

A MODIFIED STRATEGY IN ALTERNATING NON-NEGATIVE LEAST SQUARES FOR NON-NEGATIVE MATRIX FACTORIZATION

LI Xiang-li, ZHANG Wen, YU Jiang-lan

(School of Mathematics and Computing Science;

Guangxi Key Laboratory of Cryptography and Information Security;

Guangxi Key Laboratory of Automatic Detecting Technology and Instruments,

Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In this paper, we study the global convergence of non-negative least squares (ANLS) for NMF. By applying a modified strategy to guarantee the existence of the limit point, we obtain the limit point is a stationary point of NMF. In addition, we give generalized modified strategies. Numerical experimental results show the above strategies are effective.

Keywords: non-negative matrix factorization (NMF); alternating non-negative least squares (ANLS); modified strategy

2010 MR Subject Classification: 49M37

Document code: A

Article ID: 0255-7797(2018)06-1023-08

1 Introduction

Non-negative matrix factorization (NMF) [1] is different to other factorization methods because non-negative constraints are imposed. This makes factorization results more sense in the real word.

NMF attempts to decompose a given nonnegative matrix A into the product of a non-negative basis matrix W and a non-negative coefficient matrix H , which can be interpreted as follow

$$A \approx W \times H, \quad (1.1)$$

where $A \in R^{m \times n}$, $W \in R^{m \times r}$ and $H \in R^{r \times n}$. The factorization results show that NMF can generate a reduced representation of the original data. In this sense, NMF is useful for extracting parts-based representations.

* **Received date:** 2017-05-08

Accepted date: 2017-07-21

Foundation item: Supported by the National Natural Science Foundation of China (11601012; 71561008); Guangxi Natural Science Foundation (2018GXNSFAA138169); Guangxi Key Laboratory of Cryptography and Information Security (GCIS201708); Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ16112; YQ18112).

Biography: Li Xiangli (1977–), female, born at Baoji, Shaanxi, associate professor, major in optimization theory and its application.

In recent years, NMF, as a parts-based representation of data, received considerable attentions in machine learning, signal processing, computer vision and so on [2–5].

In order to decrease the approximation error between A and WH , NMF be expressed as a constrained optimization problem

$$\min_{W, H} F(W, H) = \frac{1}{2} \|A - WH\|_F^2 \quad \text{s.t.} \quad W \geq 0, \quad H \geq 0, \quad (1.2)$$

where $\|\cdot\|$ is the Frobenius norm, and $W \geq 0$ ($H \geq 0$) means that all elements of the matrix W (H) are nonnegative. Kush-Kuhn-Tucker (KKT) [6] conditions for problem (1.2) are expressed as follows

$$\begin{aligned} W \geq 0, \quad H \geq 0, \quad \nabla_W F(W, H) &= (WH - A)H^T \geq 0, \\ \nabla_H F(W, H) &= W^T(WH - A) \geq 0, \\ \langle W, \nabla_W F(W, H) \rangle &= 0, \quad \langle H, \nabla_H F(W, H) \rangle = 0. \end{aligned} \quad (1.3)$$

Alternating non-negative least squares (ANLS) [2, 7, 8] works well in practice, whose convergent speed is fast and elements not be locked. Iterate the following until a stopping criterion is satisfied

$$H^{k+1} = \arg \min_{H \geq 0} F(W^k, H), \quad (1.4)$$

$$W^{k+1} = \arg \min_{W \geq 0} F(W, H^{k+1}). \quad (1.5)$$

Clearly, $F(W^k, H)$ and $F(W, H^{k+1})$ is convex, respectively.

ANLS is widely used as an efficient computational method for NMF, but the ANLS has a serious problem that its global convergence is not guaranteed in theory. In other words, for any initial solution, the sequence of ANLS contains at least one limit point and this limit point is a stationary point of NMF. The main difficulty in proving global convergence is that the existence of the limit point.

In [9], authors proposed a modified strategy and applied it to ANLS. The modified strategy is valid for proving global convergence. Motivated by the works of [9], we present a modified strategy to guarantee the existence of the limit point, and the limit point is a stationary point of NMF. In addition, we give generalized modified strategies. We can apply it in reality.

