Projected gradient method for Non-negative Matrix Factorization

Numerical Optimization

EECS Yung-Hsiang Wang, EECS Li-Yi Chang
Lecturer: Prof. Peng-Wen Chen

Outline

- Problem description
- Methodology
  - Multiplicative update algorithm
  - Projected gradient method
- Projected gradient methods for NMF
  - Directly apply
  - Divide and conquer
- Solution integrity
- Demo

Problem description

Key idea of NMF:

$$V = WH \qquad (1)$$

$$V \in \mathbb{R}^{n\times m}, W \in \mathbb{R}^{n\times r}, H \in \mathbb{R}^{r\times m}, r < min(n,m)$$

Constrain optimization problem:

$$\min_{W,H} f(W,H) = \left||V - WH|\right|_F^2 \qquad (2)$$

$$W_{ij} \geq 0, H_{ij} \geq 0, \forall i,j$$

Gradient of objective function:

$$\begin{aligned} \nabla f(W,H) &= [\nabla_W f(W,H), \nabla_H f(W,H)] \\ &= [(WH - V)H^T,\ W^T(WH - V)] \end{aligned} \qquad (3)$$

Outline

- Problem description
- Methodology
  - Multiplicative update algorithm
  - Projected gradient method
- Projected gradient methods for NMF
  - Directly apply
  - Divide and conquer
- Solution integrity
- Demo

Multiplicative update algorithm

Multiplicative update algorithm:

$$W_{ia}^{k+1} = W_{ia}^{k} \frac{\left(V\left(H^{k}\right)^{T}\right)_{ia}}{(W^{k}H^{k}(H^{k})^{T})_{ia}}, \forall i, a$$

$$H_{bj}^{k+1} = H_{bj}^{k} \frac{\left(\left(W^{k+1}\right)^{T}V\right)_{bj}}{\left((W^{k+1})^{T}W^{k+1}H^{k}\right)_{bj}}, \forall b, j$$

When $W_{ia}^{k+1} = W_{ia}^{k} > 0$, then $\left(V\left(H^{k}\right)^{T}\right)_{ia} = \left(W^{k}H^{k}\left(H^{k}\right)^{T}\right)_{ia}$, implies that $\nabla_{W} f\left(W^{k}, H^{k}\right)_{ia} = 0$

Projected gradients method

Projected gradients method:

$$x^{k+1} = P\left[x^k - \alpha^k \nabla f\left(x^k\right)\right]$$

Where

$$P[x_i] = \begin{cases} x_i & \text{if} \quad l_i < x_i < u_i, \\ u_i & \text{if} \quad x_i \geq u_i, \\ l_i & \text{if} \quad x_i \leq l_i \end{cases}$$

u and l means the boundaries.

Outline

- Problem description
- Methodology
  - Multiplicative update algorithm
  - Projected gradient method
- Projected gradient methods for NMF
  - Directly apply
  - Divide and conquer
- Solution integrity
- Demo

Directly applying projected gradients

Projected gradient method:

$$(W^{k+1}, H^{k+1}) = P\left[(W^k, H^k) - \alpha\left(\nabla_W f(W^k, H^k), \nabla_H f(W^k, H^k)\right)\right]$$

Stopping condition:

$$\left\|\nabla^P f(W^k, H^k)\right\|_F \leq \epsilon \left\|\nabla f(W^1, H^1)\right\|_F$$

The projected gradient operator is define as follows:

$$\nabla^P f(x)_i = \begin{cases} \nabla f(x)_i & \text{if } l_i < x_i < u_i, \\ \min(0, \nabla f(x)_i) & \text{if } x_i = l_i, \\ \max(0, \nabla f(x)_i) & \text{if } x_i = u_i. \end{cases}$$

Divide and conquer algorithm

Alternating non-negative least squares method:

$$W^{k+1} = \underset{W \geq 0}{argmin}\, f(W, H^k),$$
$$H^{k+1} = \underset{H \geq 0}{argmin}\, f\left(W^{k+1}, H\right)^k.$$

In each sub-problem we have following algorithm:

Given initial point $W_{ia} \geq 0, H_{jb} \geq 0$,
(S0) Fixed H^k, update $W^{k+1} = P[W^k - \alpha \nabla f(W^k)]$ until convergence.
(S1) Fixed W^{k+1}, update $H^{k+1} = P[H^k - \alpha \nabla f(H^k)]$ until convergence.
(S2) Repeat S0 and S1 until $\nabla f\left(W^{k+1}, H^{k+1}\right) \leq \epsilon \nabla f(W^1, H^1)$
(S3) Return W and H.

