# A NONCONVEX SPLITTING METHOD FOR SYMMETRIC NONNEGATIVE MATRIX FACTORIZATION: CONVERGENCE ANALYSIS AND OPTIMALITY

Songtao Lu<sup>†</sup>, Mingyi Hong<sup>‡</sup>, and Zhengdao Wang<sup>†</sup>

<sup>†</sup>Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011, USA <sup>‡</sup>Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA 50011, USA

# ABSTRACT

Symmetric non-negative matrix factorization (SymNMF) has important applications in data analytics problems such as document clustering, community detection and image segmentation. In this paper, we propose a novel nonconvex variable splitting method for solving SymNMF. Different from the existing works, we prove that the algorithm converges to the set of Karush-Kuhn-Tucker (KKT) points of the nonconvex SymNMF problem with a global sublinear convergence rate. We also show that the algorithm can be efficiently implemented in a distributed manner. Further, we provide sufficient conditions that guarantee the global and local optimality of the obtained solutions. Extensive numerical results performed on both synthetic and real data sets suggest that the proposed algorithm yields high quality of the solutions and converges quickly to the set of local minimum solutions compared with other algorithms.

*Index Terms*— Symmetric Nonnegative Matrix Factorization, Karush-Kuhn-Tucker points, variable splitting, global and local optimality, clustering

#### 1. INTRODUCTION

Recently, symmetric non-negative matrix factorization (SymNMF) has found many applications in document clustering, community detection, image segmentation and pattern clustering in bioinformatics [1–3]. SymNMF is not only able to deal with the non-linearly separable data points, such as image data, but also capture the inherent data structure with the graph representation [1]. Mathematically, SymNMF approximates a given (usually symmetric) non-negative matrix  $\mathbf{Z} \in \mathbb{R}^{N \times K}$  is component-wise non-negative, typically with  $K \ll N$ . Such problem can be formulated as the following nonconvex optimization problem [1,3,4]:

$$\min_{\mathbf{X} \ge 0} \quad f(\mathbf{X}) = \frac{1}{2} \|\mathbf{X}\mathbf{X}^T - \mathbf{Z}\|_F^2 \tag{1}$$

where  $\|\cdot\|_F$  denotes the Frobenius norm.

Due to the importance of the SymNMF problem, many algorithms have been proposed in the literature for finding its high-quality solutions. To this end, first rewrite the SymNMF equivalently as

$$\min_{\mathbf{Y} \ge 0, \mathbf{X} = \mathbf{Y}} \frac{1}{2} \| \mathbf{X} \mathbf{Y}^{T} - \mathbf{Z} \|_{F}^{2}.$$
 (2)

A simple strategy is to ignore the equality constraint  $\mathbf{X} = \mathbf{Y}$ , and then alternatingly perform the following two steps: 1) solving  $\mathbf{Y}$  with  $\mathbf{X}$  being fixed (a non-negative least squares problem); 2) solving  $\mathbf{X}$  with  $\mathbf{Y}$  being fixed (a least squares problem). Such ANLS

algorithm has been proposed in [1] for dealing with SymNMF. Unfortunately, despite the fact that an optimal solution can be obtained in each subproblem, there is no guarantee that the Y-iterate will converge to the X-iterate. The algorithm in [1] adds a regularized term for the difference between the two factors to the objective function and explicitly enforces that the two matrices be equal at the output. While such extra step enforces symmetry, it destroys the optimality of the subproblems, resulting in the loss of global convergence guarantee. A related ANLS-based method has been introduced in [4], however the algorithm is based on the assumption that there exists an exact symmetric factorization (i.e.,  $\exists \mathbf{X} \geq 0$  such  $\mathbf{X}\mathbf{X}^T = \mathbf{Z}$ ). Without such assumption, the algorithm may not converge to Karush-Kuhn-Tucker (KKT) points of (1). A multiplicative update for SymNMF has been proposed in [3], but the algorithm not only lacks convergence guarantee (to KKT points of (1)) [5], but has a much slower convergence speed than the one proposed in [4]. In [1, 6], algorithms based on the projected gradient descent (PGD) and the projected Newton (PNewton) have been proposed, both of which directly solve the original formulation (1). However, there has been no global convergence analysis since the objective function is a nonconvex fourth-order polynomial. The more recent work applies the nonconvex coordinate descent (CD) algorithm for SymNMF, with no guarantee that the algorithm will converge to a stationary point of problem (1), since the minimizer of the fourth order polynomial is not unique in each coordinate updating [7].

