Supplementary Materials

Supplementary Materials contain A) an application of the algorithm to analyze death cause data in 367 towns in Taiwan, B) convergence analysis of the iterative scheme (2.11) and C) derivation of the stopping criteria (2.13) and (2.14).

A Application: causes of death in 367 towns in Taiwan

The data contain yearly numbers of dead people in terms of their death causes in 367 towns in Taiwan. They are collected over a 5-year period between 2008 and 2012. There are 29 death causes in the data. The 29th cause is a combination of the main death causes for about 80 to 95% of dead people in each town. The rest of 28 causes are the "rare" death causes. By rare we mean those occurring with small probabilities. The data are collected and compiled by the Ministry of Health and Welfare and belong to Open Government Data project initiated by the National Development Council of the Taiwanese government. Such a project aims to "... improve inter-organization data exchange, improve administrative efficiency, satisfy requirements of the people, and empower citizens to monitor government operations." (Chen, 2012). More details on Taiwanese government's Open Government Data project can be found in Chen (2012). In addition, the dataset used in this section can be found in Open Government Data (2012) and downloaded from `http://data.gov.tw/node/5965`.

A.1 Model estimation

Our aim is to cluster towns that have a similar pattern in the rare death causes. To begin our statistical modeling, we assume there are $p+1$ death causes. Let π_{ja} denote the probability of a dead individual in town j who dies in cause a. We model the logarithm of the odds for the death cause a against the death cause $p+1$ in town j by

$$\log\left(\frac{\pi_{ja}}{\pi_{j(p+1)}}\right) = \beta_{ja}. \tag{A.1}$$

From (A.1) we have

$$\pi_{ja} = \frac{\exp(\beta_{ja})}{1+\sum_{a'=1}^{p}\exp(\beta_{ja'})} \text{ for } a = 1, 2, \cdots, p,$$

and $\pi_{ja} = [1+\sum_{a'=1}^{p}\exp(\beta_{ja'})]^{-1}$ for $a = p+1$. The parameter vector $\beta_j = (\beta_{j1}, \beta_{j2}, \cdots, \beta_{jp})$ characterizes the distribution of death causes in town j. In this sense β_j is a profile of death causes in town j. We assume there are m towns, and each town has c observations. Under the assumptions given above, we can construct a likelihood function for parameter estimation. The minus logarithm of the likelihood function is

$$\begin{aligned} l(\beta) &= \sum_{j=1}^{m} l(\beta_j) \\ &= -\sum_{j=1}^{m}\left(\sum_{t=1}^{c}\left\{\sum_{a'=1}^{p} y_{jta'}\beta_{ja'} - n_{jt}\log\left[1+\sum_{a'=1}^{p}\exp(\beta_{ja'})\right]\right\}\right), \end{aligned}$$

where $n_{jt} = \sum_{a'=1}^{p+1} y_{jta'}$ is the number of deaths in observation t in town j. The fused group lasso estimate of $\beta = (\beta_1, \beta_2, \cdots, \beta_m)$ is defined by

$$\widehat{\beta} = \arg\min_{\beta_1,\beta_2,\cdots,\beta_m}\left\{\sum_{j=1}^{m} l(\beta_j) + \lambda\sum_{(j,k)\in\mathcal{H}} ||\beta_j - \beta_k||_2\right\}, \tag{A.2}$$

where the comparison graph $\mathcal{H}$ is defined by

$$\mathcal{H} = \left\{(j,k) : d(\widehat{\phi}_j, \widehat{\phi}_k) \leq b_{\text{thr}}\right\}.$$

Here $d(\widehat{\phi}_j, \widehat{\phi}_k)$ is a distance function, $b_{\text{thr}} \geq 0$ is a threshold value, and $\widehat{\phi}_j = (\widehat{\phi}_{j1}, \widehat{\phi}_{j2}, \cdots, \widehat{\phi}_{jp})$ is a p-dimensional vector in which each $\widehat{\phi}_{ja}$ is defined by

$$\widehat{\phi}_{ja} = \frac{\sum_{t=1}^{s} y_{jta}}{\sum_{t=1}^{s} y_{jt,p+1}},$$

i.e. the empirical odds in favor of the death cause a relative to the death cause $p+1$. In practice, we define $d(\widehat{\phi}_j, \widehat{\phi}_k) = ||\widehat{\phi}_j - \widehat{\phi}_k||_\infty$, i.e. the supremum norm of the difference between $\widehat{\phi}_j$ and $\widehat{\phi}_k$. The plot on the left hand side of Figure 1 shows the vector of empirical odds for the 367 towns.

We estimated β by the following methods:

(I) Fused group lasso estimation (FGL): Here β was estimated by (A.2). To define the comparison graph $\mathcal{H}$, we set b_{thr} equal to the median of the $367(367-1)/2 = 67161$ pairs of $d(\widehat{\phi}_j, \widehat{\phi}_k)$. The definition led to $b_{\text{thr}} = 0.012$ with the number of edges in $\mathcal{H}$ equal to 33581. The plot on the right hand side of Figure 1 shows the histogram of the 67161 values of $d(\widehat{\phi}_j, \widehat{\phi}_k)$. The red dash line in the plot indicates the median of the 67161 values of $d(\widehat{\phi}_j, \widehat{\phi}_k)$.

We carried out the FGL estimation under 20 tuning parameters values and selected the one with the smallest value in Bayesian Information Criterion (BIC) to construct the estimated similarity graph (defined in 3.7). Here nodes of the estimated similarity graph represent the towns in our data. We carried out community detection on the estimated similarity graph to find a partition of the towns. We called the resulting partition the FGL partition.

(II) MLE with k-means clustering (kMLE): We estimated $\beta_1, \beta_2, \cdots, \beta_m$ separately by the maximum likelihood estimation using data only from each corresponding town. We then used R package "kmeans" to carry out k-means clustering on the maximum likelihood estimates. We considered the number of clusters in the k-means clustering from $k = 2$ to $k = 366$. This led to 365 partitions with different numbers of clusters. We evaluated BIC of the cluster centers for the 365 partitions. We reported the result from the partition with the smallest BIC value. We called the resulting partition the kMLE partition.

To evaluate quality of a partition we used the Silhouette coefficient (Rousseeuw, 1987), which is defined by

$$\text{Silhouette coefficient} \quad = \quad \left(\frac{1}{|\{j : |\mathcal{C}_j| \geq 2\}|} \sum_{j:|\mathcal{C}_j|\geq 2} \frac{v_{2j} - v_{1i}}{\max\{v_{1j}, v_{2j}\}} \right),$$

where $v_{1j} = (|\mathcal{C}_j| - 1)^{-1} \sum_{k\in\mathcal{C}_j} (\widehat{\phi}_{ja} - \widehat{\phi}_{ka})^2$, $v_{2j} = (m - |\mathcal{C}_j|)^{-1} \sum_{k\notin\mathcal{C}_j} (\widehat{\phi}_{ja} - \widehat{\phi}_{ka})^2$, $\widehat{\phi}_{ja}$ is the empirical odds of death cause a for town j, $\mathcal{C}_j$ is the cluster that town j belongs to, and $m = 367$ is the number of towns in our data.

