Supplymentary Material for “Multi-task learning via robust regularized clustering with non-convex group penalties” by Akira Okazaki and Shuichi Kawano

Proof of Theorem 1

This proof is based on the similar steps of Wang et al. (2019) and Fan and Yin (2024). Although the modified ADMM does not explicitly require updates of B; the algorithm updates the value of U and S by the implicitly updated B in its construction (C4). Specifically, B is updated to the value that exactly minimizes the function at each update of U in the inner loop of the accelerated gradient method. Then, the modified ADMM is equivalent to the following updates of ADMM:

$$
\begin{aligned}
(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}) &= \arg\min_{\boldsymbol{w}_0, W} L_\nu(\boldsymbol{w}_0, W, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}), \\
(U^{(t+1)}, B^{(t+1)}) &= \arg\min_{U,B} L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U, O^{(t)}, B, S^{(t)}), \\
O^{(t+1)} &= \arg\min_{O} L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t)}, B^{(t+1)}, S^{(t)}), \\
S_{(m_1,m_2)}^{(t+1)} &= S^{(t)} + \nu(B^{(t+1)} - A_{\mathcal{E}} U^{(t+1)}).
\end{aligned}
$$

Thus, it suffices to prove the theorem regarding the ADMM based on the above updates.

For the simplicity of the notation, we denote the object function as

$$
Q(\boldsymbol{w}_0, W, U, O, B) = \sum_{m=1}^{T} \frac{1}{n_m} L(w_{m0}, \boldsymbol{w}_m) + \frac{\lambda_1}{2} \|W - U - O\|_F^2 + \lambda_2 \sum_{(m_1,m_2)\in\mathcal{E}} r_{m_1,m_2} \|\boldsymbol{b}_{(m_1,m_2)}\|_2 + \sum_{m=1}^{T} P(\boldsymbol{o}_m, \lambda_3, \gamma),
$$

the partial of the object function as

$$
H(\boldsymbol{w}_0, W, U, O) = \sum_{m=1}^{T} \frac{1}{n_m} L(w_{m0}, \boldsymbol{w}_m) + \frac{\lambda_1}{2} \|W - U - O\|_F^2 + \sum_{m=1}^{T} P(\boldsymbol{o}_m; \lambda_3, \gamma),
$$

and inner product as $\langle \cdot, \cdot \rangle$ that represents $\boldsymbol{a}^\top \boldsymbol{b}$ for vectors and $\mathrm{tr}(A^\top B)$ for matrices .

To prove the theorem, we show the following four steps:

1. $L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)})$ is lower bounded for all $t \in \mathbb{N}$.

2. $\left\{\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}\right\}$ is bounded.

3. $\left\{\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}\right\}$ converges to a limit point, that is, $\lim_{t\to\infty} \left\{\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}\right\} \to \left\{\boldsymbol{w}_0^{(*)}, W^{(*)}, U^{(*)}, O^{(*)}, B^{(*)}, S^{(*)}\right\}$.

4. Any limit point is a stationary point, that is, $\partial L_\nu(\boldsymbol{w}_0^{(*)}, W^{(*)}, U^{(*)}, O^{(*)}, B^{(*)}, S^{(*)}) \ni \mathbf{0}$.

Note that the objective function $Q(\boldsymbol{w}_0, W, U, O, B)$ is coercive over the feasible set, that is, $Q(\boldsymbol{w}_0, W, U, O, B) \to \infty$ if $A_{\mathcal{E}}U - B = 0$ and $\|\boldsymbol{w}_0, \mathrm{vec}(W), \mathrm{vec}(U), \mathrm{vec}(O), \mathrm{vec}(B)\|_2 \to \infty$. Moreover, as $\mathrm{Im}(A_{\mathcal{E}}) \subseteq \mathrm{Im}(I_{|\mathcal{E}|})$ with $\mathrm{Im}(\cdot)$ being the image of a matrix, there exists B' such that $A_{\mathcal{E}}U^{(t)} - B' = 0$. Therefore, we have

$$Q(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B') \geq \min_{\boldsymbol{w}_0, W, U, O, B} \{Q(\boldsymbol{w}_0, W, U, O, B) : A_{\mathcal{E}}U - B = 0\} > -\infty. \tag{1}$$