Outline

- Problem description
- Methodology
  - Multiplicative update algorithm
  - Projected gradient method
- Projected gradient methods for NMF
  - Directly apply
  - Divide and conquer
- Solution integrity
- Demo

Solution integrity

Karush-Kuhn-Tucker (KKT) optimality condition:

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ C_i(x^*) = 0, \ \forall i \in \epsilon \\ C_i(x^*) \geq 0, \ \forall i \in \mathcal{I} \\ \lambda_i^* \geq 0, \ \forall i \in \mathcal{I} \\ \lambda_i^* C_i(x^*) = 0, \ \forall i \in \epsilon \cup \mathcal{I} \end{cases} \qquad \Omega = \left\{ x \;\middle|\; \begin{array}{l} C_i(x) = 0, \ i \in \epsilon \\ C_i(x) \geq 0, \ i \in \mathcal{I} \end{array} \right\}$$

It is equal to following conditions,

$$\begin{cases} W_{ia} \geq 0, \quad H_{jb} \geq 0. \\ \nabla_w f(W, H)_{ia} \geq 0, \quad \nabla_H f(W, H)_{jb} \geq 0 \\ W_{ia} \cdot \nabla_w f(W, H)_{ia} = 0, \quad H_{jb} \cdot \nabla_H f(W, H)_{jb} = 0, \end{cases} \qquad \forall i, j, a, b$$

Solution integrity

1. According to Gonzales and Zhang (2005) points the flaw of multiplicative update method usually do not satisfy KKT condition.
2. The alternating non-negative least squares method is guarantee by corollary 2 of Grippo and Sciandrone (2000), we can claim that the result of this method is a stationary point.

We proof the KKT condition of directly apply projected gradient method with stopping condition will be elaborate in next slide.

Gonzalez, E. F., & Zhang, Y. (2005). Accelerating the Lee-Seung algorithm for non-negative matrix factorization. Dept. Comput. & Appl. Math., Rice Univ., Houston, TX, Tech. Rep. TR-05-02, 1-13.

Grippo, Luigi, and Marco Sciandrone. "On the convergence of the block nonlinear Gauss–Seidel method under convex constraints." *Operations research letters* 26.3 (2000): 127-136.

Solution integrity

The projected gradient operator is define as follows:

$$\nabla^P f(x)_i = \begin{cases} \nabla f(x)_i & \text{if } l_i < x_i < u_i, \\ \min(0, \nabla f(x)_i) & \text{if } x_i = l_i, \\ \max(0, \nabla f(x)_i) & \text{if } x_i = u_i. \end{cases}$$

This condition is an equivalent form of the KKT condition for bounded problems.

$$l_i \leq x_i \leq u_i, \forall i, \qquad \left|\left|\nabla^P f(x)\right|\right| = 0.$$

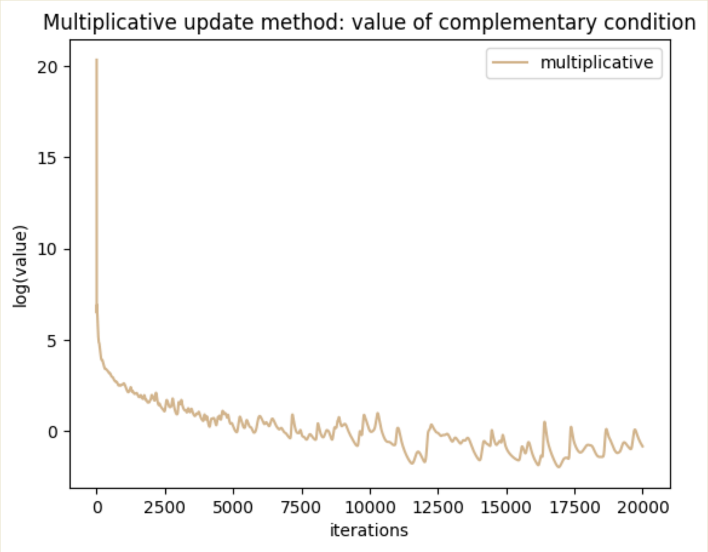
Here is the recall of KKT condition for this problem.