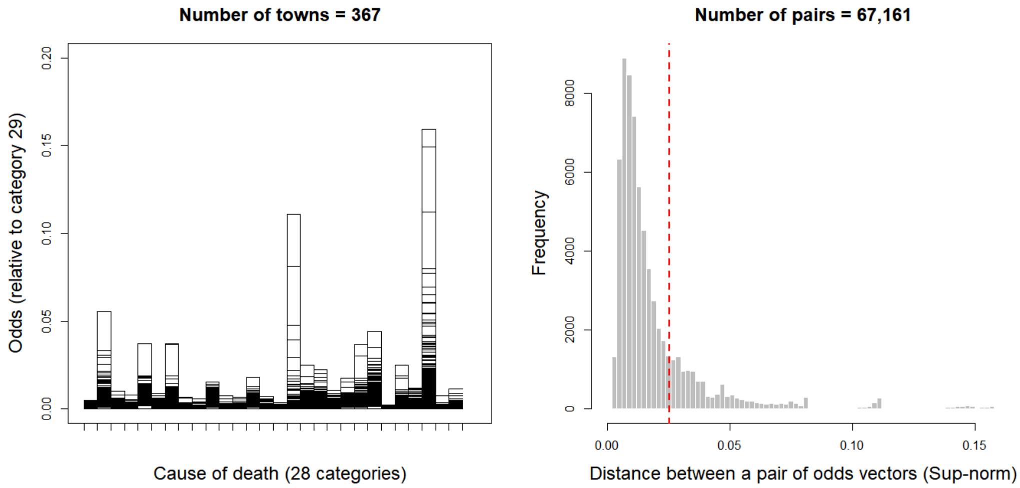

Figure 1: Left: Empirical odds of the 28 causes of death relative to cause of death 29; Right: Distribution of the supremum norm of the pairwise difference.

A.2 Results

Plots in Figure 2 show results from our estimation. The red dash lines indicate the tuning parameter value and the number of clusters corresponding to the smallest BIC value, for the FGL partition and the kMLE partition, respectively. Plots on the bottom show the distributions of the estimated cluster sizes. From the plots we can see the FGL partition has clusters with relatively smaller sizes in comparison with the kMLE partition. Table 1 suggests that the FGL partition may provide a better partition result since its Silhouette coefficient value is 0.746 in comparison with the kMLE partition, which has a Silhouette coefficient value equal to 0.295. In addition, the FGL partition has a relatively larger number of clusters (33 clusters) in comparison with the kMLE partition, which has 12 clusters in total.

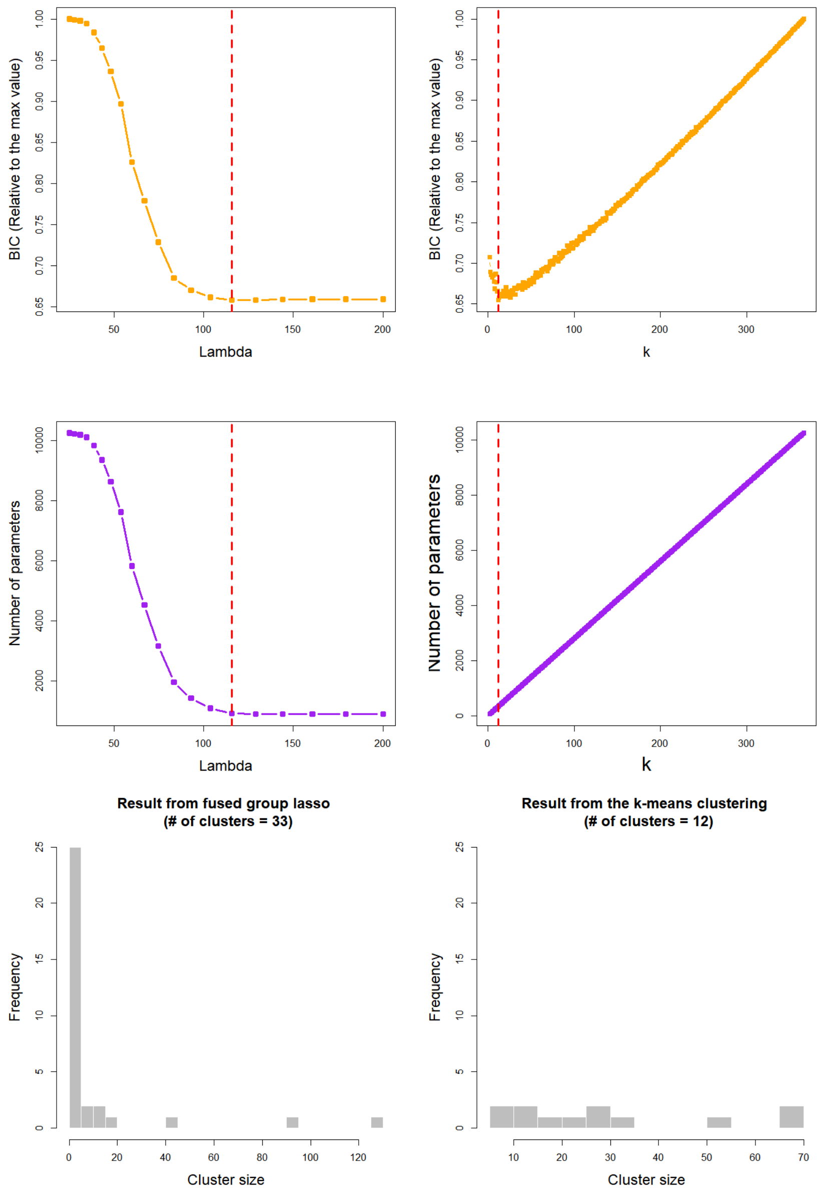

Figure 2: Results for the fused group lasso estimation and the MLE with k-means clustering. Left: The fused group lasso estimation; Right: The MLE with k-means clustering; Top: Trace plots of relative BIC against the tuning parameter λ (left); against the number of clusters (right); Middle: Scatter plots of the number of estimated parameters against the tuning parameter λ (left); against the number of clusters (right); Bottom: Histograms of the estimated cluster sizes. The plots show that the fused group lasso (FGL) estimation produces a relatively larger number of clusters when comparison with the clusters produced by the k-mean MLE (kMLE) estimation.