An important research question for SymNMF is whether it is possible to design algorithms that lead to *globally* optimal solutions. At the first sight such problem appears very challenging since finding the exact SymNMF is NP-hard [8]. However, some promising recent findings suggest that when the structure of the underlying factors are appropriately utilized, it is possible to obtain rather strong results. For example, in [9], the authors have shown that for the low rank factorized stochastic optimization problem where the two low rank matrices are symmetric, a modified stochastic gradient descent algorithm is capable of converging to a global optimum with constant probability from a random starting point. Related works include [10, 11]. However, when the factors are required to be non-negative, it is no longer clear whether the existing analysis can still be used to show convergence to global optimality, even local optimality (a milder result). For the non-negative principle component problem (that is, finding the leading non-negative eigenvector, i.e., K = 1), under the spiked model, reference [12] shows that certain approximate message passing algorithm is able to find the global optimal solution asymptotically. Unfortunately, this analysis does not generalize to an arbitrary symmetric observation matrix and any K. To our best knowledge, there is a lack of characterization of global and local optimal solutions for the SymNMF problem.

In this paper, we focus on analyzing the SymNMF problem. First, we propose a novel algorithm for the SymNMF, which utilizes certain nonconvex splitting technique and is capable of converging to the set of KKT points with provable global convergence rate. The main idea of the algorithm is to relax the symmetry requirement at the beginning and gradually enforce it as the algorithm proceeds. Second, we provide a number of easy-to-check sufficient conditions guaranteeing the local or global optimality of the obtained solutions. Numerical results on both synthetic and real data show that the proposed algorithm achieves fast and stable convergence (often to local minimum solutions) with low computational complexity.

More specifically, the main contributions of this paper are:

1) We design a novel algorithm, named the nonconvex splitting SymNMF (NS-SymNMF), which converges to the set of KKT solutions of SymNMF with a global sublinear rate. To our best knowledge, it is the first SymNMF solver that possesses provable global convergence rate.

2) We provide a set of easily checkable sufficient conditions (which only involve finding the smallest eigenvalue of certain matrix) that characterize the global and local optimality of the SymNMF. By utilizing such conditions, we demonstrate numerically that with high probability, our proposed algorithm converges not only to a KKT point but to a local optimal solution as well.

Due the limited space of the paper, all proofs are omitted and will be included in the journal version.

#### 2. THE PROPOSED ALGORITHM

The proposed algorithm leverages the reformulation (2). Our main idea is to gradually tighten the difficult equality constraint  $\mathbf{X} = \mathbf{Y}$  as the algorithm proceeds so that when convergence is approached, such equality is eventually satisfied. To this end, let us construct the augmented Lagrangian for (2), given by

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}; \mathbf{\Lambda}) = \frac{1}{2} \|\mathbf{X}\mathbf{Y}^{T} - \mathbf{Z}\|_{F}^{2} + \langle \mathbf{Y} - \mathbf{X}, \mathbf{\Lambda} \rangle + \frac{\rho}{2} \|\mathbf{Y} - \mathbf{X}\|_{F}^{2}$$
(3)

where  $\mathbf{\Lambda} \in \mathbb{R}^{N \times K}$  is a matrix of dual variables;  $\langle \cdot \rangle$  denotes the inner product operator;  $\rho > 0$  is a penalty parameter whose value will be determined later.