By the optimality condition for the updates of $(U^{(t+1)}, B^{(t+1)})$, it holds that for a subgradient $\boldsymbol{d}_{(m_1,m_2)}^{(t+1)} \in \partial(\lambda_2 r_{m_1,m_2} \|\boldsymbol{b}_{(m_1,m_2)}^{(t+1)}\|_2)$:

$$\begin{aligned}
&\boldsymbol{d}_{(m_1,m_2)}^{(t+1)} + \boldsymbol{s}_{(m_1,m_2)}^{(t)} + \nu(\boldsymbol{b}_{(m_1,m_2)}^{(t+1)} - (\boldsymbol{u}_{m_1}^{(t+1)} - \boldsymbol{u}_{m_2}^{(t+1)})) = \mathbf{0}, \quad \text{for } (m_1, m_2) \in \mathcal{E}, \\
&-\lambda_1(W^{(t+1)} - U^{(t+1)} - O^{(t)}) + A_{\mathcal{E}}^\top (S^{(t)} + \nu(B^{(t+1)} - A_{\mathcal{E}}U^{(t+1)})) = 0.
\end{aligned}$$

In addition, by the update of $S^{(t+1)}$, we have

$$\boldsymbol{s}_{(m_1,m_2)}^{(t+1)} = -\boldsymbol{d}_{(m_1,m_2)}^{(t+1)}, \quad \text{for } (m_1, m_2) \in \mathcal{E}.$$

For the subgradients $\boldsymbol{d}_{(m_1,m_2)}^{(t)} \in \partial(\lambda_2 r_{m_1,m_2} \|\boldsymbol{b}_{(m_1,m_2)}^{(t)}\|_2)$ and $\boldsymbol{d}'_{(m_1,m_2)} \in \partial(\lambda_2 r_{m_1,m_2} \|\boldsymbol{b}'_{(m_1,m_2)}\|_2)$ from the convexity, we have

$$\begin{aligned}
\lambda_2 r_{m_1,m_2} \|\boldsymbol{b}_{(m_1,m_2)}^{(t)}\|_2 - \lambda_2 r_{m_1,m_2} \|\boldsymbol{b}'_{(m_1,m_2)}\|_2 &\geq \langle \boldsymbol{d}_{(m_1,m_2)}^{(\prime)}, \boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)} \rangle, \\
\lambda_2 r_{m_1,m_2} \|\boldsymbol{b}_{(m_1,m_2)}^{(t)}\|_2 + \langle \boldsymbol{d}_{(m_1,m_2)}^{(t)}, \boldsymbol{b}'_{(m_1,m_2)} - \boldsymbol{b}_{(m_1,m_2)}^{(t)} \rangle &\geq \lambda_2 r_{m_1,m_2} \|\boldsymbol{b}'_{(m_1,m_2)}\|_2 \\
&\quad + \langle \boldsymbol{d}'_{(m_1,m_2)} - \boldsymbol{d}_{(m_1,m_2)}^{(t)}, \boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)} \rangle.
\end{aligned} \tag{2}$$