Method	BIC	# of clusters	Silhouette coefficient
Maximum likelihood estimation (MLE)	165,264.5	367	0
Fused group lasso (FGL)	110,496.9	33	0.746
MLE with k-means clustering (kMLE)	108,079.2	12	0.295

Table 1: Performance results for partitions based on the three estimations.

B Convergence analysis

In this section we conduct convergence analysis on the sequence generated by the iterative scheme (2.11). In particular, we focus on analyzing convergence of the objective function in (2.1) when being evaluated at the ergodic average of the sequence. In addition, we will also analyze convergence of the sequence itself by investigating whether the sequence can satisfy the linear constraints in the primal problem (2.1) and the dual problem (See Section B.1). We first describe the dual of the primal problem (2.1). This dual will serve as a conceptual tool for developing assumptions for our convergence analysis. We then introduce notation definitions and two assumptions for the convergence analysis before presenting the main results for convergence of the objective function and convergence of the sequence.

B.1 The dual problem

The dual of the primal problem (2.1) is

$$\begin{aligned}
\max_{\tau,\xi} \overline{L}(\tau,\xi) \quad &= \quad \max_{\tau,\xi} \left\{ -\sum_{i=1}^{m} l_i^* \left(-\rho \sum_{j \in \mathcal{N}(i)} \xi_{ij} \right) - \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}^* \left(-\frac{\rho \tau_{ij}}{\lambda_{ij}} \right) \right\} \\
\text{subject to} \quad & \quad \tau_{ij} + \xi_{ij} = 0 \text{ for } (i,j) \in \mathcal{H} \\
& \quad \tau_{ij} - \xi_{ji} = 0 \text{ for } (i,j) \in \mathcal{H}.
\end{aligned} \tag{B.1}$$

where $l_i^*(z) = \max_{\beta_i}\{\langle z, \beta_i\rangle - l_i(\beta_i)\}$ and $g_{ij}^*(z) = \max_{\alpha_{ij}}\{\langle z, \alpha_{ij}\rangle - g_{ij}(\alpha_{ij})\}$ are called the conjugate function of $l_i(\beta_i)$ and $g_{ij}(\alpha_{ij})$, respectively. The objective function $\overline{L}(\tau,\xi)$ in (B.1) is called the Lagrangian dual. Its derivation is done by evaluating the Lagrangian (2.2) with the primal minimizer provided that the dual variables are fixed. Details of the derivation are given in Appendix D in

Supplementary Materials. Note that (B.1) is itself an optimization problem with respect to the dual variables $\{\tau_{ij}\}_{(i,j)\in\mathcal{H}}$ and $\{\xi_{ij},\xi_{ji}\}_{(i,j)\in\mathcal{H}}$, and therefore here we will simply call (B.1) the dual problem in subsequent analysis. The dual problem (B.1) is a constrained maximization problem with a set of constraints $\tau_{ij}+\xi_{ij}=0$ and $\tau_{ij}-\xi_{ji}=0$ for $(i,j)\in\mathcal{H}$. These constraints serve as a guide for establishing the stopping criterion associated with the dual error (2.14) as they give us information on how the dual variables should behave in order to achieve the optimal value of the primal problem.

B.2 Assumptions

Before stating the assumptions for our convergence analysis, we give some notation definitions first. Let $\iota\{\mathcal{A}\}$ be an indicator function such that $\iota\{\mathcal{A}\}=0$ if $\mathcal{A}$ is true and $\iota\{\mathcal{A}\}=\infty$ otherwise. With the definition of $\iota\{\mathcal{A}\}$, we can express the objective function in the primal problem (2.1) as

$$\begin{aligned}\Psi(\beta,\alpha,\theta) &= \sum_{j=1}^{m} l_j(\beta_j)+\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g_{ij}(\alpha_{ij})+\sum_{(i,j)\in\mathcal{H}}\iota\{\alpha_{ij}=\theta_{ij}-\theta_{ji}\}\\ &\quad+\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\iota\{\beta_i=\theta_{ij}\},\end{aligned}\tag{B.2}$$

Similarly, we can express the objective function in the dual problem (B.1) as

$$\begin{aligned}\overline{\Psi}(\tau,\xi) &= \left\{-\sum_{i=1}^{m} l_i^*\Big(-\rho\sum_{j\in\mathcal{N}(i)}\xi_{ij}\Big)-\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g_{ij}^*\Big(-\frac{\rho\tau_{ij}}{\lambda_{ij}}\Big)\right\}\\ &\quad-\sum_{(i,j)\in\mathcal{H}}\iota\{\tau_{ij}=-\xi_{ij}\}-\sum_{(i,j)\in\mathcal{H}}\iota\{\tau_{ij}=\xi_{ji}\}.\end{aligned}\tag{B.3}$$

With definitions of (B.2) and (B.3), we considered the following two assumptions when conducting the convergence analysis:

> **Assumption 1 (convexity assumption):** Both $l_i(\beta_i)$ and $g_{ij}(\alpha_{ij})$ are convex and lower-semicontinuous functions.
>
> **Assumption 2 (saddle point assumption):** There exists a point $(\beta^*,\alpha^*,\theta^*,\tau^*,\xi^*)$ such that
>
> $$L(\beta^*,\alpha^*,\theta^*,\tau,\xi)\le L(\beta^*,\alpha^*,\theta^*,\tau^*,\xi^*)\le L(\beta,\alpha,\theta,\tau^*,\xi^*)$$

for all $(\beta, \alpha, \theta, \tau, \xi)$, where $L(\beta, \alpha, \theta, \tau, \xi)$ is the Lagrangian defined in (2.2).

The point $(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*)$ in Assumption 2 is called a saddle point of $L(\beta, \alpha, \theta, \tau, \xi)$. Assumption 1 will allow us to adopt nice properties of convex functions when conducting the convergence analysis on the sequence generated by the BS-based iterative scheme (2.11). However, as a result of that, the corresponding convergence results will only be applicable to situations when (2.11) is applied to solve convex optimization problems. Assumptions 1 and 2 imply that there exists a point that will solve the primal problem and the dual problem simultaneously. In addition, both the primal objective (B.2) and the dual objective (B.3) will have the same value when being evaluated at that point. To state the implication mathematically, we first define

$$\Psi^{\text{opt}}_{\text{primal}} = \min_{\beta,\alpha,\theta} \Psi(\beta, \alpha, \theta), \tag{B.4}$$

$$\Psi^{\text{opt}}_{\text{dual}} = \max_{\tau,\xi} \overline{\Psi}(\tau, \xi). \tag{B.5}$$

Under Assumptions 1 and 2, the first three elements of the saddle point $(\beta^*, \alpha^*, \theta^*)$ will minimize (2.1) while the last two elements (τ^*, ξ^*) will maximize (B.1). In addition, the primal-dual gap will be zero at $(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*)$, i.e.

$$\Psi^{\text{opt}}_{\text{primal}} = \Psi(\beta^*, \alpha^*, \theta^*) = \overline{\Psi}(\tau^*, \xi^*) = \Psi^{\text{opt}}_{\text{dual}},$$

Assumptions 1 and 2 had been previously adopted by Shefi and Teboulle (2014) in proving rates of ergodic convergence for ADMM-type algorithms in solving convex optimization problems.