At this point, it may be tempting to directly apply the well-known alternating direction method of multipliers (ADMM) method [13] to the augmented Lagrangian (3), which alternatingly minimizes the primal variables **X**, **Y**, followed by a dual ascent step  $\Lambda \leftarrow \Lambda + \rho(\mathbf{Y} - \mathbf{X})$ . Unfortunately, the classical result for ADMM presented in [13–15] only works for convex problems, hence they do not apply to our nonconvex problem (2) (note this is a linearly constrained *nonconvex* problem where the nonconvexity arises in the objective function). Recent results such as [16–18] for analyzing ADMM for nonconvex problems do not apply either, because in these works the basic requirements are: 1) the objective function is separable over the block variables; 2) the smooth part of the augmented Lagrangian function has Lipschitz continuous gradient with respect to all variable blocks. Unfortunately none of these conditions are satisfied in our problem.

Next we begin presenting the proposed algorithm. We start by considering the following reformulation of problem (1)

$$\min_{\mathbf{X},\mathbf{Y}} \quad \frac{1}{2} \|\mathbf{X}\mathbf{Y}^{T} - \mathbf{Z}\|_{F}^{2} \qquad (4)$$
s.t.  $\mathbf{Y} \ge 0, \ \mathbf{X} = \mathbf{Y}, \ \|\mathbf{Y}_{i}\|_{2}^{2} \le \tau, \ \forall i,$ 

where 
$$\mathbf{Y}_i$$
 denotes the *i*th row of the matrix  $\mathbf{Y}$ ;  $\tau > 0$  is some  
given constant. It is easy to check that when  $\tau$  is sufficiently large  
(with a lower bound dependent on  $\mathbf{Z}$ ), then problem (4) is *equivalent*  
to problem (1), in the sense that the KKT points  $\mathbf{X}^*$  of the two  
problems are identical, which satisfy the optimality conditions giver  
by [19, Proposition 2.1.2]

$$\langle (\mathbf{X}^* (\mathbf{X}^*)^T - \mathbf{Z}) \mathbf{X}^*, \mathbf{X} - \mathbf{X}^* \rangle > 0, \quad \forall \mathbf{X} > 0.$$
 (5)

To be precise, we have the following result.

**Lemma 1.** Suppose  $\tau \geq 2 ||\mathbf{Z}||_F$ , then the global optimal solutions as well as the KKT points of the problems (1) and (4) have a one-to-one correspondence.

The proposed algorithm, named the nonconvex splitting SymNMF (NS-SymNMF), alternates between the primal updates of variables  $\mathbf{X}$  and  $\mathbf{Y}$ , and the dual update for  $\mathbf{\Lambda}$ . Below we present its detailed steps (superscript *t* is used to denote the iteration number).

$$\mathbf{Y}^{(t+1)} = \arg \min_{\mathbf{Y} \ge 0, \|\mathbf{Y}_i\|_2^2 \le \tau, \forall i} \frac{1}{2} \|\mathbf{X}^{(t)}\mathbf{Y}^T - \mathbf{Z}\|_F^2 + \frac{\rho}{2} \|\mathbf{Y} - \mathbf{X}^{(t)} + \mathbf{\Lambda}^{(t)} / \rho\|_F^2 + \frac{\beta^{(t)}}{2} \|\mathbf{Y} - \mathbf{Y}^{(t)}\|_F^2, \quad (6)$$

$$\mathbf{X}^{(t+1)} = \arg\min_{\mathbf{X}} \frac{1}{2} \| \mathbf{X} (\mathbf{Y}^{(t+1)})^{T} - \mathbf{Z} \|_{F}^{2} + \frac{\rho}{2} \| \mathbf{X} - \mathbf{\Lambda}^{(t)} / \rho - \mathbf{Y}^{(t+1)} \|_{F}^{2},$$
(7)

$$\mathbf{\Lambda}^{(t+1)} = \tilde{\mathbf{\Lambda}}^{(t)} + \rho(\mathbf{Y}^{(t+1)} - \mathbf{X}^{(t+1)}), \tag{8}$$

$$\beta^{(t+1)} = \frac{6}{\rho} \| \mathbf{X}^{(t+1)} (\mathbf{Y}^{(t+1)})^T - \mathbf{Z} \|_F^2.$$
<sup>(9)</sup>

We remark that this algorithm is very close in form to the standard ADMM method applied to problem (4) (which lacks convergence guarantees). The key difference is the use of the proximal term  $||\mathbf{Y} - \mathbf{Y}^{(t)}||_F^2$  multiplied by an *iteration dependent* penalty parameter  $\beta^{(t)} \geq 0$ , whose value is proportional to the size of the objective value. Intuitively, if the algorithm converges to a solution with small objective value (which appears to be often the case in practice), then the parameter  $\beta^{(t)}$  vanishes in the limit. It turns out that introducing such proximal term is crucial in guaranteeing the convergence of NS-SymNMF.

