## ADMM, ALM and LADMM for solving NMF\_l20

To better demonstrate that our optimization scheme is necessary, we compare against other ways of implementing and optimizing a L2,0 constraints, including Alternating Direction Method of Multipliers (ADMM) [1,2], Augmented Lagrangian Multiplier (ALM) [3] and Linearized Alternating Direction Method of Multipliers (LADMM) [4]. We compare the PALM and maPALM algorithms for solving NMF-L20 with ADMM, ALM, and LADMM algorithms.

## Alternating Direction Method of Multipliers (ADMM) Algorithm

We first show the row sparse NMF (NMF\_L20) model as follows

(1)

We use variable splitting in which a slack variable is introduced to get a new formulation

(2)

Inspired by ref. [1,2], we adopt the ADMM framework to solve the NMF\_L20 model. To use the ADMM algorithm, we construct the augmented Lagrangian function of Eq. (2) which is given by

(3)

where is the Lagrangian multipliers and is a penalty parameter. The ADMM algorithm iteratively updates the variables as follows:

(4)

(5)

(6)

(7)

The Eq. (4) and (5) are convex optimization problems, it is very clear how to update for and . Therefore, we omit the detailed steps here. For Eq. (6), the closed-form solution can be written as

(8)

The is defined in Eq. (18) in the main text. With all the sub-problems solved, the procedure of the ADMM algorithm is summarized in Algorithm 1. In Algorithm 1, regarding the value of ρ, we refer to [2].

**Algorithm 1 ADMM for solving the NMF\_L20 model**

Input: Data matrix , (nonzero rows, i.e., the number of features)

Parameter (Refer to [2])

Cluster amount

Output: and .

1: Initialize () and

2: repeat

3: Update H using (4)

4: Update W using (5)

5: Update E using (8)

6: Update B using (7)

7: until convergence

8: Return and .

## Augmented Lagrangian Multiplier (ALM) Algorithm

Inspired by ref. [3], we adopt the ALM framework to solve the NMF\_L20 model. Similar to the derivation of the ADMM algorithm, we first obtain the augmented Lagrangian function

(9)

The ALM method iteratively updates the variables as follows:

(10)

(11)

(12)

(13)

ALM generally uses a sequence of penalty parameters which are nondecreasing and possibly unbounded. The procedure of the ALM is summarized in Algorithm 2.

**Algorithm 2 ALM for solving the NMF\_L20 model**

Input: Data matrix , (nonzero rows, i.e., the number of features)

Cluster amount

Output: and .

1: Initialize () and

2: Initialize (Refer to [2])

3: repeat

4: Update H using (10)

5: Update W using (11)

6: Update E using (12)

7: Update B using (13)

8: Update

9: until convergence

10: Return and .

Note that the difference between ALM and ADMM algorithms is in step 8.

## Linearized ADMM (LADMM) Algorithm

Due to the non-convexity of L20-norm in the NMF\_L20 model, we adopt the LADMM framework to solve the NMF\_L20 model, which is inspired by ref. [4]. To use the LADMM method, we construct the augmented Lagrangian function as follows:

(14)

where is the Lagrangian multipliers. Compared with the ADMM framework, the LADMM framework applies the Proximal Gradient Descent (PGD) method for updating and in each iteration. Therefore, the LADMM algorithm iteratively updates the variables as follows:

(15)

(16)

(17)

(18)

The procedure of the LADMM is summarized in Algorithm 3.

**Algorithm 3 LADMM for solving the NMF\_L20 model**

Input: Data matrix , (nonzero rows, i.e., the number of features)

Parameter (Refer to [2])

Cluster amount

Output: and .

1: Initialize () and

2: repeat

3: Update H using (15)

4: Update W using (16)

5: Update E using (17)

6: Update B using (18)

7: until convergence

8: Return and .

## Reference

1. Wang Y, Yin W, Zeng J. Global convergence of ADMM in nonconvex nonsmooth optimization[J]. Journal of Scientific Computing, 2019, 78(1): 29-63.
2. Xu Y, Yin W, Wen Z, et al. An alternating direction algorithm for matrix completion with nonnegative factors[J]. Frontiers of Mathematics in China, 2012, 7(2): 365-384.
3. Cai X, Nie F, Huang H. Exact top-k feature selection via l2, 0-norm constraint[C]//Twenty-third international joint conference on artificial intelligence. 2013.
4. Liu Q, Shen X, Gu Y. Linearized ADMM for nonconvex nonsmooth optimization with convergence analysis[J]. IEEE Access, 2019, 7: 76131-76144.