algorithm updates the variables in a cyclic fashion, where the block coordinate descent framework can be directly applied.

A Nonsymmetric Formulation for MV-WSNMF

With leveraging the lemma introduced in [2, Lemma. 2], and splitting the variable $\mathbf{H}$ as $\mathbf{H}\approx\mathbf{W}$, problem (3) can be reformulated as

$$\begin{aligned} &\underset{\mathbf{W},\mathbf{H},\mathbf{S},\mathbf{F}}{\text{minimize}} \ \mathcal{L}=\|\mathbf{A}-\mathbf{W}\mathbf{S}\mathbf{H}^T\|_F^2+\alpha\|\mathbf{W}-\mathbf{H}\|_F^2 \\ &\qquad\qquad +\beta\left(\mathbf{tr}(\mathbf{F}(\mathbf{S}+\epsilon\mathbf{I}_K))-\log\det(\mathbf{F})\right) \\ &\text{subject to } \ \mathbf{H},\mathbf{W}\geq 0, \mathbf{F}\succeq 0, \end{aligned} \tag{4}$$

where α is the regularization parameter that balances the accurate reconstruction and the difference between $\mathbf{W}$ and $\mathbf{H}$.

Algorithm for Nonsymmetric MV-WSNMF

Note that we can cyclically update $\mathbf{S}$, $\mathbf{F}$, $\mathbf{H}$ and $\mathbf{W}$ until a stopping criteria is satisfied.

Algorithm 1: The Variable Splitting Algorithm

Input: similarity matrix $\mathbf{A}$, number of clusters K
Output: $\mathbf{H}$
initialize $\mathbf{H}$
$\mathbf{W}\leftarrow\mathbf{H}$
repeat

$$\mathbf{S}\leftarrow\frac{1}{2}(\mathbf{W}^T\mathbf{W})^\dagger(\mathbf{W}^T\mathbf{A}\mathbf{H}+\mathbf{H}^T\mathbf{A}\mathbf{W}-\beta\mathbf{F})(\mathbf{H}^T\mathbf{H})^\dagger$$

$$\mathbf{F}\leftarrow(\mathbf{S}+\epsilon\mathbf{I}_K)^{-1}$$

$$\mathbf{W}\leftarrow\underset{\mathbf{W}\geq 0}{\text{argmin}}\left\|\begin{bmatrix}\mathbf{H}\mathbf{S}\\ \sqrt{\alpha}\mathbf{I}_K\end{bmatrix}\mathbf{W}^T-\begin{bmatrix}\mathbf{A}\\ \sqrt{\alpha}\mathbf{H}^T\end{bmatrix}\right\|_F^2$$

$$\mathbf{H}\leftarrow\underset{\mathbf{H}\geq 0}{\text{argmin}}\left\|\begin{bmatrix}\mathbf{W}\mathbf{S}\\ \sqrt{\alpha}\mathbf{I}_K\end{bmatrix}\mathbf{H}^T-\begin{bmatrix}\mathbf{A}\\ \sqrt{\alpha}\mathbf{W}^T\end{bmatrix}\right\|_F^2$$

until *stopping criteria satisfied*;

Proposition. *Every limit point of the solution sequence generated by the Algorithm 1 is a stationary point of problem* (4).

Numerical Results

We evaluate the performances of MV-WSNMF on a number of synthetic data sets and compare with the existing state-of-the-art clustering algorithms, including K-Means, NMF, *graph-regularized NMF* (GNMF), *locally consistency concept factorization* (LCCF), normalized Cut (NCut), SNMF, WSNMF, and MV-WSNMF.

