

Supplementary Material

In this section, we present definitions that were not explicitly stated in the main manuscript, alongside additional validation experiments that showcase the robustness of our algorithms, and theoretical proofs of Lemma 2 and Theorems 3 and 4.

A Definitions

Definition 5 (Laplacian operator) The Laplacian operator [8] $\mathcal{L} : \mathbb{R}_+^{p(p-1)/2} \rightarrow \mathbb{R}^{p \times p}$, which takes a nonnegative vector \mathbf{w} and outputs a Laplacian matrix $\mathcal{L}\mathbf{w}$, is defined as

$$[\mathcal{L}\mathbf{w}]_{ij} = \begin{cases} -w_{i+s(j)}, & \text{if } i > j, \\ [\mathcal{L}\mathbf{w}]_{ji}, & \text{if } i < j, \\ -\sum_{i \neq j} [\mathcal{L}\mathbf{w}]_{ij}, & \text{if } i = j, \end{cases} \quad (23)$$

where $s(j) = \frac{j-1}{2}(2p-j) - j$.

Definition 6 (Degree operator) The degree operator $\mathfrak{d} : \mathbb{R}^{p(p-1)/2} \rightarrow \mathbb{R}^p$, which takes a nonnegative vector \mathbf{w} and outputs the diagonal of a Degree matrix, is defined as

$$\mathfrak{d}\mathbf{w} = \text{diag}(\mathcal{L}\mathbf{w}). \quad (24)$$

Definition 7 (Adjoint of Laplacian operator) The adjoint of Laplacian operator [8] $\mathcal{L}^* : \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p(p-1)/2}$ is defined as

$$(\mathcal{L}^* \mathbf{P})_{s(i,j)} = \mathbf{P}_{i,i} - \mathbf{P}_{i,j} - \mathbf{P}_{j,i} + \mathbf{P}_{j,j} \quad (25)$$

where $s(i,j) = i - j + \frac{j-1}{2}(2p-j)$, $i > j$.

Definition 8 (Adjoint of degree operator) The adjoint of degree operator $\mathfrak{d}^* : \mathbb{R}^p \rightarrow \mathbb{R}^{p(p-1)/2}$ is given as

$$(\mathfrak{d}^* \mathbf{y})_{s(i,j)} = \mathbf{y}_i + \mathbf{y}_j, \quad (26)$$

where $s(i,j) = i - j + \frac{j-1}{2}(2p-j)$, $i > j$.

Definition 9 (Modularity) The modularity of a graph \mathcal{G} [49] is defined as $Q : \mathcal{G} \rightarrow \mathbb{R}$:

$$Q(\mathcal{G}) \triangleq \frac{1}{2|\mathcal{E}|} \sum_{i,j \in \mathcal{V}} \left(\mathbf{w}_{ij} - \frac{d_i d_j}{2|\mathcal{E}|} \right) \mathbb{1}(t_i = t_j), \quad (27)$$

where d_i is the weighted degree of the i -th node, i.e. $d_i \triangleq [\mathfrak{d}(\mathbf{w})]_i$, t_i is the type (or label) of the i -th node, and $\mathbb{1}(\cdot)$ is the indicator function.

B Additional Experiments

B.1 Empirical Convergence

In this section, we illustrate the empirical convergence performance of the proposed algorithms. All the experiments were carried out in a MacBook Pro 13in. 2019 with Intel Core i7 2.8GHz, 16GB of RAM. In Figure 9, we observe that the Lagrangian function quickly approaches a stationary value after a transient phase typical of ADMM-like algorithms.

C Proofs

C.1 Proof of Lemma 2

Proof We define an index set Ω_t :

$$\Omega_t \triangleq \left\{ l : [\mathcal{L}\mathbf{x}]_{tt} = \sum_{l \in \Omega_t} x_l \right\}, \quad t \in [1, 2, \dots, p]. \quad (28)$$

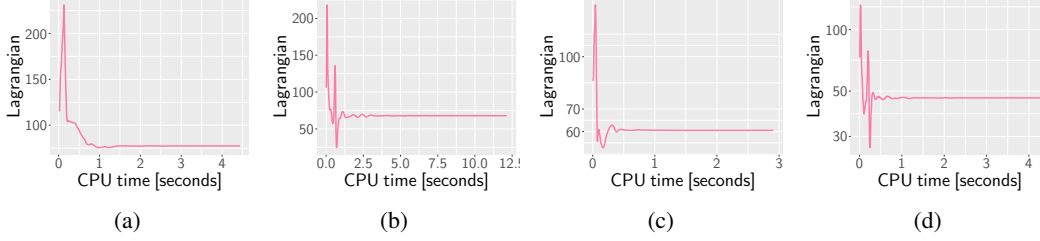


Figure 9: Empirical convergence for Algorithm 1 (panels (a) and (c)), *i.e.* connected graphs, and Algorithm 2 (panels (b) and (d)), *i.e.* k -component graphs, with cryptocurrencies data (panels (a) and (b)) and foreign exchange data (panels (c) and (d)).

Then we have

$$\begin{aligned}
\lambda_{\max}(\mathcal{L}^* \mathcal{L}) &= \sup_{\|\mathbf{x}\|=1} \mathbf{x}^\top \mathcal{L}^* \mathcal{L} \mathbf{x} = \sup_{\|\mathbf{x}\|=1} \|\mathcal{L} \mathbf{x}\|_F^2 = \sup_{\|\mathbf{x}\|=1} 2 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{i=1}^p [(\mathcal{L} \mathbf{x})_{ii}]^2 \\
&= \sup_{\|\mathbf{x}\|=1} 4 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{t=1}^p \sum_{i,j \in \Omega_t, i \neq j} x_i x_j \leq 4 + \sup_{\|\mathbf{x}\|=1} \frac{1}{2} \sum_{t=1}^p \sum_{i,j \in \Omega_t, i \neq j} x_i^2 + x_j^2 \\
&= (4 + 2(|\Omega_t| - 1)) \sup_{\|\mathbf{x}\|=1} \sum_{k=1}^{p(p-1)/2} x_k^2 = 2p,
\end{aligned}$$

with equality if and only if $x_1 = \dots = x_{p(p-1)/2} = \sqrt{\frac{2}{p(p-1)/2}}$ or $x_1 = \dots = x_{p(p-1)/2} = -\sqrt{\frac{2}{p(p-1)/2}}$. The last equality follows the fact that $|\Omega_t| = p - 1$.