In addition, since (τ^*, ξ^*) solves the dual problem (B.1), the values $\tau^* = \{\tau^*_{ij}\}_{(i,j)\in\mathcal{H}}$ and $\xi^* = \{\xi^*_{ij}, \xi^*_{ji}\}_{(i,j)\in\mathcal{H}}$ must be bounded (Otherwise we can assign arbitrary values for the dual variables to make (B.3) as large as possible). As a result of that, there should exist sequences $\{v_{ij}\}_{(i,j)\in\mathcal{H}}$, $\{w_{ij}\}_{(i,j)\in\mathcal{H}}$ and $\{w_{ji}\}_{(i,j)\in\mathcal{H}}$ such that $||\tau^*_{ij}||_2 \leq v_{ij} < \infty$, $||\xi^*_{ij}||_2 \leq w_{ij} < \infty$ and $||\xi^*_{ji}||_2 \leq w_{ji} < \infty$ for all $(i,j) \in \mathcal{H}$. For practical purposes, we define

$$\mathcal{D} = \left\{(\tau, \xi) : ||\tau_{ij}||_2 \leq v_{ij}, ||\xi_{ij}||_2 \leq w_{ij}, ||\xi_{ji}||_2 \leq w_{ji}\right\}. \tag{B.6}$$

B.3 Theoretical results

Below we provide the main results from our convergence analysis on the sequence generated by the iterative scheme (2.11). The first result states that under Assumptions 1 and 2, the objective function in (2.1) evaluated at the ergodic average

$$(\beta_i^{\mathrm{erg},r}, \alpha_i^{\mathrm{erg},r}) = \left(\frac{1}{r}\sum_{s=1}^{r}\beta_i^s, \frac{1}{r}\sum_{s=1}^{r}\alpha_i^s\right), \tag{B.7}$$

where (β_i^r, α_i^r) is generated by (2.11), will approach to the optimal value defined in (B.4) with a rate proportional to r^{-1}, where r is the iteration counter for running (2.11).

Theorem B.1. *Under Assumptions 1 and 2, we have*

$$\left\{\sum_{i=1}^{m} l_i(\beta_i^{\mathrm{erg},r}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{\mathrm{erg},r})\right\} - \Psi_{\mathrm{primal}}^{\mathrm{opt}} \leq \frac{\rho F_0^{**}}{r}, \tag{B.8}$$

where $\Psi_{\mathrm{primal}}^{\mathrm{opt}}$ *is defined in (B.4) and*

$$\begin{aligned} F_0^{**} &= \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\theta_{ij}^0 - \beta_i^*||_2^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^0 - \theta_{ji}^0) - \alpha_{ij}^*||_2^2 \\ &\quad + \sum_{(i,j)\in\mathcal{H}} ||\theta_{ji}^0 - \theta_{ij}^*||_2^2 + \sum_{(i,j)\in\mathcal{H}} ||\theta_{ji}^0 - \theta_{ji}^*||_2^2 \\ &\quad + \frac{1}{2}\max_{\tau,\xi\in\mathcal{D}}\left(\sum_{(i,j)\in\mathcal{H}} ||\tau_{ij}^0 - \tau_{ij}||_2^2 + \sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\xi_{ij}^0 - \xi_{ij}||_2^2\right), \end{aligned} \tag{B.9}$$

and $\mathcal{D}$ *is defined in (B.6).*

The second result states that the sequence generated by (2.11) will gradually satisfy the primal constraints in (2.1) and the dual constraints in (B.1) as the iteration counter r increases.

Theorem B.2. *Under Assumptions 1 and 2, we have*

$$\Delta_1^r = \sqrt{\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\beta_i^r - \theta_{ij}^r||_2^2} \leq \sqrt{\frac{2F_0^*}{r}}, \tag{B.10}$$

$$\Delta_2^r = \sqrt{\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^r - (\theta_{ij}^r - \theta_{ji}^r)||_2^2} \leq \sqrt{\frac{2F_0^*}{r}}, \tag{B.11}$$

$$\Delta_3^r = \sqrt{\sum_{(i,j)\in\mathcal{H}} ||\xi_{ij}^r + \tau_{ij}^r||_2^2} \leq \sqrt{\frac{4F_0^*}{r}}, \tag{B.12}$$

$$\Delta_4^r = \sqrt{\sum_{(i,j)\in\mathcal{H}} ||\xi_{ji}^r - \tau_{ij}^r||_2^2} \leq \sqrt{\frac{4F_0^*}{r}}, \tag{B.13}$$

where

$$
\begin{aligned}
F_0^* \quad = \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\theta_{ij}^0-\beta_i^*||_2^2+\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||(\theta_{ij}^0-\theta_{ji}^0)-\alpha_{ij}^*||_2^2 \\
& +\sum_{(i,j)\in\mathcal{H}}||\theta_{ji}^0-\theta_{ij}^*||_2^2+\sum_{(i,j)\in\mathcal{H}}||\theta_{ji}^0-\theta_{ji}^*||_2^2 \\
& +\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\tau_{ij}^0-\tau_{ij}^*||_2^2+\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\xi_{ij}^0-\xi_{ij}^*||_2^2. \qquad \text{(B.14)}
\end{aligned}
$$

In Section 3 we provide simulation experiments to examine the results stated in Theorem B.1 and Theorem B.2.