Then, we have

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) \\
&= Q(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}) \\
&+ \sum_{(m_1,m_2)\in\mathcal{E}} \Big\{ \boldsymbol{s}_{(m_1,m_2)}^{(t)\top}(\boldsymbol{b}_{(m_1,m_2)}^{(t)} - (\boldsymbol{u}_{m_1}^{(t)} - \boldsymbol{u}_{m_2}^{(t)})) + \frac{\nu}{2}\|\boldsymbol{b}_{(m_1,m_2)}^{(t)} - (\boldsymbol{u}_{m_1}^{(t)} - \boldsymbol{u}_{m_2}^{(t)})\|_2^2 \Big\} \\
&= H(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}) + \sum_{(m_1,m_2)\in\mathcal{E}} r_{m_1,m_2}\|\boldsymbol{b}_{(m_1,m_2)}^{(t)}\|_2 + \\
&+ \sum_{(m_1,m_2)\in\mathcal{E}} \Big\{ \boldsymbol{d}_{(m_1,m_2)}^{(t)\top}(\boldsymbol{b}'_{(m_1,m_2)} - \boldsymbol{b}_{(m_1,m_2)}^{(t)}) + \frac{\nu}{2}\|\boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)}\|_2^2 \Big\} \\
&\geq H(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}) + \sum_{(m_1,m_2)\in\mathcal{E}} \Big\{ r_{m_1,m_2}\|\boldsymbol{b}'_{(m_1,m_2)}\|_2 + \langle \boldsymbol{d}'_{(m_1,m_2)} - \boldsymbol{d}_{(m_1,m_2)}^{(t)}, \boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)} \rangle \\
&+ \frac{\nu}{2}\|\boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)}\|_2^2 \Big\} \\
&\geq Q(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B') \\
&+ \sum_{(m_1,m_2)\in\mathcal{E}} \Big\{ \frac{\nu}{2}\|\boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)}\|_2^2 - 2\lambda_2 r_{m_1,m_2}\|\boldsymbol{b}_{(m_1,m_2)}^{(t)} - \boldsymbol{b}'_{(m_1,m_2)}\|_2 \Big\} \\
&> -\infty
\end{aligned} \tag{3}
$$

The second inequality is based on the fact $\|\partial_{\boldsymbol{b}}\|\boldsymbol{b}\|_2\|_2 \leq 1$ for $\forall \boldsymbol{b} \in \mathbb{R}^p$ and Cauchy–Schwarz inequality. The last inequality is derived from the (1). This completes the proof of the first step.

To bound $L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t+1)})$, we show the following bounds regarding each update:

(i) $L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)})$,

(ii) $L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t)}, B^{(t+1)}, S^{(t)})$,

(iii) $L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t)}, B^{(t+1)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t)})$,

(iv) $L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t+1)})$.

For (i), we have

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) \\
&= \sum_{m=1}^{T} \frac{1}{n_m} \left\{ L(w_{m0}^{(t)}, \boldsymbol{w}_m^{(t)}) - L(w_{m0}^{(t+1)}, \boldsymbol{w}_m^{(t+1)}) \right\} + \frac{\lambda_1}{2} \left\{ \|W^{(t)} - U^{(t)} - O^{(t)}\|_F^2 - \|W^{(t+1)} - U^{(t)} - O^{(t)}\|_F^2 \right\} \\
&\geq \frac{\lambda_1}{2} \|W^{(t)} - W^{(t+1)}\|_F^2.
\end{aligned} \tag{4}
$$

The inequality is derived from the convexity of the loss function. For (ii), we have