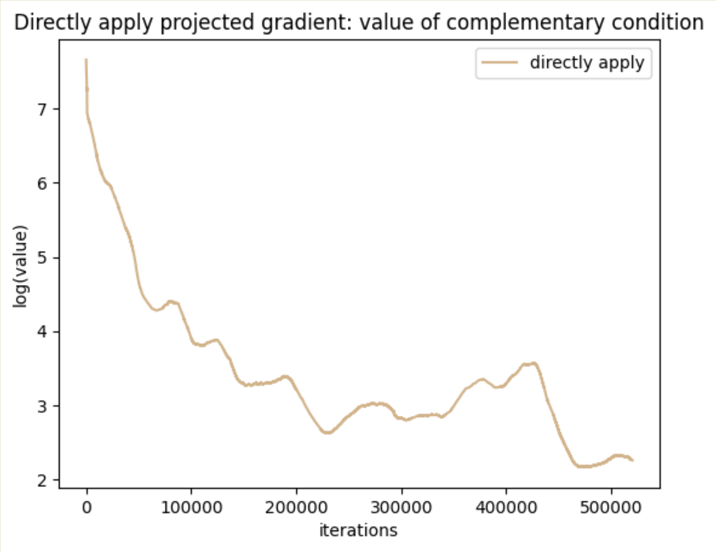
$$\begin{cases} W_{ia} \geq 0, \quad H_{jb} \geq 0. \\ \nabla_w f(W,H)_{ia} \geq 0, \quad \nabla_H f(W,H)_{jb} \geq 0 & \forall i,j,a,b \\ W_{ia} \cdot \nabla_w f(W,H)_{ia} = 0, \quad H_{jb} \cdot \nabla_H f(W,H)_{jb} = 0, \end{cases}$$

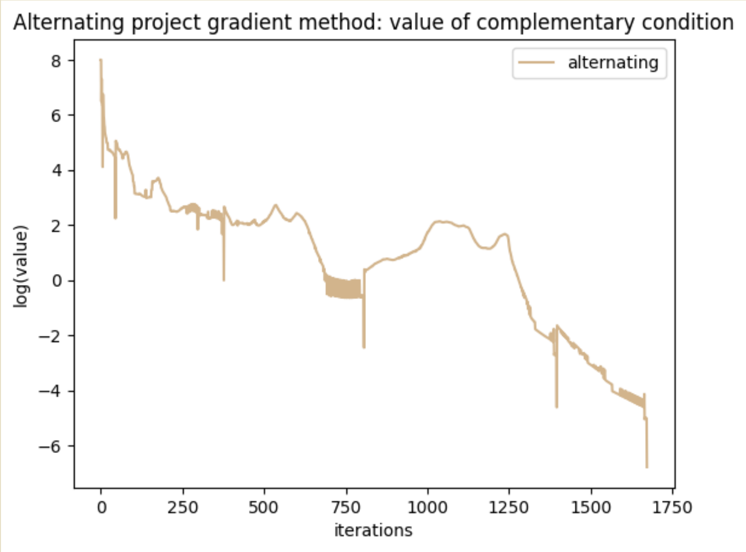
Outline

- Problem description
- Methodology
  - Multiplicative update algorithm
  - Projected gradient method
- Projected gradient methods for NMF
  - Directly apply
  - Divide and conquer
- Solution integrity
- Demo