Similarly, we can obtain

$$\begin{aligned}
\lambda_{\max}(\mathfrak{d}^* \mathfrak{d}) &= \sup_{\|\mathbf{x}\|=1} \mathbf{x}^\top \mathfrak{d}^* \mathfrak{d} \mathbf{x} = \sup_{\|\mathbf{x}\|=1} \|\mathfrak{d} \mathbf{x}\|^2 = \sup_{\|\mathbf{x}\|=1} 2 \sum_{k=1}^{p(p-1)/2} x_k^2 + \sum_{t=1}^p \sum_{i,j \in \Omega_t, i \neq j} x_i x_j \\
&\leq 2 + \sup_{\|\mathbf{x}\|=1} \frac{1}{2} \sum_{t=1}^p \sum_{i,j \in \Omega_t, i \neq j} x_i^2 + x_j^2 = (2 + 2(|\Omega_t| - 1)) \sup_{\|\mathbf{x}\|=1} \sum_{k=1}^{p(p-1)/2} x_k^2 = 2p - 2,
\end{aligned}$$

with equality if and only if $x_1 = \dots = x_{p(p-1)/2} = \sqrt{\frac{2}{p(p-1)/2}}$ or $x_1 = \dots = x_{p(p-1)/2} = -\sqrt{\frac{2}{p(p-1)/2}}$.

Finally, we have

$$\begin{aligned}
\lambda_{\max}(\mathcal{L}^* \mathcal{L} + \mathfrak{d}^* \mathfrak{d}) &= \sup_{\|\mathbf{x}\|=1} \mathbf{x}^\top (\mathcal{L}^* \mathcal{L} + \mathfrak{d}^* \mathfrak{d}) \mathbf{x} \\
&\leq \sup_{\|\mathbf{x}\|=1} \mathbf{x}^\top (\mathcal{L}^* \mathcal{L}) \mathbf{x} + \sup_{\|\mathbf{y}\|=1} \mathbf{y}^\top (\mathfrak{d}^* \mathfrak{d}) \mathbf{y} \\
&= 4p - 2.
\end{aligned} \tag{29}$$

Note that the equality in (29) can be achieved because the eigenvectors of $\mathcal{L}^* \mathcal{L}$ and $\mathfrak{d}^* \mathfrak{d}$ associated with the maximum eigenvalue are the same. Therefore, we conclude that $\lambda_{\max}(\mathcal{L}^* \mathcal{L} + \mathfrak{d}^* \mathfrak{d}) = 4p - 2$, completing the proof. \blacksquare

C.2 Proof of Theorem 3

Proof To prove Theorem 3, we first establish the boundedness of the sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ generated by Algorithm 1 in Lemma 10, and the monotonicity of $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)$ in Lemma 11.

Lemma 10 *The sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ generated by Algorithm 1 is bounded.*

Proof Let \mathbf{w}^0 , \mathbf{Y}^0 and \mathbf{y}^0 be the initialization of the sequences $\{\mathbf{w}^l\}$, $\{\mathbf{Y}^l\}$ and $\{\mathbf{y}^l\}$, respectively, and $\|\mathbf{w}^0\|$, $\|\mathbf{Y}^0\|_{\text{F}}$ and $\|\mathbf{y}^0\|$ are bounded. We prove the boundedness of the sequence by induction.

Recall that the sequence $\{\Theta^l\}$ is established by

$$\Theta^l = \frac{1}{2\rho} \mathbf{U}^{l-1} \left(\mathbf{\Gamma}^{l-1} + \sqrt{(\mathbf{\Gamma}^{l-1})^2 + 4\rho \mathbf{I}} \right) \mathbf{U}^{l-1\top}, \quad (30)$$

where $\mathbf{\Gamma}^{l-1}$ contains the largest $p - k$ eigenvalues of $\rho \mathcal{L} \mathbf{w}^{l-1} - \mathbf{Y}^{l-1}$, and \mathbf{U}^{l-1} contains the corresponding eigenvectors. When $l = 1$, $\|\mathbf{\Gamma}^0\|_{\text{F}}$ is bounded since both $\|\mathbf{w}^0\|$ and $\|\mathbf{Y}^0\|_{\text{F}}$ are bounded. Therefore, we can conclude that $\|\Theta^1\|_{\text{F}}$ is bounded.

The sequence $\{\mathbf{w}^l\}$ is established by solving the subproblems of the form

$$\begin{aligned} \min_{\mathbf{w} \geq \mathbf{0}} & \frac{\rho}{2} \mathbf{w}^\top (\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w} - \langle \mathbf{w}, \mathcal{L}^* (\mathbf{Y}^{l-1} + \rho \Theta^l) - \mathfrak{d}^* (\mathbf{y}^{l-1} - \rho \mathbf{d}) \rangle \\ & + \frac{p + \nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathcal{L} \mathbf{w} \mathbf{x}_i}{\nu} \right). \end{aligned}$$

Let

$$g_l(\mathbf{w}) = \frac{\rho}{2} \mathbf{w}^\top (\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w} + \langle \mathbf{w}, \mathbf{a}^l \rangle + \frac{p + \nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathcal{L} \mathbf{w} \mathbf{x}_i}{\nu} \right), \quad (31)$$

where $\mathbf{a}^l = -\mathcal{L}^* (\mathbf{Y}^{l-1} + \rho \Theta^l) + \mathfrak{d}^* (\mathbf{y}^{l-1} - \rho \mathbf{d})$. Note that $\|\mathbf{a}^1\|$ is bounded because $\|\mathbf{Y}^0\|_{\text{F}}$, $\|\mathbf{y}^0\|$, and $\|\Theta^1\|_{\text{F}}$ are bounded.

From Lemma 2, we know that $\mathcal{L}^* \mathcal{L}$ is a positive definite matrix with minimum eigenvalue $\lambda_{\min}(\mathcal{L}^* \mathcal{L}) = 2$. On the other hand, $\mathfrak{d}^* \mathfrak{d}$ is a positive semi-definite matrix as follows,

$$\lambda_{\min}(\mathfrak{d}^* \mathfrak{d}) = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^\top \mathfrak{d}^* \mathfrak{d} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathfrak{d} \mathbf{x}, \mathfrak{d} \mathbf{x} \rangle}{\mathbf{x}^\top \mathbf{x}} \geq 0. \quad (32)$$

Since $\log \left(1 + \frac{\mathbf{x}_i^\top \mathcal{L} \mathbf{w} \mathbf{x}_i}{\nu} \right) \geq 0$ for any $\mathbf{w} \geq \mathbf{0}$, we have

$$\lim_{\|\mathbf{w}\| \rightarrow +\infty} g_1(\mathbf{w}) \geq \lim_{\|\mathbf{w}\| \rightarrow +\infty} \rho \mathbf{w}^\top \mathbf{w} + \langle \mathbf{w}, \mathbf{a}^1 \rangle = +\infty. \quad (33)$$

Thus $g_1(\mathbf{w})$ is coercive. Recall that we solve the optimization (31) by the MM framework. Hence, the objective function value is monotonically decreasing as a function of the iterations, and \mathbf{w}^l is a stationary point of (31). Then the coercivity of $g_1(\mathbf{w})$ yields the boundedness of $\|\mathbf{w}^1\|$.