The rest of this paper is organized as follows. Our modified strategy is given in Section 2. Convergence analysis is established in Section 3. In Section 4, we give two generalized modified strategies. Finally, we conclude this paper in Section 5.

We sum up here briefly our notations. Let

$$A = (a_{ij})_{m \times n}, \|A\|_F = \sqrt{\sum_i \sum_j a_{ij}^2}, \text{Tr}(A) = \sum_i a_{ii}.$$

Let $B = R^{m \times n}$, $\langle A, B \rangle = \text{Tr}(AB^T)$. $A_{\cdot i}$ denotes the i -th column of A , A_i denotes the i -th row of A . Let $\|x\|$ denote any norm of a vector, and $\|x\|_2$ denote 2-norm of a vector x .

2 Algorithm for Updating and Element of Matrix

In this section, based on the literature [9], we will present our modified strategy. Let (W^k, H^k) be generated by alternating non-negative least squares (ANLS).

First, we state the modified strategy of [9] be given as follows

$$\bar{W}^k = W^k \Lambda_1, \bar{H}^k = \Lambda_2 H^k,$$

where $\Lambda_1 = \text{diag}(\lambda_j^1)$, $\Lambda_2 = \text{diag}(\lambda_j^2)$, $j = 1, 2, \dots, r$,

$$\lambda_j^1 = \begin{cases} 0, & \text{if } \|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\| = 0, \\ \sqrt{\alpha \frac{\|H_{j\cdot}^k\|}{\|W_{\cdot j}^k\|}}, & \text{otherwise,} \end{cases} \quad (2.1)$$

$$\lambda_j^2 = \begin{cases} 0, & \text{if } \|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\| = 0, \\ \sqrt{\frac{\|W_{\cdot j}^k\|}{\alpha \|H_{j\cdot}^k\|}}, & \text{otherwise,} \end{cases} \quad (2.2)$$

where α is a positive constant. The above strategy can ensure ANLS is convergent.

Motivated by the work of [9], we give another modified strategy. In NMF, we have the fact that

$$WH = \sum_{j=1}^r W_{\cdot j} H_{j\cdot} = \sum_{\|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\| \neq 0} \lambda_j W_{\cdot j} \cdot \frac{1}{\lambda_j} H_{j\cdot}, \quad \forall \lambda_j > 0.$$

For $\|W_{\cdot j}\| \cdot \|H_{j\cdot}\| \neq 0$, let $\bar{W}_{\cdot j} = \lambda_j W_{\cdot j}$, $\bar{H}_{j\cdot} = \frac{1}{\lambda_j} H_{j\cdot}$. Furthermore, $\|\bar{W}_{\cdot j}\| = \|\bar{H}_{j\cdot}\| + \alpha$ ($\alpha > 0$). We have

$$\lambda_j = \frac{\alpha + \sqrt{\alpha^2 + 4\|W_{\cdot j}\| \cdot \|H_{j\cdot}\|}}{2\|W_{\cdot j}\|}.$$

Hence, our modified strategy be described as follows: $\bar{W}^k = W^k \Lambda_1, \bar{H}^k = \Lambda_2 H^k$, where $\Lambda_1 = \text{diag}(\lambda_j^1)$, $\Lambda_2 = \text{diag}(\lambda_j^2)$, $j = 1, 2, \dots, r$,

$$\lambda_j^1 = \begin{cases} 0, & \text{if } \|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\| = 0, \\ \frac{\alpha + \sqrt{\alpha^2 + 4\|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\|}}{2\|W_{\cdot j}^k\|}, & \text{otherwise,} \end{cases} \quad (2.3)$$

$$\lambda_j^2 = \begin{cases} 0, & \text{if } \|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\| = 0, \\ \frac{2\|W_{\cdot j}^k\|}{\alpha + \sqrt{\alpha^2 + 4\|W_{\cdot j}^k\| \cdot \|H_{j\cdot}^k\|}}, & \text{otherwise.} \end{cases} \quad (2.4)$$

3 Algorithm

Based on the above analysis, in this section, we report ANLS with Strategy 1 for NMF as follows.

