IEICE
TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

VOL. E103-A NO. 2
FEBRUARY 2020

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

LETTER

Shift Invariance Property of a Non-Negative Matrix Factorization

Hideyuki IMAI[†,††a)], ***Member***

SUMMARY We consider a property about a result of non-negative matrix factorization under a parallel moving of data points. The shape of a cloud of original data points and that of data points moving parallel to a vector are identical. Thus it is sometimes required that the coefficients to basis vectors of both data points are also identical from the viewpoint of classification. We show a necessary and sufficient condition for such an invariance property under a translation of the data points.

key words: *non-negative matrix factorization, semi non-negative matrix factorization, parallel moving*

1. Introduction

Non-negative matrix factorization (abbreviated to NMF) is a matrix decomposition method first introduced by Lee and Seung [1], [2]. In an early period of NMF, it focused on decomposing parts based representation such as a picture part classification [1] or an audio signal processing [3], since NMF would reveal the intrinsic parts underlying the object. It has been applied also for clustering or object classification such as document classification [4] or image classification [5]. However, we do not always regard NMF as a clustering function because it depend not only on distances between individuals, but on the selection of the origin.

In this paper, we consider an invariance property under a parallel moving of data points, or in other words, a translation of the data matrix, and show a necessary and sufficient condition such that coefficient matrices of NMF are identical under a translation of the data matrix.

2. Non-Negative Matrix Factorization

A matrix whose elements are non-negative is called a non-negative matrix. The basic idea of NMF is to decompose a non-negative data matrix into a product of two non-negative matrices. It enables us to represent a column vector of a data matrix as an additive combination of basis vectors.

Let $X = (x_{ij})$ be a $p \times n$ non-negative matrix, where p and n represent the numbers of variables (or features) and individuals, respectively.

Thus, the aim of NMF is to find matrices satisfying

$$X = UV, \tag{1}$$

where U and V are $p \times r$ and $r \times n$ non-negative matrices, respectively. The columns of the matrix U are regarded as basis vectors, r is the number of basis vectors, and the columns of the matrix V are regarded as coefficients to basis vectors. In practice, r is often chosen such that $r \ll \min(p, n)$. In this paper, we assume that $\mathrm{rank}(V) = r$.

Unfortunately, the decomposition (1) is in general not unique. Uniqueness of NMF has been studied since it was proposed and some conditions for uniqueness with its geometric interpretation are found in Donoho and Stodden [6], and Huang et al. [7].

3. Shift Invariance Property

Consider a translation of the data matrix that has the form

$$X(\boldsymbol{a}) = X + \boldsymbol{a}\mathbf{1}_n^\top, \tag{2}$$

where $\boldsymbol{a} \in \mathbb{R}^p$ is a non-zero vector, $\mathbf{1}_n$ denotes the n-dimensional column vector with all components one, and $A^\top$ denotes the transpose of a vector or a matrix A. As in Eq. (1), we write NMF of $X(\boldsymbol{a})$ as

$$X(\boldsymbol{a}) = U(\boldsymbol{a})V(\boldsymbol{a}), \tag{3}$$

where the sizes of non-negative matrices $U(\boldsymbol{a})$ and $V(\boldsymbol{a})$ are equal to those of U and V. Because the shapes of the data points X and $X(\boldsymbol{a})$ in p dimensional Euclidean space are identical, the coefficients matrices V and $V(\boldsymbol{a})$ are needed to be identical when we use NMF as a clustering function. We say NMF of X is shift invariant to $\boldsymbol{a} \in \mathbb{R}^p$ if $X(\boldsymbol{a})$ has a non-negative factorization $X(\boldsymbol{a}) = U(\boldsymbol{a})V$ under a translation (2). It is shown about a shift invariance property that:

> Let $X = UV$ be a non-negative factorization with $\mathrm{rank}(V) = r$. NMF of X is shift invariant to $\boldsymbol{a} \in \mathbb{R}^p$, that is, $X(\boldsymbol{a})$ has a non-negative factorization
>
> $$X(\boldsymbol{a}) = U(\boldsymbol{a})V, \tag{4}$$
>
> if and only if there exists a constant vector $\boldsymbol{w} = [w_1, \ldots, w_r]^\top$ such that
>
> $$V^\top \boldsymbol{w} = \mathbf{1}_n, \tag{5}$$
>
> and $U + \boldsymbol{a}\boldsymbol{w}^\top$ has non-negative entries.

Proof. If Eq. (4) holds, we have

$$V^\top (U(\boldsymbol{a}) - U)^\top = \mathbf{1}_n \boldsymbol{a}^\top.$$

Manuscript received August 27, 2019.

[†]The author is with Faculty of Information Science and Technology, Hokkaido University, Sapporo-shi, 060-0814 Japan.

[††]The author is with Global Station for Big Data and Cybersecurity, Hokkaido University, Sapporo-shi, 060-0814 Japan.

a) E-mail: imai@ist.hokudai.ac.jp

DOI: 10.1587/transfun.2019EAL2121

Let a_i be a non-zero elements of $\boldsymbol{a} = [a_1, \ldots, a_p]^\top$ and $U(\boldsymbol{a}) - U = [\boldsymbol{u}_1(\boldsymbol{a}), \ldots, \boldsymbol{u}_r(\boldsymbol{a})]^\top$, then

$$V^\top \boldsymbol{w} = \mathbf{1}_n,$$

holds for $\boldsymbol{w} = a_i^{-1}\boldsymbol{u}_i(\boldsymbol{a})$. Since $\mathrm{rank}(V) = r$, $\boldsymbol{w}$ is a constant vector and $U + \boldsymbol{a}\boldsymbol{w}^\top = U(\boldsymbol{a})$ has non-negative entries.