According to the dual variable updates, we obtain

$$\mathbf{Y}^1 = \mathbf{Y}^0 + \rho (\Theta^1 - \mathcal{L} \mathbf{w}^1), \quad (34)$$

and

$$\mathbf{y}^1 = \mathbf{y}^0 + \rho (\mathfrak{d} \mathbf{w}^1 - \mathbf{d}). \quad (35)$$

It is obvious that both $\|\mathbf{Y}^1\|_{\text{F}}$ and $\|\mathbf{y}^1\|$ are bounded. Therefore, it holds for $l = 1$ that $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ is bounded.

Now we assume that $\{(\Theta^{l-1}, \mathbf{w}^{l-1}, \mathbf{Y}^{l-1}, \mathbf{y}^{l-1})\}$ is bounded for some $l \geq 1$, and check the boundedness of $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$. Similar to the proof in (30), we can prove that $\|\Theta^l\|_{\text{F}}$ is bounded. By (31), we can also obtain the boundedness of $\|\mathbf{w}^l\|$. We can also obtain that $\|\mathbf{Y}^l\|$ and $\|\mathbf{y}^l\|$ are bounded according to the boundedness of $\|\Theta^l\|_{\text{F}}$, $\|\mathbf{w}^l\|$, $\|\mathbf{Y}^{l-1}\|$ and $\|\mathbf{y}^{l-1}\|$. Thus, $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ is bounded, completing the induction. Therefore, we establish the boundedness of the sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$. ■

Lemma 11 The sequence $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)$ generated by Algorithm 1 is lower bounded, and

$$L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \leq L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l), \quad \forall l \in \mathbb{N}_+, \quad (36)$$

holds for any sufficiently large ρ .

Proof According to Problem (1) in the manuscript, we have

$$\begin{aligned} L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) &= \frac{p+\nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathcal{L} \mathbf{w}^l \mathbf{x}_i}{\nu} \right) - \log \det(\Theta^l + \mathbf{J}) + \langle \mathbf{y}^l, \mathfrak{d} \mathbf{w}^l - \mathbf{d} \rangle \\ &\quad + \frac{\rho}{2} \|\mathfrak{d} \mathbf{w}^l - \mathbf{d}\|_2^2 + \langle \mathbf{Y}^l, \Theta^l - \mathcal{L} \mathbf{w}^l \rangle + \frac{\rho}{2} \|\Theta^l - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2. \end{aligned} \quad (37)$$

We can see that the lower boundedness of the sequence $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)$ can be established by the boundedness of $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ in Lemma 10.

We first establish that

$$L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) \leq L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l), \quad \forall l \in \mathbb{N}_+. \quad (38)$$

One has

$$\begin{aligned} &L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \\ &= \underbrace{\langle \mathbf{y}^l, \mathfrak{d} \mathbf{w}^l - \mathbf{d} \rangle - \langle \mathbf{y}^{l+1}, \mathfrak{d} \mathbf{w}^{l+1} - \mathbf{d} \rangle}_{I_1} + \underbrace{\langle \mathbf{Y}^l, \Theta^{l+1} - \mathcal{L} \mathbf{w}^l \rangle - \langle \mathbf{Y}^{l+1}, \Theta^{l+1} - \mathcal{L} \mathbf{w}^{l+1} \rangle}_{I_2} \\ &\quad + r(\mathcal{L} \mathbf{w}^l) - r(\mathcal{L} \mathbf{w}^{l+1}) + \frac{\rho}{2} \|\mathfrak{d} \mathbf{w}^l - \mathbf{d}\|_2^2 - \frac{\rho}{2} \|\mathfrak{d} \mathbf{w}^{l+1} - \mathbf{d}\|_2^2 \\ &\quad + \frac{\rho}{2} \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2 - \frac{\rho}{2} \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^{l+1}\|_{\mathbb{F}}^2, \end{aligned} \quad (39)$$

$$\text{where } r(\mathbf{L}) = \frac{p+\nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathbf{L} \mathbf{x}_i}{\nu} \right).$$

From the dual variables updates, it is easy to see that

$$I_1 + I_2 = \langle \mathfrak{d}^* \mathbf{y}^l - \mathcal{L}^* \mathbf{Y}^l, \mathbf{w}^l - \mathbf{w}^{l+1} \rangle - \rho \|\mathfrak{d} \mathbf{w}^{l+1} - \mathbf{d}\|_2^2 - \rho \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^{l+1}\|_{\mathbb{F}}^2. \quad (40)$$

According to the convergence result of the majorization-minimization framework [35, 36], \mathbf{w}^{l+1} is a stationary point of the following problem

$$\underset{\mathbf{w} \geq \mathbf{0}}{\text{minimize}} \quad r(\mathcal{L} \mathbf{w}) + \frac{\rho}{2} \mathbf{w}^\top (\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w} - \langle \mathbf{w}, \mathcal{L}^* (\mathbf{Y}^l + \rho \Theta^{l+1}) - \mathfrak{d}^* (\mathbf{y}^l - \rho \mathbf{d}) \rangle. \quad (41)$$

The set of the stationary points for the optimization (41) is defined by

$$\mathcal{X} = \{ \mathbf{w} \mid \nabla g_l(\mathbf{w})^\top (\mathbf{z} - \mathbf{w}) \geq 0, \forall \mathbf{z} \geq \mathbf{0} \}, \quad (42)$$

where $g_l(\mathbf{w})$ is the objective function in (41). The existence of the limit point can be guaranteed by the the coercivity of $g_l(\mathbf{w})$, which has been established in the proof of Lemma 10. By taking $\mathbf{z} = \mathbf{w}^l$ and $\mathbf{w} = \mathbf{w}^{l+1}$ in (42), we obtain

$$(\mathcal{L}^* (\nabla r(\mathcal{L} \mathbf{w}^{l+1})) + \rho (\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w}^{l+1} - \mathcal{L}^* (\mathbf{Y}^l + \rho \Theta^{l+1}) + \mathfrak{d}^* (\mathbf{y}^l - \rho \mathbf{d}))^\top (\mathbf{w}^l - \mathbf{w}^{l+1}) \geq 0.$$

Thus, we have

$$\begin{aligned} \langle \mathfrak{d}^* \mathbf{y}^l - \mathcal{L}^* \mathbf{Y}^l, \mathbf{w}^l - \mathbf{w}^{l+1} \rangle &\geq - \langle \nabla r(\mathcal{L} \mathbf{w}^{l+1}), \mathcal{L} \mathbf{w}^l - \mathcal{L} \mathbf{w}^{l+1} \rangle \\ &\quad + \rho \langle -(\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w}^{l+1} + \mathcal{L}^* \Theta^{l+1} + \mathfrak{d}^* \mathbf{d}, \mathbf{w}^l - \mathbf{w}^{l+1} \rangle, \end{aligned} \quad (43)$$