#### 3. CONVERGENCE ANALYSIS

In this section we provide convergence analysis result of the NS-SymNMF.

#### 3.1. Convergence and Convergence Rate Analysis

Below we present our first main result, which says that when the penalty parameter  $\rho$  is sufficiently large, the NS-SymNMF algorithm converges globally to the set of KKT point of (1).

Theorem 1. Suppose the following is satisfied

$$\rho > 6N\tau. \tag{10}$$

Then the following statements are true for NS-SymNMF:

1. The equality constraint is satisfied in the limit, i.e.,

$$\lim_{t \to \infty} \|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)}\| \to 0.$$

# 2. The sequence $\{\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}\}$ is bounded, and every one of its limit point is a KKT point of problem (1).

Our second result characterizes the convergence rate of the algorithm. To this end, we need to construct a function that measures the optimality of the iterates  $\{\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}\}$ . Define the *proximal gradient* of the augmented Lagrangian function as

$$\widetilde{\nabla}\mathcal{L}(\mathbf{X},\mathbf{Y},\mathbf{\Lambda}) \triangleq \begin{bmatrix} \mathbf{Y}^{T} - \operatorname{Proj}_{\mathbf{Y}}[\mathbf{Y}^{T} - \nabla_{\mathbf{Y}}(\mathcal{L}(\mathbf{Y},\mathbf{X},\mathbf{\Lambda})] \\ \nabla_{\mathbf{X}}\mathcal{L}(\mathbf{X},\mathbf{Y},\mathbf{\Lambda}) \end{bmatrix}$$
(11)

where the operator  $\operatorname{Proj}_{\mathbf{Y}} \triangleq \operatorname{Proj}_{\mathbf{Y} \ge 0, \|\mathbf{Y}_i\|_2^2 \le \tau, \forall i}$ , i.e., it is the projection operator onto the feasible set of  $\mathbf{Y}$ . Here we propose to use the following quantity to measure the progress of the algorithm

$$\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) \triangleq \|\widetilde{\nabla}\mathcal{L}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})\|_{F}^{2} + \|\mathbf{X}^{(t)} - \mathbf{Y}^{(t)}\|_{F}^{2},$$
(12)

It can be verified that  $\lim_{t\to\infty} \mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) = 0$ , then a KKT point of the problem (1) is obtained.

Below we show that the function  $\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})$  reduces to zero in a sublinear manner.

**Theorem 2.** For a given small constant  $\epsilon$ , let  $T(\epsilon)$  denote the iteration index satisfying the following inequality

$$T(\epsilon) \triangleq \min\{t \mid \mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)}) \le \epsilon, t \ge 0\}.$$
(13)

Then there exists some constant C > 0 such that

$$\epsilon \le \frac{C\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{\Lambda}^{(1)})}{T(\epsilon)}.$$
(14)

The above result says that in order for  $\mathcal{P}(\mathbf{X}^{(t)}, \mathbf{Y}^{(t)}, \mathbf{\Lambda}^{(t)})$  to reach below  $\epsilon$ , it takes  $\mathcal{O}(1/\epsilon)$  number of iterations. It follows that NS-SymNMF converges sublinearly.

#### 3.2. Sufficient Global and Local Optimality Conditions

Since problem (1) is not convex, the KKT points obtained by NS-SymNMF could be different from the global optimal solutions. Therefore it is important to characterize the conditions under which these two different types of solutions coincide. Below we provide an easily checkable sufficient condition to ensure that a KKT point  $(\mathbf{X}^*, \mathbf{\Omega}^*)$  is also a globally optimal solution for problem (1) (where  $\mathbf{\Omega}$  is the dual matrix for the constraint  $\mathbf{X} \ge 0$  in problem (1)).