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t)}, B^{(t+1)}, S^{(t)}) \\
&= \frac{\lambda_1}{2} \left\{ \|W^{(t+1)} - U^{(t)} - O^{(t)}\|_F^2 - \|W^{(t+1)} - U^{(t+1)} - O^{(t)}\|_F^2 \right\} + \lambda_2 \sum_{(m_1,m_2)\in\mathcal{E}} \left\{ \|\boldsymbol{b}_{(m_1,m_2)}^{(t)}\|_2 - \|\boldsymbol{b}_{(m_1,m_2)}^{(t+1)}\|_2 \right\} \\
&+ \langle S^{(t)}, B^{(t)} - A_{\mathcal{E}}U^{(t)} - (B^{(t+1)} - A_{\mathcal{E}}U^{(t+1)}) \rangle \\
&+ \frac{\nu}{2} \left\{ \|B^{(t)} - A_{\mathcal{E}}U^{(t)}\|_F^2 - \|B^{(t+1)} - A_{\mathcal{E}}U^{(t+1)}\|_F^2 \right\} \\
&= \frac{\lambda_1}{2} \|U^{(t)} - U^{(t+1)}\|_F^2 + \lambda_2 \sum_{(m_1,m_2)\in\mathcal{E}} \left\{ \|\boldsymbol{b}_{(m_1,m_2)}^{(t)}\|_2 - \|\boldsymbol{b}_{(m_1,m_2)}^{(t+1)}\|_2 \right\} \\
&+ \nu \langle B^{(t+1)} - A_{\mathcal{E}}U^{(t+1)}, A_{\mathcal{E}}(U^{(t)} - U^{(t+1)}) \rangle \\
&+ \langle D^{(t+1)}, B^{(t+1)} - B^{(t)} \rangle + \nu \langle A_{\mathcal{E}}U^{(t)} - B^{(t+1)}, B^{(t)} - B^{(t+1)} \rangle \\
&+ \frac{\nu}{2} \left\{ \|B^{(t)} - A_{\mathcal{E}}U^{(t)}\|_F^2 - \|B^{(t+1)} - A_{\mathcal{E}}U^{(t)}\|_F^2 \right\} \\
&\geq \frac{\lambda_1}{2} \|U^{(t)} - U^{(t+1)}\|_F^2 + \frac{\nu}{2} \|(B^{(t)} - A_{\mathcal{E}}U^{(t)}) - (B^{(t+1)} - A_{\mathcal{E}}U^{(t+1)})\|_F^2
\end{aligned} \tag{5}
$$

where $D^{(t+1)}$ is an $|\mathcal{E}| \times p$ matrix whose each row is $\boldsymbol{d}_{(m_1,m_2)}^{(t+1)}$. The second equality is derived from the optimality condition of $(U^{(t+1)}, B^{(t+1)})$, and the cosine rule $\|A - B\|_F^2 - \|A - C\|_F^2 + 2\langle A - C, B - C\rangle = \|B - C\|_F^2$. The inequality is derived from the convexity of $\|\boldsymbol{b}_{(m_1,m_2)}^{(t+1)}\|_2$. For (iii), we have

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t)}, B^{(t+1)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t)}) \\
&= \frac{\lambda_1}{2} \left\{ \|W^{(t+1)} - U^{(t+1)} - O^{(t)}\|_F^2 - \|W^{(t+1)} - U^{(t+1)} - O^{(t+1)}\|_F^2 \right\} + \sum_{m=1}^{T} \left\{ P(\boldsymbol{o}_m^{(t)}, \lambda_3, \gamma) - P(\boldsymbol{o}_m^{(t+1)}, \lambda_3, \gamma) \right\} \\
&\geq \frac{\lambda_1 - C_p}{2} \|O^{(t+1)} - O^{(t)}\|_F^2.
\end{aligned} \tag{6}
$$

The inequality is derived from the weakly convexity of the regularization term with the constant $C_p > 0$. For (iv), we have

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t+1)}) \\
&= \langle S^{(t)} - S^{(t+1)}, B^{(t+1)} - A_{\mathcal{E}} U^{(t+1)} \rangle \\
&= -\frac{1}{\nu} \| S^{(t)} - S^{(t+1)} \|_F^2 \\
&\geq -\frac{1}{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \| A_{\mathcal{E}}^\top (S^{(t)} - S^{(t+1)}) \|_F^2 \\
&= -\frac{\lambda_1^2}{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \| W^{(t)} - U^{(t)} - O^{(t-1)} - (W^{(t+1)} - U^{(t+1)} - O^{(t)}) \|_F^2 \\
&\geq -\frac{\lambda_1^2}{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \left\{ \| W^{(t)} - W^{(t+1)} \|_F^2 + \| U^{(t)} - U^{(t+1)} \|_F^2 + \| O^{(t-1)} - O^{(t)} \|_F^2 \right\},
\end{aligned} \tag{7}
$$

where $\lambda_{++}(\cdot)$ denotes a strictly positive minimum eigenvalue of a matrix. By combining the above upper bound of each update, we obtain