Plugging (40) and (43) into (39), we obtain

$$\begin{aligned} &L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \\ &\geq \frac{\rho}{2} \|\mathfrak{d} \mathbf{w}^{l+1} - \mathfrak{d} \mathbf{w}^l\|_2^2 + \frac{\rho}{2} \|\mathcal{L} \mathbf{w}^{l+1} - \mathcal{L} \mathbf{w}^l\|_2^2 - \rho \|\mathfrak{d} \mathbf{w}^{l+1} - \mathbf{d}\|_2^2 - \rho \|\mathcal{L} \mathbf{w}^{l+1} - \Theta^{l+1}\|_{\mathbb{F}}^2 \\ &\quad + r(\mathcal{L} \mathbf{w}^l) - r(\mathcal{L} \mathbf{w}^{l+1}) - \langle \nabla r(\mathcal{L} \mathbf{w}^{l+1}), \mathcal{L} \mathbf{w}^l - \mathcal{L} \mathbf{w}^{l+1} \rangle \\ &\geq \frac{\rho}{2} \|\mathfrak{d} \mathbf{w}^{l+1} - \mathfrak{d} \mathbf{w}^l\|_2^2 + \frac{\rho - L_r}{2} \|\mathcal{L} \mathbf{w}^{l+1} - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2 - \frac{1}{\rho} \|\mathbf{y}^{l+1} - \mathbf{y}^l\|_2^2 - \frac{1}{\rho} \|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|_{\mathbb{F}}^2, \end{aligned} \quad (44)$$

where the last inequality is due to the fact that $r(\mathbf{L})$ is a concave function and has L_r -Lipschitz continuous gradient, in which $L_r > 0$ is a constant, thus we have

$$r(\mathcal{L}\mathbf{w}^l) - r(\mathcal{L}\mathbf{w}^{l+1}) - \langle \nabla r(\mathcal{L}\mathbf{w}^{l+1}), \mathcal{L}\mathbf{w}^l - \mathcal{L}\mathbf{w}^{l+1} \rangle \geq -\frac{L_r}{2} \|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\mathbb{F}}^2. \quad (45)$$

By calculation, we obtain that if ρ is sufficiently large such that

$$\rho \geq \max \left(2L_r, \max_l \frac{C \left(\|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|_{\mathbb{F}}^2 + \|\mathbf{y}^{l+1} - \mathbf{y}^l\|_2^2 \right)^{\frac{1}{2}}}{\left(\|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\mathbb{F}}^2 + \|\partial\mathbf{w}^{l+1} - \partial\mathbf{w}^l\|_2^2 \right)^{\frac{1}{2}}} \right), \quad (46)$$

holds with some constant $C > 2$, then, together with (38), we conclude that

$$L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) \geq L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) \geq L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}), \quad (47)$$

for any $l \in \mathbb{N}_+$. ■

Now we are ready to prove Theorem 3. By Lemma 10, the sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ generated by Algorithm 1 is bounded. Therefore, there exists at least one convergent subsequence $\{(\Theta^{l_s}, \mathbf{w}^{l_s}, \mathbf{Y}^{l_s}, \mathbf{y}^{l_s})\}_{s \in \mathbb{N}}$, which converges to the limit point denoted by $\{(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})\}$. By Lemma 11, we obtain that the sequence $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)$ defined in (37) is monotonically decreasing and lower bounded, implying that $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)$ is convergent.

We can then obtain $\lim_{l \rightarrow +\infty} L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) = L_\rho(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})$. Then, (44), (46), and (47) together yield

$$\begin{aligned} L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \\ \geq \frac{C^2 - 4}{4} \rho \left(\|\mathcal{L}\mathbf{w}^{l+1} - \Theta^{l+1}\|_{\mathbb{F}}^2 + \|\partial\mathbf{w}^{l+1} - \mathbf{d}\|_2^2 \right). \end{aligned} \quad (48)$$

The convergence of $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{Y}^l, \mathbf{y}^l)$ yields

$$\lim_{l \rightarrow +\infty} \|\mathcal{L}\mathbf{w}^l - \Theta^l\|_{\mathbb{F}} = 0, \quad \text{and} \quad \lim_{l \rightarrow +\infty} \|\partial\mathbf{w}^l - \mathbf{d}\|_2 = 0. \quad (49)$$

By the updating of \mathbf{Y}^{l+1} and \mathbf{y}^{l+1} , we can get

$$\lim_{l \rightarrow +\infty} \|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|_{\mathbb{F}} = 0, \quad \text{and} \quad \lim_{l \rightarrow +\infty} \|\mathbf{y}^{l+1} - \mathbf{y}^l\|_2 = 0. \quad (50)$$

Together with (44), we obtain

$$\lim_{l \rightarrow +\infty} \|\partial\mathbf{w}^{l+1} - \partial\mathbf{w}^l\|_2 = 0 \quad \text{and} \quad \lim_{l \rightarrow +\infty} \|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\mathbb{F}} = 0. \quad (51)$$

For the limit point $\{(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})\}$ of any subsequence $\{(\Theta^{l_s}, \mathbf{w}^{l_s}, \mathbf{Y}^{l_s}, \mathbf{y}^{l_s})\}_{s \in \mathbb{N}}$, Θ^{l_∞} minimizes the following subproblem

$$\begin{aligned} \Theta^{l_\infty} &= \arg \min_{\Theta \succeq \mathbf{0}} -\log \det(\Theta + \mathbf{J}) + \langle \Theta, \mathbf{Y}^{l_\infty-1} \rangle + \frac{\rho}{2} \|\Theta - \mathcal{L}\mathbf{w}^{l_\infty-1}\|_{\mathbb{F}}^2 \\ &= \arg \min_{\Theta \succeq \mathbf{0}} -\log \det(\Theta + \mathbf{J}) + \langle \Theta, \mathbf{Y}^{l_\infty} \rangle + \frac{\rho}{2} \|\Theta - \mathcal{L}\mathbf{w}^{l_\infty}\|_{\mathbb{F}}^2 \\ &\quad - \langle \Theta, \mathbf{Y}^{l_\infty} - \mathbf{Y}^{l_\infty-1} \rangle + \rho \langle \Theta, \mathcal{L}\mathbf{w}^{l_\infty} - \mathcal{L}\mathbf{w}^{l_\infty-1} \rangle. \end{aligned}$$

By (50) and (51), we conclude that Θ^{l_∞} satisfies the condition of stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{V}, \mathbf{Y}, \mathbf{y})$ with respect to Θ . Similarly, from the convergence results of the MM framework, we know that \mathbf{w}^{l_∞} is a stationary point of the subproblem

$$\begin{aligned} \underset{\mathbf{w} \geq \mathbf{0}}{\text{minimize}} \quad & \frac{\rho}{2} \mathbf{w}^\top (\partial^* \partial + \mathcal{L}^* \mathcal{L}) \mathbf{w} + \langle \mathbf{w}, \partial^* (\mathbf{y}^{l_\infty} - \rho \mathbf{d}) \rangle - \langle \mathbf{w}, \mathcal{L}^* (\mathbf{Y}^{l_\infty} + \rho \Theta^{l_\infty}) \rangle \\ & + \langle \mathcal{L}\mathbf{w}, \mathbf{Y}^{l_\infty} - \mathbf{Y}^{l_\infty-1} \rangle + \langle \partial\mathbf{w}, \mathbf{y}^{l_\infty-1} - \mathbf{y}^{l_\infty} \rangle + r(\mathcal{L}\mathbf{w}). \end{aligned}$$

By (50) and (51), \mathbf{w}^{l_∞} satisfies the condition of stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{Y}, \mathbf{y})$ with respect to \mathbf{w} .