**Theorem 3.** Suppose that  $(\mathbf{X}^*, \mathbf{\Omega}^*)$  is a KKT point of (1). Then,  $\mathbf{X}^*$  is also a global optimal point if the following is satisfied

$$\mathbf{S} \triangleq \mathbf{X}^* (\mathbf{X}^*)^T - \frac{\mathbf{Z}^T + \mathbf{Z}}{2} \succeq 0.$$
(15)

Admittedly, the condition given in Theorem 3 could be strong hence may not be satisfied for some problem instances. In this section we provide a milder condition which guarantees that a KKT point to be *locally optimal*. Such results are also very useful in practice since they can help identify spurious saddle points such as  $\mathbf{X} = \mathbf{0}$  if  $\mathbf{Z}^T + \mathbf{Z}$  is not negative semidefinite.

We have the following characterization of the local optimal solution of the SymNMF problem.

**Theorem 4.** Suppose that  $(\mathbf{X}^*, \mathbf{\Omega}^*)$  is a KKT point of (1). Define a matrix  $\mathcal{T} \in \mathbb{R}^{KN \times KN}$  whose (m, n)th block is a matrix of size  $N \times N$ 

$$\mathcal{T}_{m,n} \triangleq \left( (\mathbf{X}_m^{\prime*})^T \mathbf{X}_n^{\prime*} - \delta \|\mathbf{X}_n^{\prime*}\|_2^2 \right) \mathbf{I} + \mathbf{X}_n^{\prime*} (\mathbf{X}_m^{\prime*})^T + \delta_{m,n} \mathbf{S},$$
(16)

where **S** is defined in (15),  $\delta_{m,n}$  is the Kronecker delta function, and  $\mathbf{X}'_m^*$  denotes the mth column of  $\mathbf{X}^*$ . If there exists some  $\delta > 0$  such that  $\mathcal{T} \succ 0$ , then there exists some  $\epsilon > 0$  small enough such that  $(\mathbf{X}^*, \mathbf{\Omega}^*)$  is a local minimum point of (1), meaning that for all  $\mathbf{X} \ge 0$  satisfying  $\|\mathbf{X} - \mathbf{X}^*\|_F \le \epsilon$ , we have

$$f(\mathbf{X}) \ge f(\mathbf{X}^*) + \frac{\gamma}{2} \|\mathbf{X} - \mathbf{X}^*\|_F^2.$$
(17)

*Here the constant*  $\gamma > 0$  *is given by* 

$$\gamma = -\left(\frac{2K^2}{\delta} + K(K-2)\right)\epsilon^2 + 2\lambda_{\min}(\mathcal{T}) > 0 \qquad (18)$$

where  $\lambda_{\min}(\mathcal{T}) > 0$  is the smallest eigenvalue of  $\mathcal{T}$ .

In the special case of K = 1, the sufficient condition set forth in Theorem 4 can be significantly simplified.

**Corollary 1.** Suppose that  $(\mathbf{x}^*, \mathbf{\Omega}^*)$  is the KKT point of (1) when K = 1. If there exists some  $\delta > 0$  such that

$$\mathcal{T}_{1} \triangleq (1-\delta) \|\mathbf{x}^{*}\|_{2}^{2} \mathbf{I} + 2\mathbf{x}^{*} (\mathbf{x}^{*})^{T} - \frac{\mathbf{Z}^{T} + \mathbf{Z}}{2} \succ 0.$$
(19)

Then  $\mathbf{X}^*$  is a local minimum point of (1).

We comment that the condition given in Theorem 4 is much milder than that in Theorem 3. Further such condition is also very easy to check as it only involves finding the smallest eigenvalue of a  $KN \times KN$  matrix for each value of  $\delta^{-1}$ . In our numerical result (to be presented shortly), we set  $\delta$  to a series of consecutive values when performing the test. We have observed that the solutions generated by the proposed NS-SymNMF algorithm satisfy the condition provided in Theorem 4 with high probability.