Algorithm 1 (s.1) Give the starting point $(\bar{W}^0, \bar{H}^0) \geq 0$ and $\alpha \geq 0$. Set $k = 1$.

(s.2) Stop if $(\bar{W}^k, \bar{H}^k)$ satisfies (1.3),

$$W^{k+1} = \arg \min_{W \geq 0} F(W, \bar{H}^k), \quad (3.1)$$

$$H^{k+1} = \arg \min_{H \geq 0} F(W^{k+1}, H). \quad (3.2)$$

(s.3) Modified (W^{k+1}, H^{k+1}) to $(\bar{W}^{k+1}, \bar{H}^{k+1})$ using Strategy 1.

(s.4) Let $k = k + 1$. Go to s.2.

4 Convergence Analysis

Similar to [9], the function of (1.4) ((1.5)) has a global minimizer on $R_+^{n \times r}$. In this section, we come to study the global convergence of the algorithm.

Lemma 4.1 Let $\{\bar{W}^k, \bar{H}^k\}$ be generated by Algorithm 1. Then $\{\bar{W}^k, \bar{H}^k\}$ is bounded.

Proof H^k is a stationary of $F(W^k, H)$, we have $\langle (W^k)^T (W^k H^k - A), H^k = 0 \rangle$,

$$\begin{aligned} \|A - W^k H^k\|_F^2 &= \|A\|_F^2 - 2\langle A, W^k H^k \rangle + \|W^k H^k\|_F^2 \\ &= \|A\|_F^2 - 2\langle A - W^k H^k, W^k H^k \rangle - \|W^k H^k\|_F^2 \\ &= \|A\|_F^2 - \|W^k H^k\|_F^2 \geq 0. \end{aligned}$$

Therefore, $\|\bar{W}^k \bar{H}^k\|_F = \|W^k H^k\|_F \leq \|A\|_F$.

Based on $\|\bar{W}_{\cdot j}\| = \|\bar{H}_{j\cdot}\| + \alpha$, we deduce $\|\bar{W}_{\cdot j}\| = \|\bar{H}_{j\cdot}\| + \alpha$. Meanwhile there exist $c_1, c_2, c_3, c_4 > 0$, such that

$$c_2 \|\bar{W}_{\cdot j}\|_2 \leq \|\bar{W}_{\cdot j}\| \leq c_1 \|\bar{W}_{\cdot j}\|_2, \quad c_3 \|\bar{H}_{j\cdot}\|_2 \leq \|\bar{H}_{j\cdot}\| \leq c_4 \|\bar{H}_{j\cdot}\|_2.$$

We can get

$$\|\bar{W}_{\cdot j}\|_2 \geq \frac{\|\bar{W}_{\cdot j}\|}{c_1} = \frac{\|\bar{H}_{j\cdot}\| + \alpha}{c_1} \geq \frac{\|c_3 \bar{H}_{j\cdot}\|_2 + \alpha}{c_1}$$

and

$$\|\bar{H}_{j\cdot}\|_2 \geq \frac{\|\bar{H}_{j\cdot}\|}{c_4} = \frac{\|\bar{W}_{\cdot j}\| - \alpha}{c_4} \geq \frac{\|c_2 \bar{W}_{\cdot j}\|_2 - \alpha}{c_4}.$$

Furthermore,

$$\begin{aligned} \|\bar{W}^k \bar{H}^k\|_F^2 &= \left\langle \sum_{j=1}^r \bar{W}_{\cdot j}^k \bar{H}_{j\cdot}^k, \sum_{j=1}^r \bar{W}_{\cdot j}^k \bar{H}_{j\cdot}^k \right\rangle \geq \sum_{j=1}^r \|\bar{W}_{\cdot j}^k \bar{H}_{j\cdot}^k\|_F^2 = \sum_{j=1}^r \|\bar{W}_{\cdot j}^k\| \|\bar{H}_{j\cdot}^k\|_2^2 \\ &\geq \sum_{j=1}^r \left(\frac{\|c_3 \bar{H}_{j\cdot}^k\|_2 + \alpha}{c_1} \right)^2 \|\bar{H}_{j\cdot}^k\|_2^2 = \sum_{j=1}^r \frac{(\|c_3 \bar{H}_{j\cdot}^k\|_2^2 + \alpha \|\bar{H}_{j\cdot}^k\|_2)^2}{c_1^2}, \end{aligned}$$