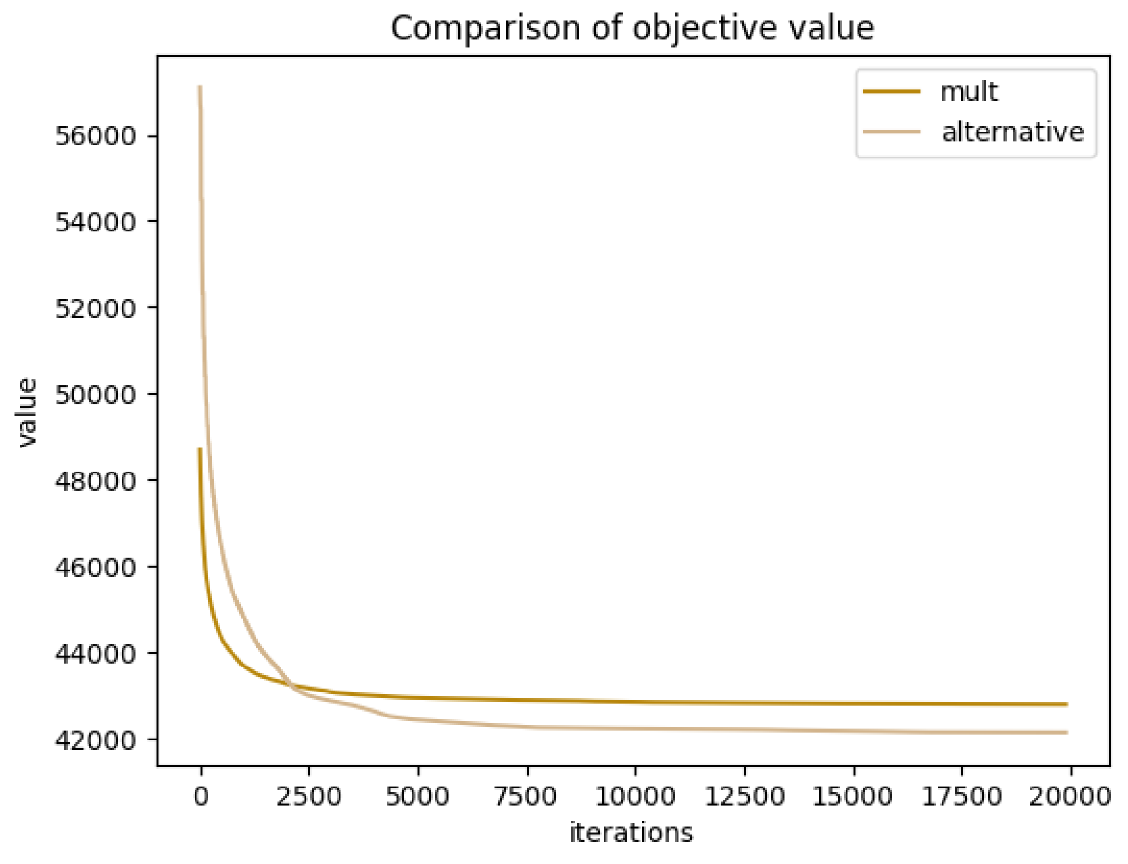
Complementary condition

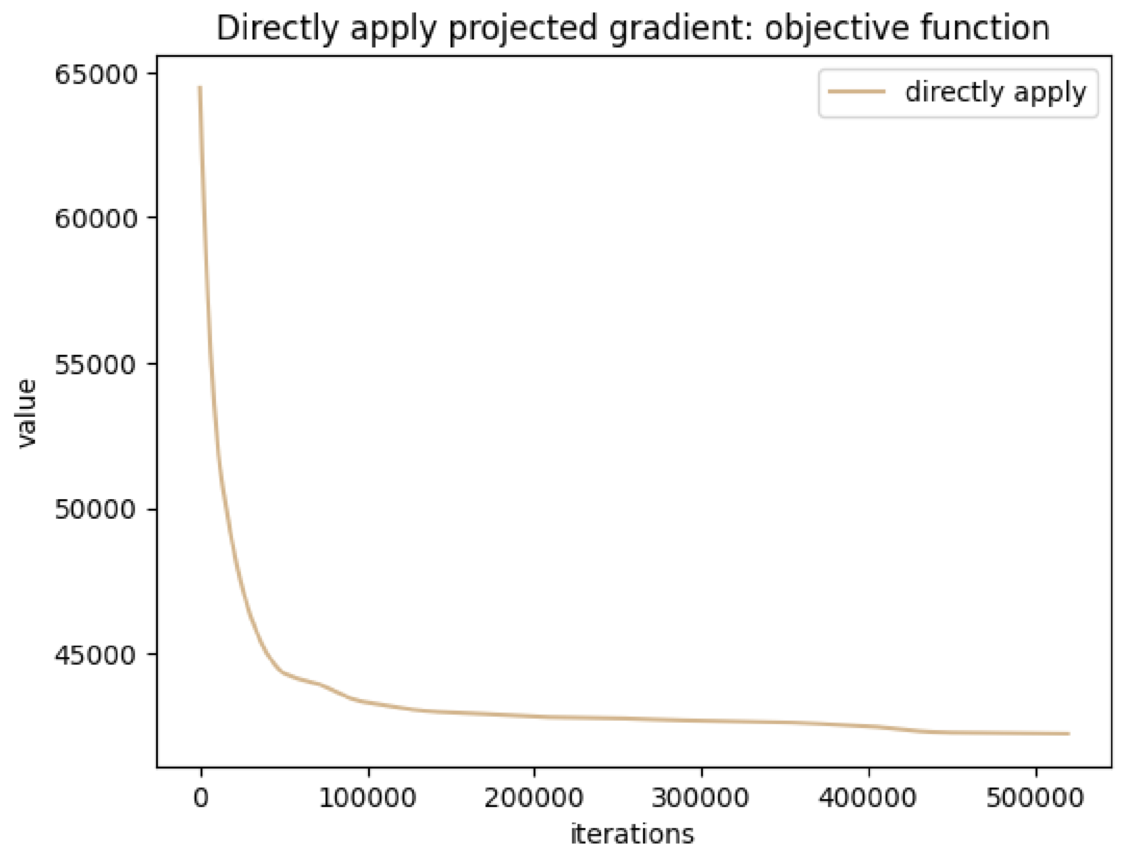






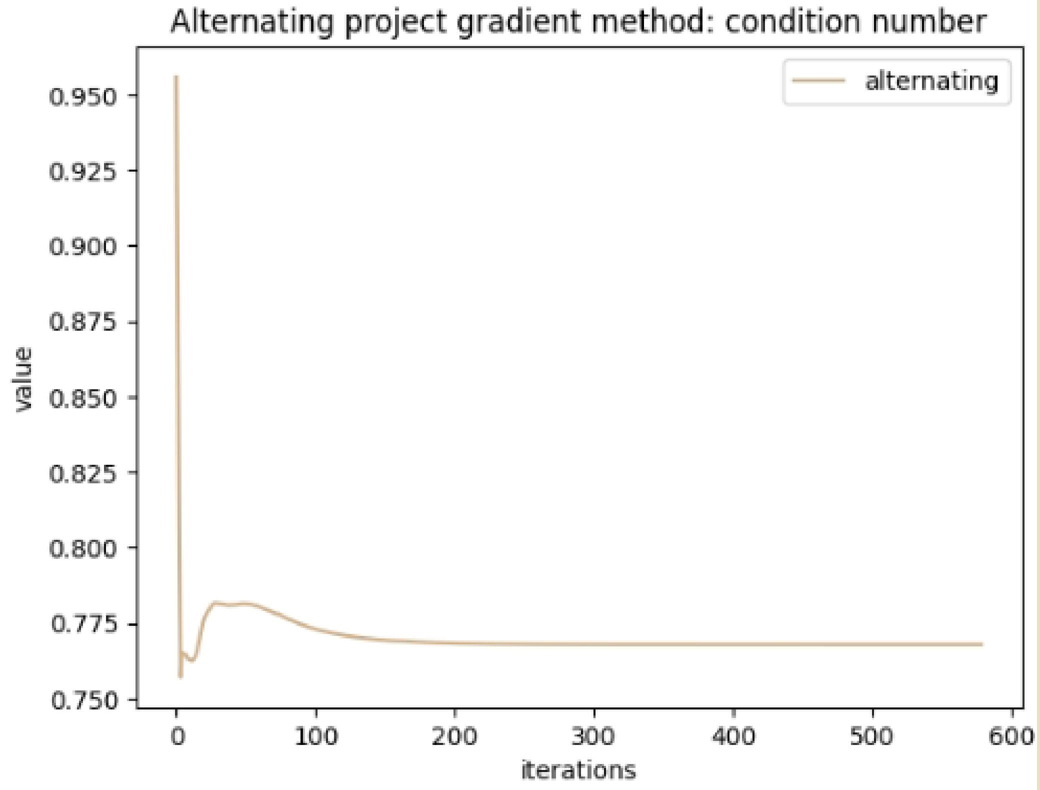
Objective value

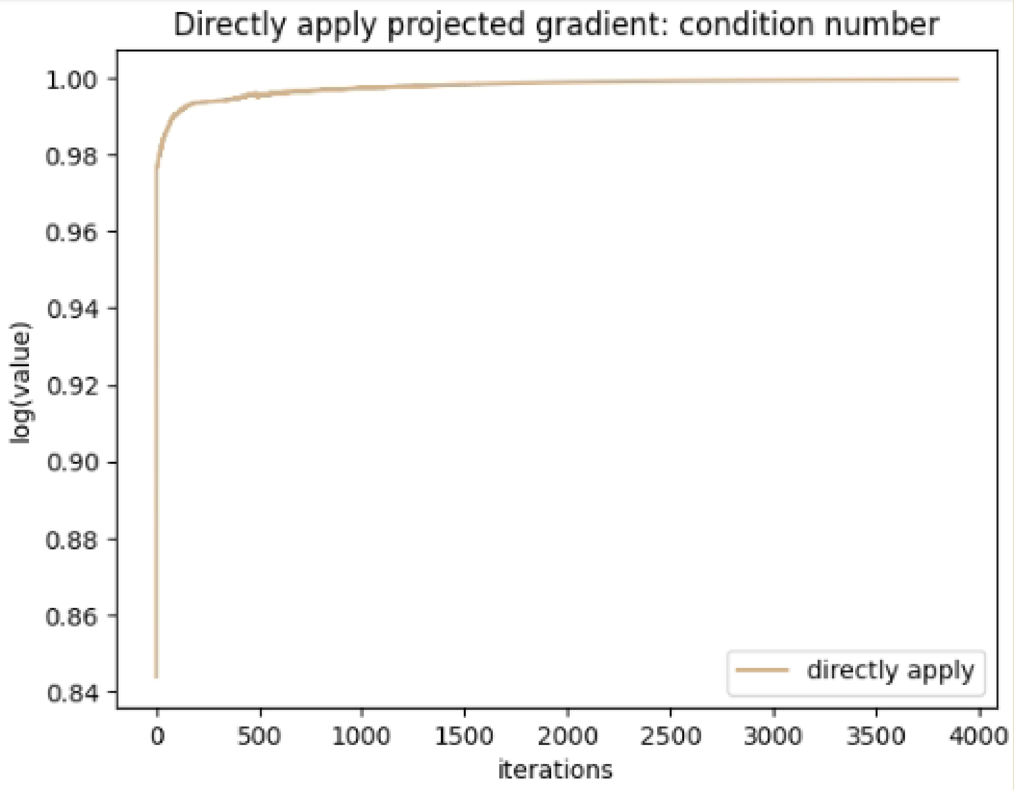




m = 100, n=70, r=30