C Derivation of the stopping criteria (2.13) and (2.14)

Below we derive the stopping criteria (2.13) and (2.14). First note that since β_i^{r+1} minimizes the objective function in (2.4), it should satisfy the following equation:

$$
0=\nabla_i l_i(\beta_i^{r+1})+\rho|\mathcal{N}(i)|\beta_i^{r+1}-\rho\sum_{j\in\mathcal{N}(i)}(\theta_{ij}^r-\xi_{ij}^r). \qquad \text{(C.1)}
$$

On the other hand, from (2.9) we have $\xi_{ij}^r=\xi_{ij}^{r+1}-(\beta_i^{r+1}-\theta_i^{r+1})$. Plugging in this relation into (C.1) yields

$$
0 \quad = \quad \nabla_i l_i(\beta_i^{r+1})+\rho\sum_{i\in\mathcal{N}(i)}(\theta_{ij}^{r+1}-\theta_{ij}^r+\xi_{ij}^{r+1}),
$$

which implies β_i^{r+1} is also a solution to the following optimization problem:

$$
\text{minimize}_{\beta_i} \qquad l_i(\beta_i)+\rho\sum_{i\in\mathcal{N}(i)}\left\langle\theta_{ij}^{r+1}-\theta_{ij}^r+\xi_{ij}^{r+1},\beta_i\right\rangle. \qquad \text{(C.2)}
$$

Similarly since α_{ij}^{r+1} minimizes the objective function in (2.5), it should satisfy the following equation:

$$
0\in\lambda_{ij}\partial g_{ij}(\alpha_{ij}^{r+1})+\rho[\alpha_{ij}^{r+1}-(\theta_{ij}^r-\theta_{ji}^r)+\tau_{ij}^r]. \qquad \text{(C.3)}
$$

From (2.8) we have $\tau_{ij}^r=\tau_{ij}^{r+1}-[\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})]$. Plugging in this relation into (C.3) yields

$$
0 \quad \in \quad \lambda_{ij}\partial g_{ij}(\alpha_{ij}^{r+1})+\rho[(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})-(\theta_{ij}^r-\theta_{ji}^r)+\tau_{ij}^{r+1}],
$$

which implies α_{ij}^{r+1} is also a solution to the following optimization problem:

$$\text{minimize}_{\alpha_{ij}} \qquad \lambda_{ij} g_{ij}(\alpha_{ij}) + \rho \left\langle (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^{r} - \theta_{ji}^{r}) + \tau_{ij}^{r+1}, \alpha_{ij} \right\rangle . \tag{C.4}$$

Now assume $(\beta^*, \alpha^*, \theta^*)$ solves the primal problem (2.1). Then from (C.2) and (C.4) we should have

$$\begin{aligned}
& \sum_{i=1}^{m} l_i(\beta_i^{r+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{r+1}) \\
\leq \; & \sum_{i=1}^{m} l_i(\beta_i^{*}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{*}) \\
& -\rho \sum_{i=1}^{m} \sum_{i\in\mathcal{N}(i)} \left\langle \theta_{ij}^{r+1} - \theta_{ij}^{r}, \beta_i^{r+1} - \beta_i^* \right\rangle - \rho \sum_{i=1}^{m} \sum_{i\in\mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle & \text{(C.5)} \\
& -\rho \sum_{(i,j)\in\mathcal{H}} \Bigg(\left\langle (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^{r} - \theta_{ji}^{r}), \alpha_{ij}^{r+1} - \alpha_{ij}^* \right\rangle \\
& - \left\langle \tau_{ij}^{r+1}, \alpha_{ij}^{r+1} - \alpha_{ij}^* \right\rangle \Bigg). & \text{(C.6)}
\end{aligned}$$

We first consider the last term of (C.5). Note that

$$\begin{aligned}
\sum_{i=1}^{m} \sum_{i\in\mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle \; = \; & \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1}, \beta_j^{r+1} - \beta_j^* \right\rangle \\
= \; & \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ij}^{r+1} + \tau_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1} - \tau_{ij}^{r+1}, \beta_j^{r+1} - \beta_j^* \right\rangle \\
& - \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \beta_j^{r+1} - \beta_j^* \right\rangle \\
= \; & \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ij}^{r+1} + \tau_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1} - \tau_{ij}^{r+1}, \beta_j^{r+1} - \beta_j^* \right\rangle \\
& + \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, (\beta_j^{r+1} - \beta_i^{r+1}) - (\beta_j^* - \beta_i^*) \right\rangle . \qquad \text{(C.7)}
\end{aligned}$$

Now adding the last term of (C.7) to the last term of (C.6) yields

$$\begin{aligned}
& \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, (\beta_j^{r+1} - \beta_i^{r+1}) - (\beta_j^* - \beta_i^*) \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \alpha_{ij}^{r+1} - \alpha_{ij}^* \right\rangle \\
= \; & \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) + (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\beta_i^{r+1} - \beta_j^{r+1}) - (\alpha_{ij}^* - (\beta_i^* - \beta_j^*)) \right\rangle .
\end{aligned} \tag{C.8}$$

Since $(\beta^*, \alpha^*, \theta^*)$ solves the primal problem (2.1), we should have $\alpha^*_{ij} - (\beta^*_i - \beta^*_j) = \alpha^*_{ij} - (\theta^*_{ij} - \theta^*_{ji}) = 0$. Therefore adding the last term of (C.5) to the last term of (C.6) and using the result (C.8) yields

$$\begin{aligned}
& \sum_{i=1}^{m} \sum_{i \in \mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \alpha_{ij}^{r+1} - \alpha_{ij}^* \right\rangle \\
= & \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ij}^{r+1} + \tau_{ij}^{r+1}, \beta_i^{r+1} - \beta_i^* \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1} - \tau_{ij}^{r+1}, \beta_j^{r+1} - \beta_j^* \right\rangle \\
& + \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, [\alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1})] + (\theta_{ij}^{r+1} - \beta_i^{r+1}) + (\beta_j^{r+1} - \theta_{ji}^{r+1}) \right\rangle .
\end{aligned} \tag{C.9}$$

The result (C.9) implies that

$$\left\{ \sum_{i=1}^{m} l_i(\beta_i^{r+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{r+1}) \right\} - \left\{ \sum_{i=1}^{m} l_i(\beta_i^*) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^*) \right\} \le \rho(G_1 + G_2 + G_3 + G_4),$$

where

$$\begin{aligned}
G_1 &= \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} ||\theta_{ij}^{r+1} - \theta_{ij}^r||_2 \Big(||\beta_i^{r+1} - \beta_i^*||_2 + ||\alpha_{ij}^{r+1} - \alpha_{ij}^*||_2 \Big), \\
G_2 &= \sum_{(i,j)\in\mathcal{H}} ||\xi_{ij}^{r+1} + \tau_{ij}^{r+1}||_2 ||\beta_i^{r+1} - \beta_i^*||_2, \\
G_3 &= \sum_{(i,j)\in\mathcal{H}} ||\xi_{ji}^{r+1} - \tau_{ij}^{r+1}||_2 ||\beta_j^{r+1} - \beta_j^*||_2, \\
G_4 &= \sum_{(i,j)\in\mathcal{H}} ||\tau_{ij}^{r+1}||_2 \Big(||\alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1})||_2 + ||\beta_i^{r+1} - \theta_{ij}^{r+1}||_2 + ||\beta_j^{r+1} - \theta_{ji}^{r+1}||_2 \Big).
\end{aligned}$$

