Supplementary Material
for
Multi-View Clustering via Joint Nonnegative Matrix Factorization

Jialu Liu[1], Chi Wang[1], Jing Gao[2], and Jiawei Han[1]

[1]University of Illinois at Urbana-Champaign
[2]University at Buffalo

Computational Complexity Study

Recall that we conduct the experiments on a desktop with Intel Core I7 2600 and 16GB memory. The default setting is 10000 data points, 4 clusters, and 2 views. During the experiment, we fix two aspects and change the remaining one. Figure 1 shows the additional curves depicting the running time of MultiNMF and CoreguSC in terms of varying clusters and views.

Proof of Convergence

We prove that our MultiNMF algorithm converges to a local minimum.

The objective function O in Eq. 3.6 is bounded by zero. To prove the convergence of our algorithm, we need to show that the objective function is non-increasing under the updating steps in Eqs. Eq. 3.8 to Eq. 3.11. The update for V^* in Eq. 3.11 gives an exact analytical solution for the minimization of O when $U^{(v)}$ and $V^{(v)}$ are fixed. Therefore, we just need to prove that O keeps non-increasing when Eqs. 3.8 and 3.10 are applied.

Our proof will make use of an *auxiliary function* similar to what is used in the Expectation Maximization algorithm. We begin with the definition of the auxiliary function and a lemma on it [1].

Definition 1. *$G(A, A')$ is an auxiliary function for function $F(A)$ if the following conditions are satisfied: $G(A, A') \geq F(A), G(A, A) = F(A)$.*

Lemma 0.1. *If G is an auxiliary function of F, then F is non-increasing under the update*

$$A^{t+1} = \arg\min_A G(A, A^t) \tag{0.1}$$

where t means the t-th iteration.

Proof: $F(A^t) = G(A^t, A^t) \geq G(A^{t+1}, A^t) \geq F(A^{t+1})$. ∎

Now we will show that the update steps for $U^{(v)}$ and $V^{(v)}$ in Eqs. 3.8 and 3.10 are both the update in Eq. 0.1 with a proper auxiliary function, respectively.

We use $O^{(v)}_{i,k}$ and $\Theta^{(v)}_{k,j}$ to denote the part of O which is only relevant to $U^{(v)}_{i,k}$ and $V^{(v)}_{k,j}$ respectively. Since our updates are essentially element-wise, it is sufficient to show that each $O^{(v)}_{i,k}$ and $\Theta^{(v)}_{j,k}$ is non-increasing under Eqs. 3.8 and 3.10 respectively. For brevity, we omit the superscript (v) in O, Θ, X, U, V and Q in all the following lemmas and proofs because there is no cross-view entry in any of them. But we keep the subscript v in λ_v to imply that each formula contains an omitted superscript (v).

Lemma 0.2. *The function*

$$G(U_{i,k}, U^t_{i,k}) = O_{i,k}(U^t_{i,k}) + O'_{i,k}(U^t_{i,k})(U_{i,k} - U^t_{i,k}) + \left[\frac{(U^t V^T V)_{i,k}}{U^t_{i,k}} + \frac{\lambda_v \sum_{w=1}^M U^t_{w,k} \sum_{j=1}^N V^2_{j,k}}{U^t_{i,k}}\right](U_{i,k} - U^t_{i,k})^2$$

is an auxiliary function of $O_{i,k}$, where O' is the first-order derivative of $O_{i,k}$ on $U_{i,k}$.

Proof: It is obvious that $G(U^t_{i,k}, U^t_{i,k}) = O_{i,k}(U^t_{i,k})$. We just need to prove $G(U_{i,k}, U^t_{i,k}) \geq O_{i,k}(U_{i,k})$.