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) - L_\nu(\boldsymbol{w}_0^{(t+1)}, W^{(t+1)}, U^{(t+1)}, O^{(t+1)}, B^{(t+1)}, S^{(t+1)}) \\
&\geq \lambda_1 \left(\frac{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}}) - 2\lambda_1}{2\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \right) \left(\| W^{(t)} - W^{(t+1)} \|_F^2 + \| U^{(t)} - U^{(t+1)} \|_F^2 \right) \\
&+ \frac{\nu}{2} \left\| (B^{(t)} - A_{\mathcal{E}} U^{(t)}) - (B^{(t+1)} - A_{\mathcal{E}} U^{(t+1)}) \right\|_F^2 \\
&+ \frac{\lambda_1 - C_p}{2} \| O^{(t)} - O^{(t+1)} \|_F^2 - \frac{\lambda_1^2}{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \| O^{(t-1)} - O^{(t)} \|_F^2.
\end{aligned} \tag{8}
$$

Then, we have

$$
\begin{aligned}
&L_\nu(\boldsymbol{w}_0^{(0)}, W^{(0)}, U^{(0)}, O^{(0)}, B^{(0)}, S^{(0)}) - L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) \\
&\geq \sum_{l=0}^{t} \left\{ \lambda_1 \left(\frac{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}}) - 2\lambda_1}{2\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \right) \left(\| W^{(l)} - W^{(l+1)} \|_F^2 + \| U^{(l)} - U^{(l+1)} \|_F^2 \right) \right. \\
&\left. + \frac{\nu}{2} \left\| (B^{(t)} - A_{\mathcal{E}} U^{(t)}) - (B^{(t+1)} - A_{\mathcal{E}} U^{(t+1)}) \right\|_F^2 \right\} \\
&+ \left(\frac{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})(\lambda_1 - C_p) - 2\lambda_1^2}{2\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \right) \sum_{l=1}^{t} \| O^{(l)} - O^{(l-1)} \|_F^2 + \frac{\lambda_1 - C_p}{2} \| O^{(t)} - O^{(t+1)} \|_F^2 - \frac{\lambda_1^2}{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \| O^{(0)} \|_F^2.
\end{aligned} \tag{9}
$$

From this, $L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)})$ is upper bounded by $L_\nu(\boldsymbol{w}_0^{(0)}, W^{(0)}, U^{(0)}, O^{(0)}, B^{(0)}, S^{(0)}) + \frac{\lambda_1^2}{\nu \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})} \| O^{(0)} \|_F^2$. Thus, $Q(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B')$ and $\| B^{(t)} - A_{\mathcal{E}} U^{(t)} \|_F^2$ are also upper bounded. Since the objective function $Q(\boldsymbol{w}_0, W, U, O, B)$ is coercive over the feasible set, $\{\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}\}$ is bounded. Then, from the following inequality:

$$
\| S^{(t)} \|_F^2 \leq \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})^{-1} \| A_{\mathcal{E}}^\top S^{(t)} \|_F^2 = \lambda_{++}(A_{\mathcal{E}}^\top A_{\mathcal{E}})^{-1} \lambda_1^2 \| W^{(t+1)} - U^{(t+1)} - O^{(t)} \|_F^2, \tag{10}
$$