# 4. IMPLEMENTATION OF THE PROPOSED ALGORITHM

#### 4.1. The Y-Subproblem

Y

To solve the  $\mathbf{Y}$ -subproblem (6), we use the gradient projection (GP) method. We define

$$\mathbf{Z}_{\mathbf{Y}}^{(t)} \triangleq (\mathbf{X}^{(t)})^{T} \mathbf{Z} + \rho(\mathbf{X}^{(t)})^{T} - (\mathbf{\Lambda}^{(t)})^{T} + \beta^{(t)} (\mathbf{Y}^{(t)})^{T}, \quad (20)$$
$$\mathbf{A}_{\mathbf{Y}}^{(t)} \triangleq (\mathbf{X}^{(t)})^{T} \mathbf{X}^{(t)} + (\rho + \beta^{(t)}) \mathbf{I}. \quad (21)$$

Performing the  $\mathbf{Y}^{(t+1)}$  update in (6) is equivalent to solving

$$\min_{\|\mathbf{y}\| \ge 0, \|\mathbf{Y}_i\|_2^2 \le \tau, \forall i} \sum_{i=1}^N \|\mathbf{A}_{\mathbf{Y}}^{(t)} \mathbf{Y}_i^T - \mathbf{Z}_{\mathbf{Y}, i}^{(t)}\|_2^2$$
(22)

where  $\mathbf{Z}_{\mathbf{Y},i}$  is the *i*th column of matrix  $\mathbf{Z}_{\mathbf{X}}$ .

This problem can be decomposed into N separable constrained least squares problems, each of which can be solved independently, therefore can be implemented in parallel. Here we use the conventional gradient projection for solving each subproblem, which generates a sequence by

$$\mathbf{Y}_i^{(t+1)} = \operatorname{Proj}_{\mathbf{Y}_i} [\mathbf{Y}_i^{(t)} - \alpha (\mathbf{A}_{\mathbf{Y}}^{(t)})^T (\mathbf{A}_{\mathbf{Y}}^{(t)} \mathbf{Y}_i^{(t)} - \mathbf{Z}_{\mathbf{Y},i}^{(t)})].$$

In the above expression,  $\alpha$  is the step size, which is chosen either as a constant  $1/\lambda_{\max}((\mathbf{A}_{\mathbf{Y}}^{(t)})^T \mathbf{A}_{\mathbf{Y}}^{(t)})$ , or by using some line search

<sup>&</sup>lt;sup>1</sup>To find such smallest eigenvalue, we can find the largest eigenvalue of  $\eta \mathbf{I} - \mathcal{T}$ , using algorithms such as the power method [20], where  $\eta$  is sufficient large based on  $\tau$  and  $\|\mathbf{Z}\|_{F}$ .



Fig. 1. The convergence behaviors of different SymNMF solvers on both synthetic and real data sets and local optimality condition check.

procedure [19]; for a given vector  $\mathbf{w}$ ,  $\operatorname{Proj}_{\mathbf{Y}_i}[\mathbf{w}]$  denotes the projection of it to the feasible set of  $\mathbf{Y}_i$ , which can be evaluated in closed-form [21, pp. 80]

$$\mathbf{w}^{+} = \operatorname{Proj}_{+}(\mathbf{w}) \triangleq \max\{\mathbf{w}, \mathbf{0}_{K \times 1}\},$$
(23)

$$\mathbf{Y}_{i} = \operatorname{Proj}_{\|\mathbf{w}^{+}\|_{2}^{2} \leq \tau}(\mathbf{w}^{+}) \triangleq \sqrt{\tau} \mathbf{w}^{+} / \max\{\sqrt{\tau}, \|\mathbf{w}^{+}\|_{2}\}.$$
(24)

We comment that other algorithms such as the accelerated version of the gradient projection [22] can also be used to solve the subproblem (22). Here we pick GP for its simplicity.