According to the definition, it is not hard to obtain the first-order and second-order derivative of $O_{i,k}$:

$$O'_{i,k}(U_{i,k}) = 2(UV^TV - XV)_{i,k} + 2\lambda_v(\sum_{w=1}^M U_{w,k} \sum_{j=1}^N V^2_{j,k} - \sum_{j=1}^N V_{j,k}V^*_{j,k})$$

$$O''_{i,k}(U_{i,k}) = 2(V^TV)_{k,k} + 2\lambda_v \sum_{j=1}^N V^2_{j,k}$$

Comparing G with the Taylor series expansion of $O_{i,k}$,

$$O_{i,k}(U_{i,k}) = O_{i,k}(U^t_{i,k}) + O'_{i,k}(U^t_{i,k})(U_{i,k} - U^t_{i,k}) + \frac{1}{2}O''_{i,k}(U^t_{i,k})(U_{i,k} - U^t_{i,k})^2$$

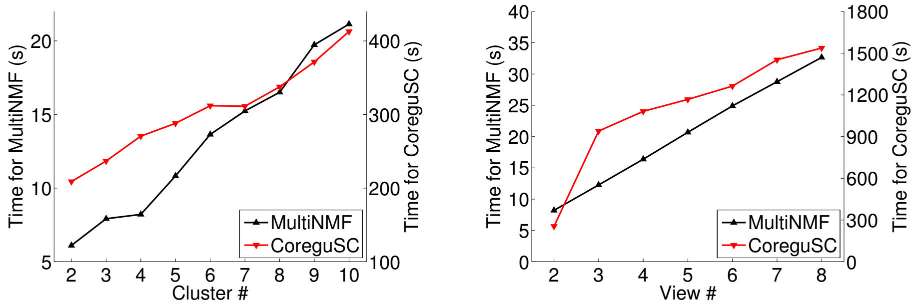

Figure 1: Running time of MultiNMF v.s. CoreguSC on synthetic dataset.

We can see that $G(U_{i,k}, U^t_{i,k}) \geq O_{i,k}(U_{i,k})$ is equivalent to:

$$\frac{(U^tV^TV)_{i,k}}{U^t_{i,k}} + \frac{\lambda_v \sum_{w=1}^M U^t_{w,k} \sum_{j=1}^N V^2_{j,k}}{U^t_{i,k}} \geq (V^TV)_{k,k} + \lambda_v \sum_{j=1}^N V^2_{j,k}$$

That is true because we have:

$$(U^tV^TV)_{i,k} = \sum_{l=1}^K U^t_{i,l}(V^TV)_{l,k} \geq U^t_{i,k}(V^TV)_{k,k}$$
$$\sum_{w=1}^M U^t_{w,k} \sum_{j=1}^N V^2_{j,k} \geq U^t_{i,k} \sum_{j=1}^N V^2_{j,k}$$

due to the non-negativity of U and V. ∎

Lemma 0.3. *Eq. 3.8 is equivalent to Eq. 0.1. with the auxiliary function* $G(U_{i,k}, U^t_{i,k})$.

Proof: It is straightforward to prove it by solving $\partial G/\partial U_{i,k} = 0$. ∎

Lemmas 2 and 3 ensure that $O_{i,k}$ is non-increasing under the update Eq. 3.8. Similarly, $\Theta_{j,k}$ is non-increasing under the update Eq. 3.10 due to the following lemma.

Lemma 0.4. *The function*

$$G(V_{j,k}, V^t_{j,k}) = \Theta_{j,k}(V^t_{j,k}) + \Theta'_{j,k}(V^t_{j,k})(V_{j,k} - V^t_{j,k}) + \left[\frac{(V^tU^TU)_{j,k} + \lambda_v V^t_{j,k}}{V^t_{j,k}}\right](V_{j,k} - V^t_{j,k})^2$$

is an auxiliary function of $\Theta_{j,k}$, *and Eq. 3.10 is equivalent to Eq. 0.1 with the auxiliary function* $G(V_{j,k}, V^t_{j,k})$.

Proof: Similar to the above. ∎

References

[1] D. Lee and S. Seung. Algorithms for Non-negative Matrix Factorization. In *NIPS*, pages 556–562, 2000.