$S^{(t)}$ is also bounded. Furthermore, from the lower boundness of the $L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)})$, the value of right hand side of (9) converges to some non-negative constant without the last term as $t \to \infty$. Therefore, we obtain $\lim_{t\to\infty} \|W^{(t)} - W^{(t+1)}\|_F^2 = 0$, $\lim_{t\to\infty} \|U^{(t)} - U^{(t+1)}\|_F^2 = 0$, $\lim_{t\to\infty} \|(B^{(t)} - A_\mathcal{E} U^{(t)}) - (B^{(t+1)} - A_\mathcal{E} U^{(t+1)})\|_F^2{=}0$, and $\lim_{t\to\infty} \|O^{(t)} - O^{(t+1)}\|_F^2 = 0$. Moreover, from triangle inequality

$$\|B^{(t)} - B^{(t+1)}\|_F^2 \leq \|(B^{(t)} - A_\mathcal{E} U^{(t)}) - (B^{(t+1)} - A_\mathcal{E} U^{(t+1)})\|_F^2 + \|A_\mathcal{E}(U^{(t)} - U^{(t+1)})\|_F^2$$

and the fact $\|A_\mathcal{E}(U^{(t)} - U^{(t+1)})\| \leq \lambda_+(A_\mathcal{E}^\top A_\mathcal{E})\|U^{(t)} - U^{(t+1)}\|_F^2$, we obtain $\lim_{t\to\infty} \|B^{(t)} - B^{(t+1)}\|_F^2 = 0$, where $\lambda_+(\cdot)$ denotes a positive maximum eigenvalue of a matrix. These convergence and (10) also imply $\lim_{t\to\infty} \|S^{(t)} - S^{(t+1)}\|_F^2 = 0$. Since the estimation of $\boldsymbol{w}_0$ calculated by Algorithm 4 only depends on $U^{(t-1)}$ and $O^{(t-1)}$, $\lim_{t\to\infty} \|\boldsymbol{w}_0^{(t)} - \boldsymbol{w}_0^{(t+1)}\|_2^2 = 0$. This concludes the steps 2 and 3.

To prove the limit point of the ADMM algorithm satisfies the optimality condition, we show

$$\lim_{t\to\infty} \left\| \partial L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}) \right\|_F = 0.$$

We denote $L_\nu(\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}) = L_\nu^{(t)}$. From the optimality condition of update for $\boldsymbol{o}_m^{(t)}$, we have $\partial_{\boldsymbol{o}_m} L_\nu^{(t)} \ni \mathbf{0}$. We have

$$\begin{aligned}
\lim_{t\to\infty} \left\| \frac{\partial L_\nu^{(t)}}{\partial(\boldsymbol{w}_0, W)} \right\|_F &\leq \lim_{t\to\infty} \sum_{m=1}^{T} \left\| \frac{\partial L_\nu^{(t)}}{\partial \boldsymbol{w}'_m} \right\|_2 \\
&= \lim_{t\to\infty} \sum_{m=1}^{T} \left\| \lambda_1(\boldsymbol{w}_m^{(t-1)} - \boldsymbol{u}_m^{(t-1)} - \boldsymbol{o}_m^{(t-1)}) - \lambda_1(\boldsymbol{w}_m^{(t)} - \boldsymbol{u}_m^{(t)} - \boldsymbol{o}_m^{(t)}) \right\|_2 = 0
\end{aligned}$$