where

$$\begin{aligned} \sum_{j=1}^r (\|c_3 \bar{H}_{j\cdot}^k\|_2^2 + \alpha \|\bar{H}_{j\cdot}^k\|_2)^2 &\geq \sum_{j=1}^r (c_3^2 \|\bar{H}_{j\cdot}^k\|_2^4 + \alpha^2 \|\bar{H}_{j\cdot}^k\|_2^2) \\ &= c_3^2 \sum_{j=1}^r \|\bar{H}_{j\cdot}^k\|_2^4 + \alpha^2 \sum_{j=1}^r \|\bar{H}_{j\cdot}^k\|_2^2 \geq c_3^2 \cdot \frac{(\sum_{j=1}^r \|\bar{H}_{j\cdot}^k\|_2^2)^2}{r} + \alpha^2 \sum_{j=1}^r \|\bar{H}_{j\cdot}^k\|_2^2. \end{aligned}$$

Suppose $\sum_{j=1}^r \|\bar{H}_{j\cdot}^k\|_2^2 = m$, we have $c_3^2 \cdot \frac{m^2}{r} + \alpha^2 \cdot m \leq c_1^2 \|A\|_F^2$. By the nature of the quadratic function, there must be a positive constant M , such that $m \leq M$. Hence,

$$\|\bar{H}^k\|_F = \sqrt{\sum_{j=1}^r \|\bar{H}_{j\cdot}^k\|_2^2} = m \leq M.$$

Similarly, we can get the conclusion that $\|\bar{W}^k\|_F$ is bounded. Therefore, $\{\bar{W}^k, \bar{H}^k\}$ is bounded.

Lemma 4.2 Let $\{\bar{W}^k, \bar{H}^k\}$ be generated by Algorithm 1, and (W^*, H^*) be a limit point of $\{\bar{W}^k, \bar{H}^k\}$. Then (W^*, H^*) is a stationary point of problem (1.2).

Since the above theorem corresponds to [9, Theorem 2] and the proof is the same as that in [9, Theorem 2], we will omit it here.

5 Generalized Modified Strategies

From the above results, α is a positive constant no matter Strategy 1 or the strategy of [9]. We might as well replace α with $g(\alpha)$, in which $g(\alpha)$ is a function of α and $g(\alpha) > 0$. Thus, we can get two generalized modified strategies.

The convergence of ANLS also is established when $g(\alpha) > 0$. In reality, we can let $g(\alpha) = \frac{1}{\alpha^2}$ or $g(\alpha) = e^{-\alpha}$ and so on. For different application problems, the choose of $g(\alpha)$ is different.

6 Numerical Experiments

In this section, we give some numerical experiments for the proposed modified strategies. We apply these strategies to the following algorithms: SBBNMF [9], BBNMF [10] and PGNMF [11]. All codes of the computer procedures are written in MATLAB and run in MATLAB 7.10, and are carried out on a PC (CPU 2.13GHz, 2G memory) with Windows 7 operation system environment.

In experiments, we will compare the following statistics: the number of inner iterations (Iter), the number of outer iterations (Oter), the objective function value (Fun), the residual function value (Res), and the CPU time in second (Time). The Res formula is similar to [9], and the initial parameters

$$\epsilon = 10^{-6}, \text{ Oter} \leq 10000, \text{ Iter} \leq 10000, \text{ Time} \leq 300.$$

In order to avoid the initial point affect the numerical result, in every experiment, we use 30 initial points, and every initial point is randomly generated. We list the following four experimental tables for different α .