Conversely, if there exists a constant vector $\boldsymbol{w} \in \mathbb{R}^r$ such that $V^\top \boldsymbol{w} = \mathbf{1}_n$ and $U + \boldsymbol{a}\boldsymbol{w}^\top$ has non-negative entries, we have

$$\begin{aligned} X(\boldsymbol{a}) &= UV + \boldsymbol{a}\mathbf{1}_n^\top = UV + \boldsymbol{a}\boldsymbol{w}^\top V \\ &= (U + \boldsymbol{a}\boldsymbol{w}^\top)V = U(\boldsymbol{a})V \end{aligned}$$

as a non-negative factorization. □

4. Application to Data Analysis

The condition Eq. (5) does not depend on a shift vector $\boldsymbol{a} \in \mathbb{R}^p$. Thus, NMF of X is shift invariant to an arbitrary shift vector $\boldsymbol{a} \in \mathbb{R}^p$ such that $U + \boldsymbol{a}\boldsymbol{w}^\top$ has non-negative entries if the condition Eq. (5) holds.

Though the condition contains an unknown vector $\boldsymbol{w} \in \mathbb{R}^r$, it seems to be free from ambiguity. Suppose that we find a solution of Eq. (1) with $V^\top \mathbf{1}_r = \mathbf{1}_n$, then we obtain a solution $U_1 = UW$ and $V_1 = W^{-1}V$ with the $r \times r$ diagonal matrix W whose diagonal elements $\{w_1, \ldots, w_r\}$. This is a solution satisfying $V_1^\top \boldsymbol{w} = \mathbf{1}_n$. Thus, we always use $\mathbf{1}_r$ as a weighting vector $\boldsymbol{w} \in \mathbb{R}^r$.

Consequently, in an application, our aim is to find non-negative matrices U^* and V^* such that

$$\begin{aligned} (U^*, V^*) &= \underset{U,V}{\mathrm{argmin}} \|X - UV\|^2 \\ &= \underset{U,V}{\mathrm{argmin}} \sum_{i=1}^{p} \sum_{j=1}^{n} \left(x_{ij} - \sum_{\alpha=1}^{r} u_{i\alpha} v_{\alpha j} \right)^2 \end{aligned}$$

subject to

$$V^\top \mathbf{1}_r = \mathbf{1}_n.$$

Under the optimization criterion, NMF of X is shift invariant to an arbitrary shift vector $\boldsymbol{a} \in \mathbb{R}^p$ such that $U^* + \boldsymbol{a}\mathbf{1}_r^\top$ has non-negative entries if $\mathrm{rank}(V^*) = r$ holds, because

$$\underset{U,V}{\mathrm{argmin}} \|X(\boldsymbol{a}) - UV\|^2 = \underset{U,V}{\mathrm{argmin}} \|X + (U - \boldsymbol{a}\mathbf{1}_r^\top)V\|^2$$

shows that $(U^* + \boldsymbol{a}\mathbf{1}_r^\top, V^*)$ is an optimal solution for a non-negative matrix $X(\boldsymbol{a}) = X + \boldsymbol{a}\mathbf{1}_n^\top$.

5. Concluding Remarks

When we use NMF as a clustering function, it is natural to require the same results when all data points move parallel to a vector. We show a necessary and sufficient condition for such a requirement.

It is to be noted that the proof of the invariance property for NMF can be applied for semi-NMF [8], because it is still valid for X and U with negative elements.

An advantage of NMF is that we have a sparse solution for both U and V. It enables us to interpret an underlying data structure such as picture parts in digital images. A shift invariant NMF, however, gives a non-sparse solution in almost all cases, and it may become a serious drawback for understanding a data structure. To overcome such disadvantages, we should impose some constraints [9].

On the basis of the result shown in the study, we will develop an effective algorithm for a shift invariant NMF, and apply it to find groups or clusters in real world data sets.

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.788–791, Oct. 1999.

[2] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization," Advances in Neural Information Processing Systems 13, T.K. Leen, T.G. Dietterich, and V. Tresp, eds., pp.556–562, MIT Press, 2001.

[3] P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp.177–180, 2003.

[4] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," Proc. 26th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR'03, p.267, 2003.

[5] D. Guillamet, J. Vitrià, and B. Schiele, "Introducing a weighted non-negative matrix factorization for image classification," Pattern Recogn. Lett., vol.24, no.14, pp.2447–2454, 2003.

[6] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?," Advances in Neural Information Processing Systems 16, S. Thrun, L.K. Saul, and B. Schölkopf, eds., pp.1141–1148, MIT Press, 2003.

[7] K. Huang, N.D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric," IEEE Trans. Signal Process., vol.62, no.1, pp.211–224, 2014.

[8] C.H.Q. Ding, T. Li, and M.I. Jordan, "Convex and semi-nonnegative matrix factorizations," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.1, pp.45–55, Jan. 2010.

[9] K. Kimura, M. Kudo, and Y. Tanaka, "A column-wise update algorithm for nonnegative matrix factorization in Bregman divergence with an orthogonal constraint," Mach. Learn., vol.103, no.2, pp.285–306, 2016.