Here G_1 and G_4 have terms involving sequence differences associated with the primal variables $\{\beta_i\}_{i=1}^m$ and $\{\alpha_{ij}\}_{(i,j)\in\mathcal{H}}$ and auxiliary variables $\{\theta_{ij}, \theta_{ji}\}_{(i,j)\in\mathcal{H}}$ while G_2 and G_3 have terms involving sequence differences associated with the dual variables $\{\tau_{ij}\}_{(i,j)\in\mathcal{H}}$ and $\{\xi_{ij}, \xi_{ji}\}_{(i,j)\in\mathcal{H}}$. These results imply we can use

$$\begin{aligned}
& \sum_{(i,j)\in\mathcal{H}} \Big(||\theta_{ij}^{r+1} - \theta_{ij}^r||_2 + ||\theta_{ji}^{r+1} - \theta_{ji}^r||_2 \\
& + ||\beta_i^{r+1} - \theta_{ij}^{r+1}||_2 + ||\beta_j^{r+1} - \theta_{ji}^{r+1}||_2 + ||\alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1})||_2 \Big) \\
\le\ & 5 q_{\mathcal{H}} \sqrt{p}\, \epsilon_{\text{primal}}
\end{aligned}$$

and

$$\sum_{(i,j)\in\mathcal{H}} \Big(||\xi_{ij}^{r+1} + \tau_{ij}^{r+1}||_2 + ||\xi_{ji}^{r+1} - \tau_{ij}^{r+1}||_2 \Big) \le 2 q_{\mathcal{H}} \sqrt{p}\, \epsilon_{\text{dual}}$$

as stopping criteria for the iterative scheme (2.11).

D Derivation of the objective function in (B.1)

The Lagrangian "dual" of the Lagrangian (2.2) is defined as

$$\overline{L}(\tau,\xi) \quad = \quad \min_{(\beta,\alpha,\theta)} L(\beta,\alpha,\theta,\tau,\xi). \tag{D.1}$$

It is a function of the dual variables (τ,ξ). By collecting terms involving β,α and θ, respectively, we can express the Lagrangian dual (D.1) as

$$\begin{aligned}\overline{L}(\tau,\xi) \quad &= \quad \min_{\beta}\sum_{i=1}^{m}\Big[l_i(\beta_i)+\sum_{j\in\mathcal{N}(i)}\rho\left\langle\xi_{ij},\beta_i\right\rangle\Big]+\min_{\alpha}\sum_{(i,j)\in\mathcal{H}}\Big[\lambda_{ij}g_{ij}(\alpha_{ij})+\rho\left\langle\tau_{ij},\alpha_{ij}\right\rangle\Big]\\ &\qquad +\rho\cdot\min_{\theta}\Big[-\sum_{(i,j)\in\mathcal{H}}\langle\tau_{ij},\theta_{ij}-\theta_{ji}\rangle-\sum_{j=1}^{m}\sum_{j\in\mathcal{N}(i)}\langle\xi_{ij},\theta_{ij}\rangle\Big].\end{aligned} \tag{D.2}$$

It is a linear combination of three terms involving objective functions evaluated at their minimizers. We can further express the first term of (D.2) as

$$\begin{aligned}\min_{\beta}\sum_{i=1}^{m}\Big[l_i(\beta_i)+\sum_{j\in\mathcal{N}(i)}\rho\left\langle\xi_{ij},\beta_i\right\rangle\Big] \quad &= \quad \sum_{i=1}^{m}\min_{\beta_i}\Big[l_i(\beta_i)+\Big\langle\beta_i,\rho\sum_{j\in\mathcal{N}(i)}\xi_{ij}\Big\rangle\Big]\\ &= \quad \sum_{i=1}^{m}-\Big\{\max_{\beta_i}\Big[\Big\langle\beta_i,-\rho\sum_{j\in\mathcal{N}(i)}\xi_{ij}\Big\rangle-l_i(\beta_i)\Big]\Big\}\\ &= \quad -\sum_{i=1}^{m}l_i^{*}\Big(-\rho\sum_{j\in\mathcal{N}(i)}\xi_{ij}\Big).\end{aligned} \tag{D.3}$$

Similarly, We can express the second term of (D.2) as

$$\begin{aligned}\min_{\alpha}\sum_{(i,j)\in\mathcal{H}}\Big[\lambda_{ij}g_{ij}(\alpha_{ij})+\rho\left\langle\tau_{ij},\alpha_{ij}\right\rangle\Big] \quad &= \quad \sum_{(i,j)\in\mathcal{H}}\min_{\alpha_{ij}}\Big[\lambda_{ij}g_{ij}(\alpha_{ij})+\rho\left\langle\tau_{ij},\alpha_{ij}\right\rangle\Big]\\ &= \quad \sum_{(i,j)\in\mathcal{H}}-\lambda_{ij}\Big\{\max_{\alpha_{ij}}\Big[\Big\langle\alpha_{ij},-\frac{\rho\tau_{ij}}{\lambda_{ij}}\Big\rangle-g_{ij}(\alpha_{ij})\Big]\Big\}\\ &= \quad -\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g^{*}\Big(-\frac{\rho\tau_{ij}}{\lambda_{ij}}\Big).\end{aligned}$$

In addition, we can express the third term of (D.2) as

$$\begin{aligned}&\rho\cdot\min_{\theta}\Big[-\sum_{(i,j)\in\mathcal{H}}\langle\tau_{ij},\theta_{ij}-\theta_{ji}\rangle-\sum_{j=1}^{m}\sum_{j\in\mathcal{N}(i)}\langle\xi_{ij},\theta_{ij}\rangle\Big]\\ = \quad &\rho\cdot\min_{\theta}\Big[-\sum_{(i,j)\in\mathcal{H}}\langle\tau_{ij},\theta_{ij}\rangle+\sum_{(i,j)\in\mathcal{H}}\langle\tau_{ij},\theta_{ji}\rangle-\sum_{(i,j)\in\mathcal{H}}\langle\xi_{ij},\theta_{ij}\rangle-\sum_{(i,j)\in\mathcal{H}}\langle\xi_{ji},\theta_{ji}\rangle\Big]\\ = \quad &\rho\sum_{(i,j)\in\mathcal{H}}\min_{\theta_{ij}}\left\langle\theta_{ij},-(\tau_{ij}+\xi_{ij})\right\rangle+\rho\sum_{(i,j)\in\mathcal{H}}\min_{\theta_{ji}}\left\langle\theta_{ji},\tau_{ij}-\xi_{ji}\right\rangle.\end{aligned} \tag{D.4}$$