To sum up, we can conclude that any limit point $\{(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})\}$ of the sequence generated by Algorithm 1 is a stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{Y}, \mathbf{y})$. ■

C.3 Proof of Theorem 4

Proof Similarly to the proof conducted for Theorem 3, to prove Theorem 4, we first establish the boundedness of the sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ generated by Algorithm 2 in Lemma 12, and the monotonicity of $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)$ in Lemma 13.

Lemma 12 *The sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ generated by Algorithm 2 is bounded.*

Proof Let $\mathbf{w}^0, \mathbf{V}^0, \mathbf{Y}^0$ and \mathbf{y}^0 be the initialization of the sequences $\{\mathbf{w}^l\}, \{\mathbf{V}^l\}, \{\mathbf{Y}^l\}$ and $\{\mathbf{y}^l\}$, respectively, and $\|\mathbf{w}^0\|, \|\mathbf{V}^0\|_F, \|\mathbf{Y}^0\|_F$ and $\|\mathbf{y}^0\|$ are bounded.

We prove the lemma by induction. Recall that the sequence $\{\Theta^l\}$ is established by

$$\Theta^l = \frac{1}{2\rho} \mathbf{U}^{l-1} \left(\Gamma^{l-1} + \sqrt{(\Gamma^{l-1})^2 + 4\rho \mathbf{I}} \right) \mathbf{U}^{l-1\top}, \quad (52)$$

where Γ^{l-1} contains the largest $p - k$ eigenvalues of $\rho \mathcal{L} \mathbf{w}^{l-1} - \mathbf{Y}^{l-1}$, and \mathbf{U}^{l-1} contains the corresponding eigenvectors. When $l = 1$, $\|\Gamma^0\|_F$ is bounded since both $\|\mathbf{w}^0\|$ and $\|\mathbf{Y}^0\|_F$ are bounded. Therefore, we can conclude that $\|\Theta^1\|_F$ is bounded.

The sequence $\{\mathbf{w}^l\}$ is established by solving the subproblems that tantamount to (31) except for the additional linear term $\eta \langle \mathbf{w}, \mathcal{L}^* (\mathbf{V}^l \mathbf{V}^{l\top}) \rangle$. Hence, $g_1(\mathbf{w}) + \eta \langle \mathbf{w}, \mathcal{L}^* (\mathbf{V}^1 \mathbf{V}^{1\top}) \rangle$ is still coercive, which implies that $\|\mathbf{w}^1\|$ is bounded.

The feasible set of \mathbf{V} is the set of ordered orthonormal k -frames in \mathbb{R}^p , thus $\|\mathbf{V}^l\|_F$ is bounded for any $l \geq 1$.

From the updates of \mathbf{Y} and \mathbf{y} , it is obvious that both $\|\mathbf{Y}^1\|_F$ and $\|\mathbf{y}^1\|$ are bounded. Therefore, it holds for $l = 1$ that $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ is bounded.

Assume that $\{(\Theta^{l-1}, \mathbf{w}^{l-1}, \mathbf{V}^{l-1}, \mathbf{Y}^{l-1}, \mathbf{y}^{l-1})\}$ is bounded for some $l \geq 1$, i.e., each term in $\{(\Theta^{l-1}, \mathbf{w}^{l-1}, \mathbf{V}^{l-1}, \mathbf{Y}^{l-1}, \mathbf{y}^{l-1})\}$ is bounded under ℓ_2 -norm or Frobenius norm. Following (52), we can obtain that $\|\Theta^l\|_F$ is bounded.

Due to the coercivity of $g_l(\mathbf{w}) + \eta \langle \mathbf{w}, \mathcal{L}^* (\mathbf{V}^l \mathbf{V}^{l\top}) \rangle$, we can get that $\|\mathbf{w}^l\|$ is bounded. It is also obvious that $\|\mathbf{Y}^l\|$ and $\|\mathbf{y}^l\|$ are bounded, because of the boundedness of $\|\Theta^l\|_F, \|\mathbf{w}^l\|, \|\mathbf{Y}^{l-1}\|$ and $\|\mathbf{y}^{l-1}\|$. Thus, $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ is bounded, completing the induction. Therefore, we can conclude that the sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ is bounded. ■

Lemma 13 *The sequence $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)$ generated by Algorithm 2 is lower bounded, and*

$$L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{V}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \leq L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l), \quad \forall l \in \mathbb{N}_+, \quad (53)$$

holds for any sufficiently large ρ .

Proof According to Problem (16) in the manuscript, we have

$$\begin{aligned} L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) &= \frac{p + \nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathcal{L} \mathbf{w}^l \mathbf{x}_i}{\nu} \right) + \text{tr} \left(\mathcal{L} \mathbf{w}^l \left(\eta \mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) \\ &\quad - \log \det^* (\Theta^l) + \langle \mathbf{y}^l, \partial \mathbf{w}^l - \mathbf{d} \rangle + \frac{\rho}{2} \|\partial \mathbf{w}^l - \mathbf{d}\|_2^2 \\ &\quad + \langle \mathbf{Y}^l, \Theta^l - \mathcal{L} \mathbf{w}^l \rangle + \frac{\rho}{2} \|\Theta^l - \mathcal{L} \mathbf{w}^l\|_F^2. \end{aligned} \quad (54)$$

We can see that the lower boundedness of the sequence $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)$ can be established by the boundedness of $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ in Lemma 12.