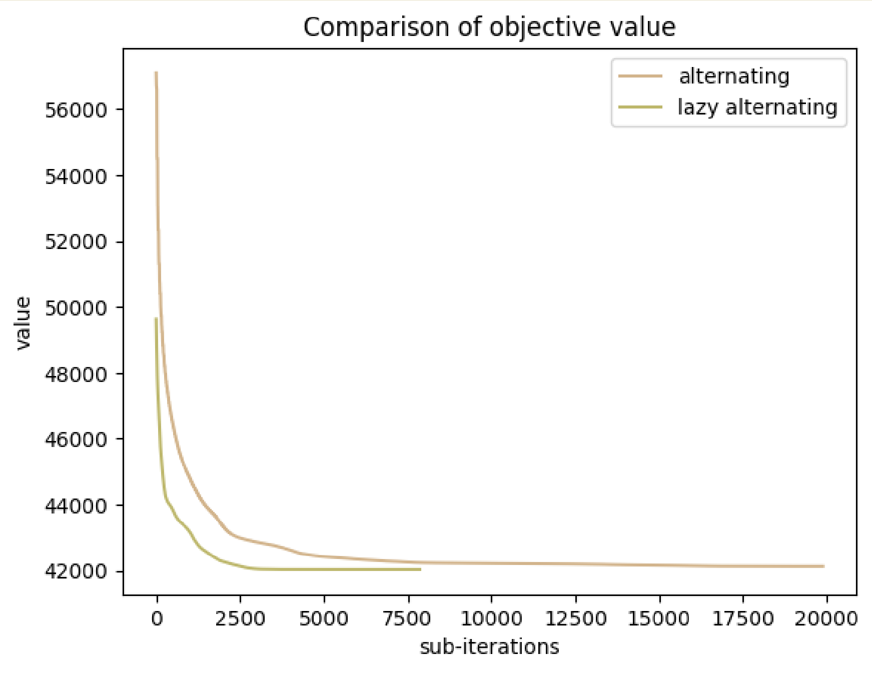
Condition number

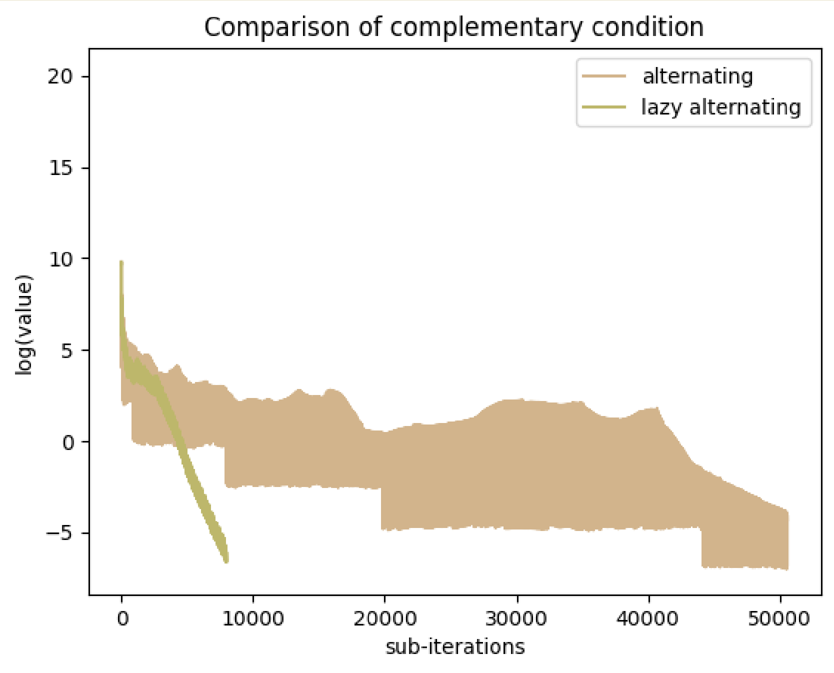




Further discussion

We apply the idea of ADMM on the alternating update method. Following is the numerical result.





Conclusion

From the above numerical results, we found that the alternating non-negative least squares method is better than multiplicative update algorithm and directly apply projected gradient method to NMF.

- Multiplicative update algorithm convergence faster than directly apply projected gradient method.
- The alternating non-negative least squares method is better because the cost is lower and performance is better than multiplicative algorithm.

Thank you for listening

Presented by Yung-Hsiang Wang, Li-Yi Chang
Jun 13th, 2024