The right hand side of (D.4) implies that

$$\min_{\theta}\Big[-\sum_{(i,j)\in\mathcal{H}}\langle\tau_{ij},\theta_{ij}-\theta_{ji}\rangle-\sum_{j=1}^{m}\sum_{j\in\mathcal{N}(i)}\langle\xi_{ij},\theta_{ij}\rangle\Big]=\begin{cases}0 \text{ if } \tau_{ij}+\xi_{ij}=0 \text{ and } \tau_{ij}-\xi_{ji}=0\\ \\ -\infty \text{ otherwise}\end{cases}, \quad \text{(D.5)}$$

which means we can always find some θ that makes (D.5) equal to $-\infty$ given that $\tau_{ij}+\xi_{ij}\neq 0$ or $\tau_{ij}-\xi_{ji}\neq 0$. However, such a choice will make the Lagrangian dual $\overline{L}(\tau,\xi)=-\infty$, which further makes the dual problem (B.1) meaningless. Therefore to construct a sensible dual problem we prefer the constraints $\tau_{ij}+\xi_{ij}=0$ and $\tau_{ij}-\xi_{ji}=0$ hold for $(i,j)\in\mathcal{H}$. Combining the results in (D.3), (D.4) and (D.5) we obtain the Lagrangian dual

$$\overline{L}(\tau,\xi)=-\sum_{i=1}^{m}l_i^*\Big(-\rho\sum_{j\in\mathcal{N}(i)}\xi_{ij}\Big)-\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g^*\Big(-\frac{\rho\tau_{ij}}{\lambda_{ij}}\Big)$$

with constraints $\tau_{ij}+\xi_{ij}=0$ and $\tau_{ij}-\xi_{ji}=0$ for $(i,j)\in\mathcal{H}$.

The constraints $\tau_{ij}+\xi_{ij}=0$ and $\tau_{ij}-\xi_{ji}=0$ provide information about the relationships between dual variables τ and ξ, and they also serve as a guideline for monitoring convergence of the sequence generated by the iterative scheme (2.11).

E Proofs of Theorem B.1 and Theorem B.2

This section contains proofs of the main results in Section B.3. Section E.1 provide notation definitions useful for proving the main results. Section E.2 contains the proof of Theorem B.1 and technical results relating to the proof. Section E.3 contains the proof of Theorem B.2 and technical results relating to the poof. Section E.4 contains other technical lemmas underlying the theoretical results given in Sections E.2 and E.3.

E.1 Notation

Below we introduce new notation and definitions of quantities that will be useful in proving theoretical results in our paper. We define

$$
\begin{aligned}
A(\beta)^{r+1} &= \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} \left\langle \theta_{ij}^{r+1} - \theta_{ij}^{r}, \beta_i - \beta_i^{r+1} \right\rangle, && \text{(E.1)}\\
B(\alpha)^{r+1} &= \sum_{(i,j) \in \mathcal{H}} \left\langle (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^{r} - \theta_{ji}^{r}), \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle, && \text{(E.2)}\\
C(\theta)^{r+1} &= \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \theta_{ij} - \theta_{ij}^{r+1} \right\rangle, && \text{(E.3)}\\
D(\theta)^{r+1} &= \sum_{(i,j) \in \mathcal{H}} \left\langle \tau_{ij}^{r+1}, (\theta_{ij} - \theta_{ji}) - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) \right\rangle. && \text{(E.4)}
\end{aligned}
$$

We further define

$$
\Omega(a, b)^{r+1} = \frac{1}{2}\Bigg(||a - b^r||_2^2 - ||a - b^{r+1}||_2^2 \Bigg). \tag{E.5}
$$

E.2 Proof of Theorem B.1

Below we provide a proof of Theorem B.1. We first give some results that are key for proving the theorem. These results are summarized in Lemma E.1 and Lemma E.2.

Lemma E.1. *Under Assumption 1 we have*

$$
\begin{aligned}
L(\beta^{r+1}, \alpha^{r+1}, \theta^{r+1}, \tau, \xi) \quad \leq \quad & L(\beta, \alpha, \theta, \tau^{r+1}, \xi^{r+1}) \\
& + \rho \Bigg[A(\beta)^{r+1} + B(\alpha)^{r+1} + C(\theta)^{r+1} + D(\theta)^{r+1} \\
& + \sum_{(i,j) \in \mathcal{H}} \Omega(\tau_{ij}, \tau_{ij})^{r+1} - \frac{1}{2} \sum_{(i,j \in \mathcal{H})} ||\tau_{ij}^{r+1} - \tau_{ij}^{r}||_2^2 \\
& + \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} \Omega(\xi_{ij}, \xi_{ij})^{r+1} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} ||\xi_{ij}^{r+1} - \xi_{ij}^{r}||_2^2 \Bigg],
\end{aligned}
\tag{E.6}
$$

where $L(\beta, \alpha, \theta, \tau, \xi)$ is defined in (2.2).

Proof of Lemma E.1. First note that by using the definition of $C(\theta)^{r+1}$ in (E.3) and the definition of $D(\theta)^{r+1}$ in (E.4), we have

$$-\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\left\langle \xi_{ij}^{r+1},\theta_{ij}^{r+1}\right\rangle = -\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\left\langle \xi_{ij}^{r+1},\theta_{ij}\right\rangle + C(\theta)^{r+1}, \tag{E.7}$$

$$-\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}^{r+1},\theta_{ij}^{r+1}-\theta_{ji}^{r+1}\right\rangle = -\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}^{r+1},\theta_{ij}-\theta_{ji}\right\rangle + D(\theta)^{r+1}, \tag{E.8}$$

Next, we have

$$\begin{aligned}\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij},\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})\right\rangle &= \sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}^{r+1},\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})\right\rangle \\ &\quad+\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}-\tau_{ij}^{r+1},\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})\right\rangle.\end{aligned} \tag{E.9}$$

By using the relation (2.8), we have $\tau_{ij}^{r+1}-\tau_{ij}^{r}=\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})$. Further by the derivation using the result (E.71) in Proposition E.1 and the definition (E.5), we can express the second term on right hand side of (E.9) as

$$\begin{aligned}&\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}-\tau_{ij}^{r+1},\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})\right\rangle \\ =&\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}-\tau_{ij}^{r+1},\tau_{ij}^{r+1}-\tau_{ij}^{r}\right\rangle \\ =&\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}\left(||\tau_{ij}-\tau_{ij}^{r}||_2^2-||\tau_{ij}-\tau_{ij}^{r+1}||_2^2-||\tau_{ij}^{r+1}-\tau_{ij}^{r}||_2^2\right) \\ =&\sum_{(i,j)\in\mathcal{H}}\Omega(\tau_{ij},\tau_{ij})^{r+1}-\frac{1}{2}\sum_{(i,j\in\mathcal{H})}||\tau_{ij}^{r+1}-\tau_{ij}^{r}||_2^2,\end{aligned} \tag{E.10}$$