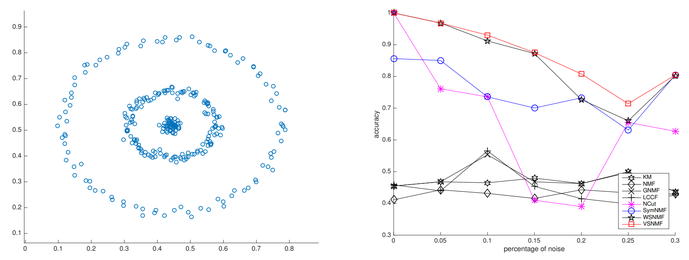

Figure 2: Clustering results for data set 1

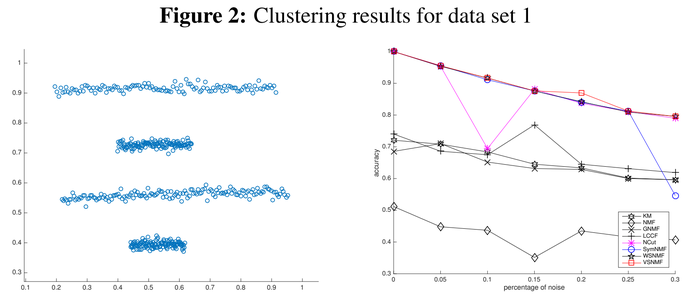

Figure 3: Clustering results for data set 2

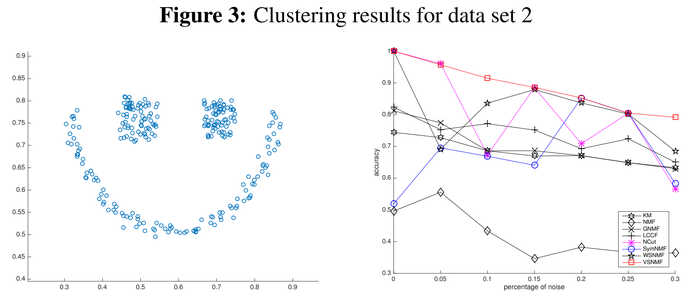

Figure 4: Clustering results for data set 3

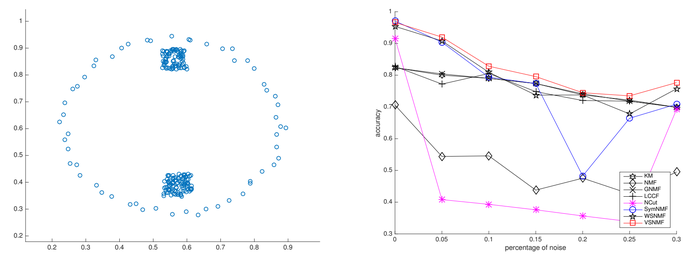

Figure 5: Clustering results for data set 4

Clustering Results

- Kernel functions play an important role in clustering problems since the data sets are transformed in a high dimensional space.
- Factorizing the similarity matrix shows more accurate clustering results. For example, NCut, and SNMF achieve better performance than the NMF-based methods.
- The weighted SNMF including WSNMF and MV-WSNMF usually have better performance in terms of accuracy with or without noise background.
- Regardless of the data sets, the MV-WSNMF achieves the highest accuracy in any percentage of noise data, which shows that MV-WSNMF is stable in noisy data by minimizing the volume of latent space.

It is concluded that with leveraging similarity matrix factorization and minimum volume regularization, MV-SNMF can learn an accurate and compact low dimensional representation of data.

Tuning Parameter Selection

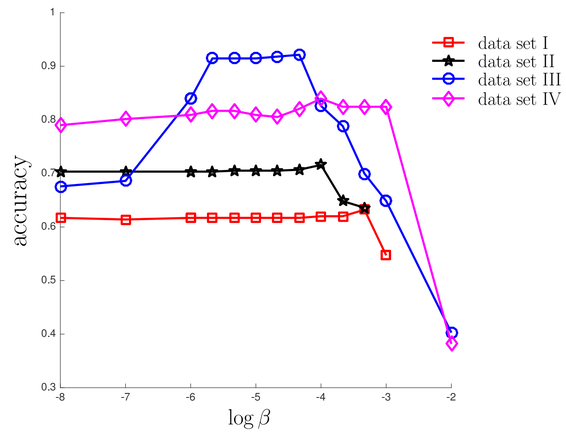

Figure 6: Performance of MV-SNMF varies as the change of β when the percentage of noise data points is 10%

In the MV-SNMF algorithm, there is a regularization parameter β. Figure 6 shows how the clustering accuracy of MV-SNMF changes under various β when the percentage of noise is 10%.

Conclusions

Motivated by the limitation of SNMF, we proposed the MV-WSNMF for improving the robustness of SNMF in clustering problems. An computational efficient four-block cyclical algorithm is proposed and guaranteed to converge to a stationary point. Numerical results show that the clustering accuracy of MV-SNMF based algorithm is much more robust against noisy data in comparison with other benchmark algorithms.

References

[1] Xiao Fu, Wing-Kin Ma, Kejun Huang, and Nicholas D Sidiropoulos. Robust volume minimization-based matrix factorization via alternating optimization. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2534–2538. IEEE, 2016.

[2] Jubin Jose, Narayan Prasad, Mohammad Khojastepour, and Sampath Rangarajan. On robust weighted-sum rate maximization in MIMO interference networks. In *Proc. of 2011 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2011.

[3] Lidan Miao and Hairong Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, 2007.