Orthogonal symmetric nonnegative matrix trifactorization

DACHE Alexandra
Université de Mons
Mons, Belgium
alexandra.dache@student.umons.ac.be

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—TO DO**

***Index Terms*—nonnegative matrix factorization (NMF), orthogonal factorization, clustering, symetric tri-factorization**

I. Introduction

Symmetric Non-Negative Matrix Trifactorization (tri-symNMF) is a model derived from NMF (Non-Negative Matrix Factorization) used in community detection in a network where communities can interact with each other.This approach finds other practical applications in topic modeling, and modeling hidden Markov chains. For more details and applications, see [1].

Given a positive symmetric matrix $X \in \mathbb{R}^{n\times n}$ and a rank $r \in \mathbb{N}$, the problem involves finding a positive matrix $W \in \mathbb{R}^{n\times r}$ and a positive symmetric matrix $S \in \mathbb{R}^{r\times r}$ such that :

$$X \approx WSW^T \tag{1}$$

The input matrix X is symmetric and can be viewed as the adjacency matrix of a graph, where the entry $X(i,j) = X(j,i)$ indicates the weight of the connection between node i and j. Consequently, this factorization aims to identify r communities and their interactions. For each element of the matrix X, we have:

$$X(i,j) \approx W(i,:)SW(:,j)^T = \sum_{k=1}^{r}\sum_{l=1}^{r} W(i,k)S(k,l)W(l,j)^T \tag{2}$$

The columns of W identify the communities, i.e., strongly correlated elements in the dataset. If $W(j,k)$ is non-zero, element j belongs to community k. The matrix S enables interactions between communities, where entry $S(k,l)$ represents the strength of the connection between communities k and l.

In this work, we focus on finding disjoint communities, meaning that each node belongs to only one community. Mathematically, this constraint translates to ensuring that there is at most one non-zero element per row of W. To satisfy this constraint, it can be imposed that W is column-wise orthogonal.Indeed, with the non-negativity of H, orthogonality implies that each line of W has a single positive entry. To measure the error, the Frobenius norm is utilized. In summary, the problem can be formulated as follows:

$$\min_{W\geq 0, S\geq 0} \|X - WSW^T\|_F^2 \quad \text{s.t.} \quad W^TW = I, \tag{3}$$

Identify applicable funding agency here. If none, delete this.

II. Uniqueness

In conjunction with the nonnegativity property of W, column-wise orthogonality ensures that every row of W contains only one positive entry.

Lemma 2.1: $W \in \mathbb{R}_+^{n\times r}$ satisfy $W^TW = I$. Then each row of W has at most a single positive entry

Proof. $W^TW = I$ implies that the columns of W are orthogonal to each other. However, two nonnegative vectors are orthogonal if and only if they have disjoint supports. Therefore, the columns of W must have disjoint supports, meaning that each row of W contains only one non-zero element.

Thanks to the lemma 2.1, we can simplify the expression 2. By denoting k_i and l_j as the indices of the non-zero element in rows i and j of W, respectively, we obtain:

$$X(i,j) \approx W(i,k_i)S(k_i,l_j)W(l_j,j)^T \tag{4}$$

Theorem 2.2: Let $X = WSW^T$ where $W \in \mathbb{R}_+^{n\times r}$ satisfies $W^TW >= I$ and the columns of $S \in \mathbb{R}_+^{r\times r}$ Then the exact orthogonal symmetric nonnegative matrix trifactorization of X of size r is unique.

Proof. Suppose X is expressed as the product of matrices F and G, where $S \in \mathbb{R}_+^{r\times n}$ and satisfies the condition $GG^T = Ir$, and the columns of $F \in \mathbb{R}_+^{m\times r}$ are non-collinear. In such a case, the exact orthogonal nonnegative matrix factorization (ONMF) of X with a size of r is guaranteed to be unique [1, p. 136] Assuming that $F = WS$ and $G = W^T$, it is necessary for the columns of the product WS to be non-multiples of each other. Since W $W^TW = I$, the condition is equivalent to ensuring that S does not have any multiple columns. Indeed, the columns of W are not multiples of each other, and each row contains only one non-zero element(lemma 2.1). The only way to have multiple columns would be if S contains multiple columns.

III. Coordinate descent method

Many iterative algorithms optimize alternately between the variable W with S fixed and the variable S with W fixed. The article [2] proposes an iterative algorithm by updating H and then S according to update formulas. The formulas involve computing significant matrix products and do not exploit the specific structure of W. Another article [3] proposes a method that adds two penalty terms based on α-divergence to incorporate the orthogonality of parameters. The proposed iterative method leverages lemma 2.1 by updating W row-wise. The calculations are simplified by working only with the non-zero elements of W.