The three tables indicates the proposed modified strategies are utility for these algorithms (SBBNMF, BBNMF, and PGNMF) of ANLS framework. In some experiments, Iter (Oter) is not an integer, because we test the average value the 30 initial point. In other words, Iter (Oter) represents the average number of iterations.

Table 1: $A = \text{abs}(\text{randn}(m, r)) * \text{abs}(\text{randn}(r, n)), \alpha = 1$

(m, n, r)		Oter	Iter	Fun	Res	Time
(100,50,4)	SBBNMF	30.50	438.00	1.94e+00	8.65e-03	0.05
	BBNMF	27.00	564.50	2.24e+00	9.28e-03	0.06
	PGNMF	29.50	1838.00	9.15e-01	5.94e-03	0.28
(50,50,8)	SBBNMF	22.00	502.50	7.49e+00	1.37e-02	0.04
	BBNMF	21.00	966.50	9.05e+00	1.51e-02	0.12
	PGNMF	29.00	2968.00	5.44e+00	1.17e-02	0.44
(100,50,20)	SBBNMF	12.00	413.50	7.09e+01	1.32e-02	0.19
	BBNMF	17.00	1550.00	9.70e+01	1.55e-02	0.96
	PGNMF	18.50	1351.00	7.55e+01	1.37e-02	0.70
(100,250,5)	SBBNMF	26.00	415.50	1.11e+01	8.30e-03	0.16
	BBNMF	22.00	791.00	1.63e+01	1.00e-02	0.41
	PGNMF	69.50	1713.50	5.06e-01	1.77e-03	0.90
(200,300,8)	SBBNMF	28.00	554.00	1.61e+02	1.31e-02	0.23
	BBNMF	28.00	1503.50	3.01e+02	1.80e-02	0.82
	PGNMF	27.50	1686.00	1.00e+02	1.04e-02	0.90
(400,400,5)	SBBNMF	27.00	530.50	8.56e+01	8.82e-03	0.21
	BBNMF	25.00	1178.50	1.14e+02	1.02e-02	0.55
	PGNMF	34.00	1526.00	1.91e+01	3.75e-03	1.47

Table 2: $A = \text{abs}(\text{randn}(m, r)) * \text{abs}(\text{randn}(r, n)), \alpha = 0$

(m, n, r)		Oter	Iter	Fun	Res	Time
(100,50,4)	SBBNMF	17.50	357.50	5.45e+00	1.56e-02	0.02
	BBNMF	16.00	495.50	6.15e+00	1.65e-02	0.03
	PGNMF	20.50	1493.00	2.74e+00	1.09e-02	0.09
(50,50,8)	SBBNMF	21.00	540.00	9.10e+00	1.51e-02	0.02
	BBNMF	21.00	967.50	9.52e+00	1.54e-02	0.05
	PGNMF	19.50	2003.50	9.41e+00	1.53e-02	0.12
(100,50,20)	SBBNMF	14.00	454.50	9.77e+01	1.56e-02	0.09
	BBNMF	17.00	1529.00	1.10e+02	1.66e-02	0.41
	PGNMF	18.50	2185.50	1.05e+02	1.62e-02	0.53
(100,250,5)	SBBNMF	16.00	393.00	4.03e+01	1.58e-02	0.06
	BBNMF	16.00	991.00	5.79e+01	1.89e-02	0.23
	PGNMF	15.00	2132.50	5.87e+01	1.90e-02	0.53
(200,300,8)	SBBNMF	22.00	596.50	2.94e+02	1.79e-02	0.22
	BBNMF	21.00	2658.50	4.28e+02	2.16e-02	1.15
	PGNMF	20.00	1410.00	4.23e+02	2.15e-02	0.65

Table 3: $A = \text{abs}(\text{randn}(m, r)) * \text{abs}(\text{randn}(r, n)), g(\alpha) = e^{-\alpha}$