We first establish that

$$L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) \leq L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l), \quad \forall l \in \mathbb{N}_+. \quad (55)$$

We have

$$\begin{aligned} L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) &= \frac{p+\nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathcal{L} \mathbf{w}^l \mathbf{x}_i}{\nu} \right) + \text{tr} \left(\mathcal{L} \mathbf{w}^l \left(\eta \mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) \\ &\quad - \log \det^* (\Theta^{l+1}) + \langle \mathbf{y}^l, \partial \mathbf{w}^l - \mathbf{d} \rangle + \frac{\rho}{2} \|\partial \mathbf{w}^l - \mathbf{d}\|_2^2 \\ &\quad + \langle \mathbf{Y}^l, \Theta^{l+1} - \mathcal{L} \mathbf{w}^l \rangle + \frac{\rho}{2} \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2. \end{aligned}$$

Then we obtain

$$\begin{aligned} L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) &= -\log \det^* (\Theta^{l+1}) + \langle \mathbf{Y}^l, \Theta^{l+1} \rangle \\ &\quad + \frac{\rho}{2} \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2 - \left(-\log \det^* (\Theta^l) + \langle \mathbf{Y}^l, \Theta^l \rangle + \frac{\rho}{2} \|\Theta^l - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2 \right). \end{aligned}$$

Note that Θ^{l+1} minimizes the objective function

$$\Theta^{l+1} = \arg \min_{\substack{\text{rank}(\Theta) = p-k \\ \Theta \succeq \mathbf{0}}} -\log \det^* (\Theta) + \langle \Theta, \mathbf{Y}^l \rangle + \frac{\rho}{2} \|\Theta - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2. \quad (56)$$

Therefore

$$L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) \leq 0 \quad (57)$$

holds for any $l \in \mathbb{N}_+$.

One has

$$\begin{aligned} &L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{V}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \\ &= r(\mathcal{L} \mathbf{w}^l) - r(\mathcal{L} \mathbf{w}^{l+1}) + \underbrace{\text{tr} \left(\eta \mathcal{L} \mathbf{w}^l \left(\mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) - \text{tr} \left(\eta \mathcal{L} \mathbf{w}^{l+1} \left(\mathbf{V}^{l+1} (\mathbf{V}^{l+1})^\top \right) \right)}_{I_1} \\ &\quad + \underbrace{\langle \mathbf{y}^l, \partial \mathbf{w}^l - \mathbf{d} \rangle - \langle \mathbf{y}^{l+1}, \partial \mathbf{w}^{l+1} - \mathbf{d} \rangle}_{I_2} + \underbrace{\langle \mathbf{Y}^l, \Theta^{l+1} - \mathcal{L} \mathbf{w}^l \rangle - \langle \mathbf{Y}^{l+1}, \Theta^{l+1} - \mathcal{L} \mathbf{w}^{l+1} \rangle}_{I_3} \\ &\quad + \frac{\rho}{2} \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^l\|_{\mathbb{F}}^2 - \frac{\rho}{2} \|\Theta^{l+1} - \mathcal{L} \mathbf{w}^{l+1}\|_{\mathbb{F}}^2 + \frac{\rho}{2} \|\partial \mathbf{w}^l - \mathbf{d}\|_2^2 - \frac{\rho}{2} \|\partial \mathbf{w}^{l+1} - \mathbf{d}\|_2^2, \quad (58) \end{aligned}$$

where $r(\mathbf{L}) = \frac{p+\nu}{n} \sum_{i=1}^n \log \left(1 + \frac{\mathbf{x}_i^\top \mathbf{L} \mathbf{x}_i}{\nu} \right)$. The term I_1 can be written as

$$\begin{aligned} I_1 &= \text{tr} \left(\eta \mathcal{L} \mathbf{w}^l \left(\mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) - \text{tr} \left(\eta \mathcal{L} \mathbf{w}^{l+1} \left(\mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) \\ &\quad + \underbrace{\text{tr} \left(\eta \mathcal{L} \mathbf{w}^{l+1} \left(\mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) - \text{tr} \left(\eta \mathcal{L} \mathbf{w}^{l+1} \left(\mathbf{V}^{l+1} (\mathbf{V}^{l+1})^\top \right) \right)}_{I_{1a}}. \end{aligned}$$

Note that \mathbf{V}^{l+1} is the optimal solution of the problem

$$\min_{\mathbf{V} \in \mathbb{R}^{p \times k}} \text{tr} (\mathbf{V}^\top \mathcal{L} \mathbf{w}^{l+1} \mathbf{V}), \quad \text{subject to } \mathbf{V}^\top \mathbf{V} = \mathbf{I}. \quad (59)$$

Thus the term $I_{1a} \geq 0$, and we can obtain

$$I_1 \geq \text{tr} \left(\eta \mathcal{L} \mathbf{w}^l \left(\mathbf{V}^l (\mathbf{V}^l)^\top \right) \right) - \text{tr} \left(\eta \mathcal{L} \mathbf{w}^{l+1} \left(\mathbf{V}^l (\mathbf{V}^l)^\top \right) \right). \quad (60)$$

For the term I_2 , we have

$$\begin{aligned} I_2 &= \langle \mathbf{y}^l, \partial \mathbf{w}^l - \mathbf{d} \rangle - \langle \mathbf{y}^{l+1}, \partial \mathbf{w}^{l+1} - \mathbf{d} \rangle - \rho \langle \partial \mathbf{w}^{l+1} - \mathbf{d}, \partial \mathbf{w}^{l+1} - \mathbf{d} \rangle \\ &= \langle \partial^* \mathbf{y}^l, \mathbf{w}^l - \mathbf{w}^{l+1} \rangle - \rho \|\partial \mathbf{w}^{l+1} - \mathbf{d}\|_2^2, \quad (61) \end{aligned}$$

where the first equality is due to the updating of \mathbf{y}^{l+1} as below

$$\mathbf{y}^{l+1} = \mathbf{y}^l + \rho (\mathfrak{d}\mathbf{w}^{l+1} - \mathbf{d}). \quad (62)$$

For the term I_3 , similarly, we have

$$\begin{aligned} I_3 &= \langle \mathbf{Y}^l, \Theta^{l+1} - \mathcal{L}\mathbf{w}^l \rangle - \langle \mathbf{Y}^l, \Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1} \rangle - \rho \langle \Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1}, \Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1} \rangle \\ &= \langle \mathcal{L}^* \mathbf{Y}^l, \mathbf{w}^{l+1} - \mathbf{w}^l \rangle - \rho \|\Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1}\|_{\mathbb{F}}^2, \end{aligned} \quad (63)$$

where the first equality follows from

$$\mathbf{Y}^{l+1} = \mathbf{Y}^l + \rho (\Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1}). \quad (64)$$

Therefore, we can obtain

$$I_2 + I_3 = \langle \mathfrak{d}^* \mathbf{y}^l - \mathcal{L}^* \mathbf{Y}^l, \mathbf{w}^l - \mathbf{w}^{l+1} \rangle - \rho \|\mathfrak{d}\mathbf{w}^{l+1} - \mathbf{d}\|_2^2 - \rho \|\Theta^{l+1} - \mathcal{L}\mathbf{w}^{l+1}\|_{\mathbb{F}}^2. \quad (65)$$