The pseudocode for the coordinate descent for orthogonal symmetric nonnegative matrix trifactorization is given in Algorithm 1. W and S are initialized while adhering to the constraints of the problem. The initialization can be done either randomly or using a clustering method.(TODO)(line 1) Next, we update W and S alternatively a certain number of times or until the error is sufficiently low (line 6), or the algorithm no longer significantly improves the error (line 9). The update of W is detailed in the Algorithm 2. W is updated row-wise. For each row, we identify the non-zero element that results in the smallest error (line 2). To achieve this, it is sufficient to keep the other elements fixed and set the remaining elements of the respective row to zero. The problem reduces to finding the minima of a polynomial of the form $az^4 + bz^2 + cx$, which can be solved exactly using Cardano's formulas. The Algorithme 3 provides the update for S. S is updated element by element while keeping the other elements fixed (line 3). The problem boils down to finding the minima of a quadratic polynomial. Finally, we normalize the columns of W while updating S accordingly (lines 13 to 19).

Algorithm 1: Coordinate descent for OTriSymNMF

Input: Input matrix $X \in \mathbb{R}^{n\times n}$, rank $r \in \mathbb{N}$
Output: Matrix $W \in \mathbb{R}^{n\times r}$ and a matrix $S \in \mathbb{R}^{r\times r}$
1: $W_0, S_0 \leftarrow$ initialisation(X,r)
2: **for** $t = 1$ to maxiter **do**
3: $W_t \leftarrow UpdateW(r, S_{t-1})$ see Algorithm 2
4: $S_t \leftarrow UpdateS(r, W_t)$ see Algoritm 3
5: $error = \|X - WSW^T\|_F^2$
6: **if** $error \simeq 0$ **then**
7: break;
8: **end if**
9: **if** $previous\_error - error \simeq 0$ **then**
10: break;
11: **end if**
12: **end for**
13: **for** $j = 1$ to n **do**
14: $W(i,j) = \frac{W(i,j)}{\|W(:,j)\|_2}, \forall i$
15: **for** k s.t. $W(k,j) \neq 0$ **do**
16: $S(k,l) = S(k,l) * \|W(:,j)\|_2$
17: $S(l,k) = S(l,k) * \|W(:,j)\|_2$
18: **end for**
19: **end for**

Algorithm 2: UpdateW

Input: Input matrix $X \in \mathbb{R}^{n\times n}$, rank $r \in \mathbb{N}$, matrix $S_{t-1} \in \mathbb{R}^{r\times r}$ and the matrix $W_{t-1} \in \mathbb{R}^{n\times r}$
Output: Matrix $W_t \in \mathbb{R}^{n\times r}$
1: **for** $i = 1$ to n **do**
2: $[\min_{k, W(i,k)\geq 0} \|X - WSW^T\|_F^2$ s.t. $W(i,j) = 0, \forall j \neq k] \rightarrow W_t(i,:)$
3: **end for**

Algorithm 3: UpdateS

Input: Input matrix $X \in \mathbb{R}^{n\times n}$, rank $r \in \mathbb{N}$, matrix $S_{t-1} \in \mathbb{R}^{r\times r}$ and the matrix $W_t \in \mathbb{R}^{n\times r}$
Output: Matrix $S_t \in \mathbb{R}^{r\times r}$
1: **for** $k = 1$ to r **do**
2: **for** $l = 1$ to r **do**
3: $[\min_{S_{i,j}\geq 0} \|X(i,:) - W(i,:)SW^T\|_F^2] \rightarrow S_t(i,j)$
4: **end for**
5: **end for**

IV. Experiments

TODO

V. Conclusion

TODO

References

[1] Nicolas Gillis. *Nonnegative Matrix Factorization*. Dec. 2020. ISBN: 978-1-61197-640-3. DOI: 10 . 1137 / 1 . 9781611976410.

[2] Chris Ding, Tao Li, Wei Peng, and Haesun Park. "Orthogonal nonnegative matrix t-factorizations for clustering". In: vol. 2006. Aug. 2006, pp. 126–135. DOI: 10. 1145/1150402.1150420.

[3] Saeid Hoseinipour, Mina Aminghafari, and Adel Mohammadpour. "Orthogonal parametric non-negative matrix tri-factorization with -divergence for co-clustering". In: *Expert Systems with Applications* 231 (2023), p. 120680. ISSN: 0957-4174. DOI: https : / / doi . org / 10 . 1016 / j . eswa . 2023 . 120680. URL: https : / / www . sciencedirect . com / science / article / pii / S095741742301182X.