### 4.2. The X-Subproblem

The subproblem for updating  $\mathbf{X}^{(t+1)}$  in (7) is equivalent to the following problem

$$\min_{\mathbf{X}} \|\mathbf{Z}_{\mathbf{X}}^{(t+1)} - \mathbf{X}\mathbf{A}_{\mathbf{X}}^{(t+1)}\|_{F}^{2}$$
(25)

where  $\mathbf{Z}_{\mathbf{X}}^{(t+1)} \triangleq \mathbf{Z}\mathbf{Y}^{(t+1)} + \mathbf{\Lambda}^{(t)} + \rho\mathbf{Y}^{(t+1)}$  and  $\mathbf{A}_{\mathbf{X}}^{(t+1)} \triangleq (\mathbf{Y}^{(t+1)})^T\mathbf{Y}^{(t+1)} + \rho\mathbf{I}$  are two fixed matrices. This is just a least-squares problem and can be solved in closed-form. The solution is given by  $\mathbf{X}^{(t+1)} = \mathbf{Z}_{\mathbf{X}}^{(t+1)}(\mathbf{A}_{\mathbf{X}}^{(t+1)})^{-1}$ . We remark that the  $\mathbf{A}_{\mathbf{X}}^{(t+1)}$  is a  $K \times K$  matrix, where K is usually small (e.g., the number of clusters for graph clustering applications). As a result, the matrix inversion of  $\mathbf{A}_{\mathbf{X}}^{(t+1)}$  is computationally cheap.

#### 5. NUMERICAL RESULTS

In this section we present the performance of the proposed NS-SymNMF algorithm, compare it with other existing algorithms, and test it on problems such as community detection.

Synthetic Data Set (Random Symmetric Matrices): We randomly generate a nonnegative matrix  $\mathbf{M}$  with i.i.d. Gaussian entries. We then generate  $\mathbf{Z} = \mathbf{M}\mathbf{M}^T$  as the given matrix to be decomposed.

*Real Data Set (An Adjacency Matrix):* We also implement the algorithm on a real data set in a clustering application. The data were collected by Facebook from survey participants [23]. The vertices  $(x_i, i = 1, ..., N)$  represent different individuals, and edges denote the relationship between friends on Facebook. If two individuals (e.g.,  $x_i, x_j$ ) are friends, then  $\mathbf{S}'_{i,j} = 1$ . Otherwise, we set  $\mathbf{S}'_{i,j} = 0$ . The network has 333 nodes and 5374 edges after data cleansing such that the degree of each node is greater than 1. Since there is no ground truth, a quantity called modularity [24] is adopted to measure the strength of division of the network.

The convergence behaviors of NS-SymNMF and other existing SymNMF solvers based on the synthetic data are shown in Figure 1(a) where the size of M is  $50 \times 4$ , K = 4, and the results are averaged over 100 Monte Carlo (MC) trials. The results with

Facebook data are shown in Figure 1(b) where there are K = 18communities considered. All tests are performed using Matlab on a computer with Intel Core i3-2350M CPU running at 2.30GHz with 4GB RAM. The step size of PGD is  $1 \times 10^{-3}$ . For NS-SymNMF, we let  $\tau = 2 \|\mathbf{Z}\|_F$  and the maximum number of iterations of GP be 40. Also we only update  $\beta^{(t)}$  once every 100 iterations for saving the CPU time, and gradually increase the value of  $\rho$  from an initial value  $(\rho = 20)$  with  $\rho \leftarrow 1.01\rho$  to meet condition (10) for accelerating the convergence rate [13, Section 3.4.1]. In Figure 1(a), it can be observed that NS-SymNMF and SNMF [4] can achieve the global optimal solution with a short time. In Figure 1(b), we can see that there is a constant gap between SNMF and the rest of methods, since the SNMF algorithm transforms the original problem to another one under the assumption that  $\mathbf{Z}$  can be exactly decomposed by  $\mathbf{X}\mathbf{X}^{T}$ . Levering the dual update for X and Y, NS-SymNMF also shows a faster convergence rate and lower objective value than ANLS [1] which only adds certain penalty term for the difference between the factors (X and Y). The drawbacks of PGD [6], PNewton [6] and CD [7] compared with NS-SymNMF are as follows. There is no rule of choosing the step-size of the PGD algorithm because of the unboundness of Lipschitz continuous gradient. PNewton has high per-iteration complexity due to the requirement of computing the Hessian matrix at each iteration. CD method updates each entry of X cyclicly such that the objective function can be only decreased locally by the updating entry at a time. Also, CD cannot be implemented in a parallel way. From the modularity result depicted in Figure 1(c), it can be observed that the proposed algorithm gives an accurate decomposition more quickly compared with other algorithms.