Recall that \mathbf{w}^{l+1} is a stationary point of the problem

$$\begin{aligned} \underset{\mathbf{w} \geq \mathbf{0}}{\text{minimize}} \quad & r(\mathcal{L}\mathbf{w}) + \frac{\rho}{2} \mathbf{w}^\top (\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w} \\ & - \langle \mathbf{w}, \mathcal{L}^* (\mathbf{Y}^l + \rho \Theta^{l+1} - \eta \mathbf{V}^l \mathbf{V}^{l\top}) - \mathfrak{d}^* (\mathbf{y}^l - \rho \mathbf{d}) \rangle. \end{aligned} \quad (66)$$

The set of the stationary points for the optimization (66) is defined by

$$\mathcal{X} = \{ \mathbf{w} | \nabla q_l(\mathbf{w})^\top (\mathbf{z} - \mathbf{w}) \geq 0, \forall \mathbf{z} \geq \mathbf{0} \}, \quad (67)$$

where $q_l(\mathbf{w})$ is the objective function in (66). By taking $\mathbf{z} = \mathbf{w}^l$ and $\mathbf{w} = \mathbf{w}^{l+1}$ in (67), we obtain

$$\begin{aligned} \langle \mathfrak{d}^* \mathbf{y}^l - \mathcal{L}^* \mathbf{Y}^l, \mathbf{w}^l - \mathbf{w}^{l+1} \rangle &\geq - \langle \nabla r(\mathcal{L}\mathbf{w}^{l+1}), \mathcal{L}\mathbf{w}^l - \mathcal{L}\mathbf{w}^{l+1} \rangle \\ &+ \rho \left\langle - (\mathfrak{d}^* \mathfrak{d} + \mathcal{L}^* \mathcal{L}) \mathbf{w}^{l+1} + \mathcal{L}^* \Theta^{l+1} + \mathfrak{d}^* \mathbf{d} - \frac{\eta}{\rho} \mathbf{V}^l \mathbf{V}^{l\top}, \mathbf{w}^l - \mathbf{w}^{l+1} \right\rangle, \end{aligned} \quad (68)$$

Substituting (65) and (68) into (58), we obtain

$$\begin{aligned} & L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{V}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \\ & \geq \frac{\rho}{2} \|\mathfrak{d}\mathbf{w}^{l+1} - \mathfrak{d}\mathbf{w}^l\|_2^2 + \frac{\rho}{2} \|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_2^2 - \rho \|\mathfrak{d}\mathbf{w}^{l+1} - \mathbf{d}\|_2^2 - \rho \|\mathcal{L}\mathbf{w}^{l+1} - \Theta^{l+1}\|_{\mathbb{F}}^2 \\ & \quad + r(\mathcal{L}\mathbf{w}^l) - r(\mathcal{L}\mathbf{w}^{l+1}) - \langle \nabla r(\mathcal{L}\mathbf{w}^{l+1}), \mathcal{L}\mathbf{w}^l - \mathcal{L}\mathbf{w}^{l+1} \rangle \\ & \geq \frac{\rho}{2} \|\mathfrak{d}\mathbf{w}^{l+1} - \mathfrak{d}\mathbf{w}^l\|_2^2 + \frac{\rho - L_r}{2} \|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\mathbb{F}}^2 - \frac{1}{\rho} \|\mathbf{y}^{l+1} - \mathbf{y}^l\|_2^2 - \frac{1}{\rho} \|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|_{\mathbb{F}}^2, \end{aligned} \quad (69)$$

where the last inequality is due to the fact that $r(\mathbf{L})$ is a concave function and has L_r -Lipschitz continuous gradient where $L_r > 0$ is a constant, and thus we obtain

$$r(\mathcal{L}\mathbf{w}^l) - r(\mathcal{L}\mathbf{w}^{l+1}) - \langle \nabla r(\mathcal{L}\mathbf{w}^{l+1}), \mathcal{L}\mathbf{w}^l - \mathcal{L}\mathbf{w}^{l+1} \rangle \geq -\frac{L_r}{2} \|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\mathbb{F}}^2. \quad (70)$$

By calculation, we obtain that if ρ is sufficiently large such that

$$\rho \geq \max \left(2L_r, \max_l \frac{C \left(\|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|_{\mathbb{F}}^2 + \|\mathbf{y}^{l+1} - \mathbf{y}^l\|_2^2 \right)^{\frac{1}{2}}}{\left(\|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\mathbb{F}}^2 + \|\mathfrak{d}\mathbf{w}^{l+1} - \mathfrak{d}\mathbf{w}^l\|_2^2 \right)^{\frac{1}{2}}} \right), \quad (71)$$

holds with some constant $C > 2$, then, together with (57), we obtain that

$$L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) \geq L_\rho(\Theta^{l+1}, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) \geq L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{V}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}), \quad (72)$$

for any $l \in \mathbb{N}_+$. ■

Now we are ready to prove Theorem 4. By Lemma 12, the sequence $\{(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)\}$ is bounded. Therefore, there exists at least one convergent subsequence $\{(\Theta^{l_s}, \mathbf{w}^{l_s}, \mathbf{V}^{l_s}, \mathbf{Y}^{l_s}, \mathbf{y}^{l_s})\}_{s \in \mathbb{N}}$, which converges to a limit point denoted by $\{(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{V}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})\}$. By Lemma 13, we obtain that $L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l)$ is monotonically decreasing and lower bounded, and thus is convergent. Note that the function $\log \det^*(\Theta)$ is continuous over the set $\mathcal{S} = \{\Theta \in \mathcal{S}_+^p | \text{rank}(\Theta) = p - k\}$.

We can then obtain

$$\lim_{l \rightarrow +\infty} L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) = L_\rho(\Theta^\infty, \mathbf{w}^\infty, \mathbf{V}^\infty, \mathbf{Y}^\infty, \mathbf{y}^\infty) = L_\rho(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{V}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty}).$$

Then, (69), (71) and (72) together yields

$$\begin{aligned} L_\rho(\Theta^l, \mathbf{w}^l, \mathbf{V}^l, \mathbf{Y}^l, \mathbf{y}^l) - L_\rho(\Theta^{l+1}, \mathbf{w}^{l+1}, \mathbf{V}^{l+1}, \mathbf{Y}^{l+1}, \mathbf{y}^{l+1}) \\ \geq \frac{C^2 - 4}{4} \rho \left(\|\mathcal{L}\mathbf{w}^{l+1} - \Theta^{l+1}\|_{\text{F}}^2 + \|\partial\mathbf{w}^{l+1} - \mathbf{d}\|_2^2 \right). \end{aligned} \quad (73)$$

Thus, we obtain

$$\lim_{l \rightarrow +\infty} \|\mathcal{L}\mathbf{w}^l - \Theta^l\|_{\text{F}} = 0, \quad \text{and} \quad \lim_{l \rightarrow +\infty} \|\partial\mathbf{w}^l - \mathbf{d}\|_2 = 0. \quad (74)$$