(m, n, r)		Oter	Iter	Fun	Res	Time
(100,50,4)	SBBNMF	16.00	394.50	3.85e+00	1.19e-02	0.03
	BBNMF	15.00	563.00	4.22e+00	1.25e-02	0.06
	PGNMF	15.00	1512.50	3.81e+00	1.19e-02	0.17
(50,50,8)	SBBNMF	20.00	492.50	1.21e+01	1.80e-02	0.02
	BBNMF	19.00	893.50	1.21e+01	1.80e-02	0.06
	PGNMF	18.50	2038.50	1.31e+01	1.87e-02	0.14
(100,250,5)	SBBNMF	15.50	387.00	2.52e+01	1.22e-02	0.06
	BBNMF	14.50	1003.00	3.50e+01	1.44e-02	0.25
	PGNMF	14.50	2134.00	3.22e+01	1.38e-02	0.50
(200,300,8)	SBBNMF	21.00	559.00	2.66e+02	1.69e-02	0.26
	BBNMF	19.50	2655.00	4.03e+02	2.09e-02	1.44
	PGNMF	19.00	1361.00	3.99e+02	2.07e-02	0.74
(400,400,8)	SBBNMF	13.00	425.50	2.30e+02	1.50e-02	0.21
	BBNMF	16.50	2722.00	3.27e+02	1.79e-02	1.29
	PGNMF	21.50	848.50	3.19e+02	1.77e-02	0.81

7 Convergence Analysis

Nonnegative matrix factorization (NMF) is not only a well-known matrix decomposition approach but also an utility and efficient feature extraction technique. In this paper, we propose a modified strategy, which can ensure the convergence of ANLS. In addition, we give generalized modified strategies.

References

- [1] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788–791.
- [2] Andrzej C, Rafal Z, Shun I A. New algorithms for non-negative matrix factorization in applications to blind source separation[J]. IEEE Intern. Confer. Acoustics, Speech Signal Proc., 2006: 870–879.
- [3] Gao B, Woo W L, Dlay S S. Variational regularized 2-d nonnegative matrix factorization[J]. Neural Networks Learn. Sys. IEEE Trans., 2012, 23(5): 703–716.
- [4] Shahnaz F, Berry M W, Pauca V P, Plemmons R J. Document clustering using nonnegative matrix factorization[J]. Inform. Proc. Mgt., 2006, 42(2): 373–386.
- [5] Wang Yuan, Jia Yunde, Hu Changbo, Matthewturk. Non-negative matrix factorization framework for face recognition[J]. Intern. J. Patt. Rec. Arti. Intel., 2008, 19(4): 495–511.
- [6] Boyd, Vandenberghe, Faybusovich. Convex optimization[J]. IEEE Trans. Auto. Control, 2006, 51(11): 1859–1859.
- [7] Paatero P, Tapper U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values[J]. Environmetrics, 1994, 5(2): 111–126.

- [8] Rafal Zdunek, Andrzej Cichocki. Non-negative matrix factorization with quasi-newton optimization[J]. Intern. Confer. Arti. Intel. Soft Comp., 2006: 870–879.
- [9] Li Xiangli, Liu Hongwei. Modified subspace barzilai-borwein gradient method for non-negative matrix factorization[J]. Kluwer Academic Publishers, 2013, 55: 173–196.
- [10] Xiao Yunhai, Hu Qingjie. Subspace Barzilai-Borwein gradient method for large-scale bound constrained optimization[J]. Appl. Math. Optim., 2008, 58: 275–290.
- [11] Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization[J]. Neural Comput., 2014, 19(10): 2756–2779.

求解非负矩阵分解的交替非负最小二乘法的一种修正策略

李向利, 张 雯, 余江兰

(桂林电子科技大学数学与计算科学学院; 广西密码学与信息安全重点实验室;
广西自动检测技术与仪器重点实验室, 广西 桂林 541004)

摘要: 本文研究了关于求解非负矩阵分解的交替非负最小二乘法的全局收敛性. 利用一种修正策略保证了极限点的存在性, 得到了极限点为非负矩阵分解问题的稳定点. 此外, 给出了推广的修正策略. 数值实验结果表明上述修正策略是有效的.

关键词: 非负矩阵分解; 交替非负最小二乘法; 修正策略

MR(2010)主题分类号: 49M37 中图分类号: O224