and

$$\begin{aligned}
\lim_{t\to\infty} \|\partial_B L_\nu^{(t)}\|_F &\leq \lim_{t\to\infty} \sum_{(m_1,m_2)\in\mathcal{E}} \left\| \partial_{\boldsymbol{b}_{(m_1,m_2)}} L_\nu^{(t)} \right\|_2 \\
&= \lim_{t\to\infty} \sum_{(m_1,m_2)\in\mathcal{E}} \left\| \boldsymbol{d}_{(m_1,m_2)}^{(t)} + \boldsymbol{s}_{(m_1,m_2)}^{(t)} + \nu(\boldsymbol{b}_{(m_1,m_2)}^{(t)} - (\boldsymbol{u}_{m_1}^{(t)} - \boldsymbol{u}_{m_2}^{(t)})) \right\|_2 \\
&= \lim_{t\to\infty} \sum_{(m_1,m_2)\in\mathcal{E}} \nu \left\| \boldsymbol{b}_{(m_1,m_2)}^{(t)} - (\boldsymbol{u}_{m_1}^{(t)} - \boldsymbol{u}_{m_2}^{(t)}) \right\|_2 \\
&\leq \lim_{t\to\infty} \nu \left\| B^{(t)} - A_\mathcal{E} U^{(t)} \right\|_F^2 = 0.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\lim_{t\to\infty}\left\|\frac{\partial L_\nu^{(t)}}{\partial U}\right\|_F &= \lim_{t\to\infty}\left\|-\lambda_1(W^{(t)}-U^{(t)}-O^{(t)})-\langle A_{\mathcal{E}}, S^{(t)}+\nu(B^{(t)}-A_{\mathcal{E}}U^{(t)})\rangle\right\|_F \\
&= \lim_{t\to\infty}\left\|\lambda_1(W^{(t)}-W^{(t+1)}-(U^{(t)}-U^{(t+1)}))+\nu\langle A_{\mathcal{E}}, (B^{(t)}-B^{(t+1)})-A_{\mathcal{E}}(U^{(t)}-U^{(t+1)})\rangle\right\|_F \\
&\leq \lim_{t\to\infty}\lambda_1\left\{\|W^{(t)}-W^{(t+1)}\|_F+\|U^{(t)}-U^{(t+1)}\|_F\right\} \\
&+\nu\left\{\|A_{\mathcal{E}}^\top(B^{(t)}-B^{(t+1)})\|_F+\|A_{\mathcal{E}}^\top A_{\mathcal{E}}(U^{(t)}-U^{(t+1)})\|_F\right\}=0.
\end{aligned}$$

The last equality is derived from the fact $\left\|A_{\mathcal{E}}^\top(B^{(t)}-B^{(t+1)})\right\|_F^2 \leq \lambda_+(A_{\mathcal{E}}^\top A_{\mathcal{E}})\left\|B^{(t)}-B^{(t+1)}\right\|_F^2$ and $\left\|A_{\mathcal{E}}^\top A_{\mathcal{E}}(U^{(t)}-U^{(t+1)})\right\|_F^2 \leq \lambda_+(A_{\mathcal{E}}^\top A_{\mathcal{E}})\left\|A_{\mathcal{E}}(U^{(t)}-U^{(t+1)})\right\|_F^2$. Finally, we have

$$\begin{aligned}
\lim_{t\to\infty}\left\|\frac{\partial L_\nu^{(t)}}{\partial S}\right\|_F &= \lim_{t\to\infty}\left\|B^{(t)}-A_{\mathcal{E}}U^{(t)}\right\|_F \\
&= \lim_{t\to\infty}\frac{1}{\nu}\left\|S^{(t+1)}-S^{(t)}\right\|_F=0.
\end{aligned}$$

Therefore, we have $\lim_{t\to\infty}\left\|\partial L_\nu^{(t)}\right\|_F=0$.

Because $\{\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}\}$ is bounded and converges to a limit point denoted by $\{\boldsymbol{w}_0^{(*)}, W^{(*)}, U^{(*)}, O^{(*)}, B^{(*)}, S^{(*)}\}$. Moreover, it holds that $\lim_{t\to\infty}\|\partial L_\nu^{(t)}\|_F=\|\partial L_\nu^{(*)}\|_F=0$. From the definition of general subgradient, we have $\partial L_\nu^{(*)}\ni\mathbf{0}$. Therefore, the sequence $\{\boldsymbol{w}_0^{(t)}, W^{(t)}, U^{(t)}, O^{(t)}, B^{(t)}, S^{(t)}\}$ has at least a limit point $\{\boldsymbol{w}_0^{(*)}, W^{(*)}, U^{(*)}, O^{(*)}, B^{(*)}, S^{(*)}\}$ and any limit point is a stationary point. This completes the proof.

References

Fan, Y. and Yin, G. (2024). Gaussian mixture models with concave penalized fusion. *Statistica Sinica*, **3**, 2115–2139.

Wang, Y., Yin, W., and Zeng, J. (2019). Global convergence of admm in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, **78**, 29–63.