Obviously, $\|\mathcal{L}\mathbf{w}^{l_s} - \Theta^{l_s}\|_{\text{F}} \rightarrow 0$ and $\|\partial\mathbf{w}^{l_s} - \mathbf{d}\|_2 \rightarrow 0$ also hold for any subsequence as $s \rightarrow +\infty$, which implies that \mathbf{Y}^{l_∞} and \mathbf{y}^{l_∞} satisfy the condition of stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{V}, \mathbf{Y}, \mathbf{y})$ with respect to \mathbf{Y} and \mathbf{y} , respectively. By (64) and (62), we also have

$$\lim_{l \rightarrow +\infty} \|\mathbf{Y}^{l+1} - \mathbf{Y}^l\|_{\text{F}} = 0, \quad \text{and} \quad \lim_{l \rightarrow +\infty} \|\mathbf{y}^{l+1} - \mathbf{y}^l\|_2 = 0. \quad (75)$$

Together with (69), we obtain

$$\lim_{l \rightarrow +\infty} \|\partial\mathbf{w}^{l+1} - \partial\mathbf{w}^l\|_2 = 0 \quad \text{and} \quad \lim_{l \rightarrow +\infty} \|\mathcal{L}\mathbf{w}^{l+1} - \mathcal{L}\mathbf{w}^l\|_{\text{F}} = 0. \quad (76)$$

Recall that \mathbf{V}^l contains the k eigenvectors associated with the k smallest eigenvalues of $\mathcal{L}\mathbf{w}^l$, and thus we can check that

$$\lim_{l \rightarrow +\infty} \|\mathbf{V}^{l+1} - \mathbf{V}^l\|_{\text{F}} = 0. \quad (77)$$

For the limit point $\{(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{V}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})\}$ of any subsequence $\{(\Theta^{l_s}, \mathbf{w}^{l_s}, \mathbf{V}^{l_s}, \mathbf{Y}^{l_s}, \mathbf{y}^{l_s})\}_{s \in \mathbb{N}}$, Θ^{l_∞} minimizes the following subproblem

$$\begin{aligned} \Theta^{l_\infty} &= \arg \min_{\substack{\text{rank}(\Theta) = p - k \\ \Theta \succeq \mathbf{0}}} -\log \det^*(\Theta) + \langle \Theta, \mathbf{Y}^{l_\infty - 1} \rangle + \frac{\rho}{2} \|\Theta - \mathcal{L}\mathbf{w}^{l_\infty - 1}\|_{\text{F}}^2 \\ &= \arg \min_{\substack{\text{rank}(\Theta) = p - k \\ \Theta \succeq \mathbf{0}}} -\log \det^*(\Theta) + \langle \Theta, \mathbf{Y}^{l_\infty} \rangle + \frac{\rho}{2} \|\Theta - \mathcal{L}\mathbf{w}^{l_\infty}\|_{\text{F}}^2 \\ &\quad - \langle \Theta, \mathbf{Y}^{l_\infty} - \mathbf{Y}^{l_\infty - 1} \rangle + \rho \langle \Theta, \mathcal{L}\mathbf{w}^{l_\infty} - \mathcal{L}\mathbf{w}^{l_\infty - 1} \rangle. \end{aligned}$$

By (75) and (76), we conclude that Θ^{l_∞} satisfies the condition of stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{V}, \mathbf{Y}, \mathbf{y})$ with respect to Θ . Similarly, \mathbf{w}^{l_∞} minimizes the subproblem

$$\begin{aligned} \mathbf{w}^{l_\infty} &= \arg \min_{\mathbf{w} \geq \mathbf{0}} r(\mathcal{L}\mathbf{w}) + \frac{\rho}{2} \mathbf{w}^\top (\partial^* \partial + \mathcal{L}^* \mathcal{L}) \mathbf{w} + \langle \mathbf{w}, \partial^* (\mathbf{y}^{l_\infty - 1} - \rho \mathbf{d}) \rangle \\ &\quad + \left\langle \mathbf{w}, \mathcal{L}^* \left(\eta \mathbf{V}^{l_\infty - 1} (\mathbf{V}^{l_\infty - 1})^\top - \mathbf{Y}^{l_\infty - 1} - \rho \Theta^{l_\infty} \right) \right\rangle \\ &= \arg \min_{\mathbf{w} \geq \mathbf{0}} r(\mathcal{L}\mathbf{w}) + \frac{\rho}{2} \mathbf{w}^\top (\partial^* \partial + \mathcal{L}^* \mathcal{L}) \mathbf{w} + \langle \mathbf{w}, \partial^* (\mathbf{y}^{l_\infty} - \rho \mathbf{d}) \rangle \\ &\quad + \left\langle \mathbf{w}, \mathcal{L}^* \left(\eta \mathbf{V}^{l_\infty} (\mathbf{V}^{l_\infty})^\top - \mathbf{Y}^{l_\infty} - \rho \Theta^{l_\infty} \right) \right\rangle + \langle \mathcal{L}\mathbf{w}, \mathbf{Y}^{l_\infty} - \mathbf{Y}^{l_\infty - 1} \rangle \\ &\quad + \langle \partial\mathbf{w}, \mathbf{y}^{l_\infty - 1} - \mathbf{y}^{l_\infty} \rangle + \eta \left\langle \mathcal{L}\mathbf{w}, \mathbf{V}^{l_\infty - 1} (\mathbf{V}^{l_\infty - 1})^\top - \mathbf{V}^{l_\infty} (\mathbf{V}^{l_\infty})^\top \right\rangle. \end{aligned}$$

By (75), (76) and (77), \mathbf{w}^{l_∞} satisfies the condition of stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{V}, \mathbf{Y}, \mathbf{y})$ with respect to \mathbf{w} . \mathbf{V}^∞ minimizes the subproblem

$$\mathbf{V}^{l_\infty} = \arg \min_{\mathbf{V} \in \mathbb{R}^{p \times k}} \text{tr}(\mathbf{V}^\top \mathcal{L} \mathbf{w}^{l_\infty} \mathbf{V}), \quad \text{subject to } \mathbf{V}^\top \mathbf{V} = \mathbf{I},$$

which implies that \mathbf{V}^{l_∞} satisfies the condition of stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{V}, \mathbf{Y}, \mathbf{y})$ with respect to \mathbf{V} . To sum up, we can conclude that any limit point $\{(\Theta^{l_\infty}, \mathbf{w}^{l_\infty}, \mathbf{V}^{l_\infty}, \mathbf{Y}^{l_\infty}, \mathbf{y}^{l_\infty})\}$ of the sequence generated by Algorithm 2 is a stationary point of $L_\rho(\Theta, \mathbf{w}, \mathbf{V}, \mathbf{Y}, \mathbf{y})$. ■