After the NS-SymNMF algorithm is converged, the optimality can be checked according to Theorem 4. To find an appropriate  $\delta$ that satisfies the condition  $\lambda_{\min}(\mathcal{T}) > 0$ , we initialize  $\delta$  as 1 and decrease it by 0.01 each time and check the minimum eigenvalue of  $\mathcal{T}$ . The satisfiability results are shown in Figure 1(d) based on 100 MC trials for the synthetic data set where the size of **M** is  $500 \times 6$  and K = 4. There always exists a  $\delta$  such that  $\mathcal{T}$  is positive definite in all cases that we have tested. This indicates that (with high probability) the proposed algorithm finds a locally optimal solution.

# 6. CONCLUSIONS

In this paper, we propose a nonconvex splitting algorithm for solving the SymNMF problem. We show that the proposed algorithm converges to the set of KKT points in a sublinear manner. Further, we provide sufficient conditions to identify global or local optimal solutions of the SymNMF problem. Numerical experiments show that the proposed method can converge quickly to local optimal solutions.

#### 7. REFERENCES

- D. Kuang, S. Yun, and H. Park, "SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering," *Journal of Global Optimization*, vol. 62, no. 3, pp. 545–574, July 2015.
- [2] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493–521, May 2011.
- [3] Z. He, S. Xie, R. Zdunek, G. Zhou, and A. Cichocki, "Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2117–2131, Dec. 2011.
- [4] K. Huang, N. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 211–224, Jan. 2014.
- [5] C. J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1589–1596, Nov. 2007.
- [6] D. Kuang, C. Ding, and H. Park, "Symmetric nonnegative matrix factorization for graph clustering," in *Proc. of SIAM Int. Conf. Data Mining*, pp. 106–117, 2012.
- [7] A. Vandaele, N. Gillis, Q. Lei, K. Zhong, and I. Dhillon, "Efficient and non-convex coordinate descent for symmetric nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5571–5584, Nov 2016.
- [8] S. A. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1364–1377, 2009.
- [9] C. D. Sa, C. Re, and K. Olukotun, "Global convergence of stochastic gradient descent for some non-convex matrix problems," in *Proc. of the 32nd International Conference on Machine Learning (ICML)*, pp. 2332–2341, 2015.
- [10] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Transactions on Information Theory*, 2016.
- [11] T. Zhao, Z. Wang, and H. Liu, "A nonconvex optimization framework for low rank matrix estimation," in *Proc. of Neural Information Processing Systems (NIPS)*, pp. 559–567, 2015.
- [12] A. Montanari and E. Richard, "Non-negative principal component analysis: Message passing algorithms and sharp asymptotics," *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1458–1484, Mar. 2016.
- [13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [14] D. P. Bertsekas, P. Hosein, and P. Tseng, "Relaxation methods for network flow problems with convex arc costs," *SIAM Journal on Control and Optimization*, vol. 25, no. 5, pp. 1219–1243, Sep. 1987.
- [15] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, 1992.

- [16] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.
- [17] G. Li and T.-K. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2434–2460, 2015.
- [18] Y. Wang and J. Z. W. Yin, "Global convergence of ADMM in nonconvex nonsmooth optimization," arXiv Preprint, arXiv:1511.06324, 2015.
- [19] D. P. Bertsekas, Nonlinear Programming, 2nd ed, Athena Scientific, Belmont, MA, 1999.
- [20] S. Lu and Z. Wang, "Accelerated algorithms for eigen-value decomposition with application to spectral clustering," in *Proc.* of 2015 49th Asilomar Conference on Signals, Systems and Computers, pp. 355–359, Nov. 2015.
- [21] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, Springer Science & Business Media, 2007.
- [22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imgaging Science*, vol. 2, no. 1, pp. 183–202, 2009.
- [23] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Proc. of Neural Information Processing Systems (NIPS)*, 2012.
- [24] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Aacademy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.