With the result (E.10) we can re-express (E.9) as

$$\begin{aligned}\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij},\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})\right\rangle &= \sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}^{r+1},\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})\right\rangle \\ &\quad+\sum_{(i,j)\in\mathcal{H}}\Omega(\tau_{ij},\tau_{ij})^{r+1}-\frac{1}{2}\sum_{(i,j\in\mathcal{H})}||\tau_{ij}^{r+1}-\tau_{ij}^{r}||_2^2.\end{aligned} \tag{E.11}$$

Following the relations (2.9) and (2.10) and a similar derivation procedure given above, we further obtain

$$\begin{aligned}&\sum_{(i,j)\in\mathcal{H}}\left\langle \xi_{ij},\beta_i^{r+1}-\theta_{ij}^{r+1}\right\rangle \\ =&\sum_{(i,j)\in\mathcal{H}}\left\langle \xi_{ij}^{r+1},\beta_i^{r+1}-\theta_{ij}^{r+1}\right\rangle+\sum_{(i,j)\in\mathcal{H}}\Omega(\xi_{ij},\xi_{ij})^{r+1}-\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\xi_{ij}^{r+1}-\xi_{ij}^{r}||_2^2,\end{aligned} \tag{E.12}$$

and

$$
\begin{aligned}
& \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}, \beta_j^{r+1} - \theta_{ji}^{r+1} \right\rangle \\
= \; & \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1}, \beta_j^{r+1} - \theta_{ji}^{r+1} \right\rangle + \sum_{(i,j)\in\mathcal{H}} \Omega(\xi_{ji}, \xi_{ji})^{r+1} - \frac{1}{2} \sum_{(i,j)\in\mathcal{H}} ||\xi_{ji}^{r+1} - \xi_{ji}^{r}||_2^2.
\end{aligned} \tag{E.13}
$$

Now adding (E.53) in Lemma E.4, (E.59) in Lemma E.5, (E.7), (E.8), (E.11), (E.12) and (E.13), and cancelling out the same terms on both sides of the inequality, we obtain

$$
\begin{aligned}
& \sum_{i=1}^{m} l_i(\beta_i^{r+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{r+1}) + \rho \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}, \alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) \right\rangle \\
& + \rho \sum_{i=1}^{m} \sum_{j\in\mathcal{N}(i)} \left\langle \xi_{ij}, \beta_i^{r+1} - \theta_{ij}^{r+1} \right\rangle \\
\leq \; & \sum_{i=1}^{m} l_i(\beta_i) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}) + \rho \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \alpha_{ij} - (\theta_{ij} - \theta_{ji}) \right\rangle \\
& + \rho \sum_{i=1}^{m} \sum_{j\in\mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \beta_i - \theta_{ij} \right\rangle \\
& + \rho \Bigg[A(\beta)^{r+1} + B(\alpha)^{r+1} + C(\theta)^{r+1} + D(\theta)^{r+1} \\
& + \sum_{(i,j)\in\mathcal{H}} \Omega(\tau_{ij}, \tau_{ij})^{r+1} - \frac{1}{2} \sum_{(i,j\in\mathcal{H})} ||\tau_{ij}^{r+1} - \tau_{ij}^{r}||_2^2 \\
& + \sum_{i=1}^{m} \sum_{j\in\mathcal{N}(i)} \Omega(\xi_{ij}, \xi_{ij})^{r+1} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j\in\mathcal{N}(i)} ||\xi_{ij}^{r+1} - \xi_{ij}^{r}||_2^2 \Bigg].
\end{aligned} \tag{E.14}
$$

By using the definition in (2.2) for the first four lines of (E.14), we recover (E.6), which completes the proof. □

Lemma E.2. *Under Assumption 1 we have*

$$\begin{aligned}
L(\beta^{r+1},\alpha^{r+1},\theta^{r+1},\tau,\xi) \quad \le \quad & L(\beta,\alpha,\theta,\tau^{r+1},\xi^{r+1}) \\
& +\rho\Big[\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\Omega(\beta_i,\theta_{ij})^{r+1}+\sum_{(i,j)\in\mathcal{H}}\Omega(\alpha_{ij},\theta_{ij}-\theta_{ji})^{r+1} \\
& +\sum_{(i,j)\in\mathcal{H}}\Omega(\tau_{ij},\tau_{ij})^{r+1}+\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\Omega(\xi_{ij},\xi_{ij})^{r+1} \\
& +\sum_{(i,j)\in\mathcal{H}}\Omega(\theta_{ij}+\theta_{ji},\theta_{ij}+\theta_{ji})^{r+1} \\
& -\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\beta_i^{r+1}-\theta_{ij}^{r}||_2^2-\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\alpha_{ij}^{r+1}-(\theta_{ij}^{r}-\theta_{ji}^{r})||_2^2 \\
& -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||(\theta_{ij}^{r+1}+\theta_{ji}^{r+1})-(\theta_{ij}^{r}+\theta_{ji}^{r})||_2^2\Big]. \qquad \text{(E.15)}
\end{aligned}$$

Proof of Lemma E.2. The proof starts from the right hand side of (E.6) in Lemma E.1. First by using the result (E.70) in Proposition E.1 with $a=\beta_i$, $b=\beta_i^{r+1}$, $c=\theta_{ij}^{r+1}$ and $d=\theta_{ij}^{r}$, we can express the ith term in $A(\alpha)^{r+1}$ as

$$\big\langle \beta_i-\beta_i^{r+1},\theta_{ij}^{r+1}-\theta_{ij}^{r}\big\rangle=\Omega(\beta_i,\theta_{ij})^{r+1}+\frac{1}{2}||\beta_i^{r+1}-\theta_{ij}^{r+1}||_2^2-\frac{1}{2}||\beta_i^{r+1}-\theta_{ij}^{r}||_2^2, \qquad \text{(E.16)}$$

where $\Omega(\beta_i,\theta_{ij})^{r+1}$ is defined in (E.5). Similarly, with $a=\alpha_{ij}$, $b=\alpha_{ij}^{r+1}$, $c=\theta_{ij}^{r+1}-\theta_{ji}^{r+1}$ and $d=\theta_{ij}^{r}-\theta_{ji}^{r}$, we also express the ijth term in $B(\alpha)^{r+1}$ as

$$\begin{aligned}
& \big\langle (\theta_{ij}^{r+1}-\theta_{ji}^{r+1})-(\theta_{ij}^{r}-\theta_{ji}^{r}),\alpha_{ij}-\alpha_{ij}^{r+1}\big\rangle \\
= \quad & \Omega(\alpha_{ij},\theta_{ij}-\theta_{ji})^{r+1}+\frac{1}{2}||\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})||_2^2-\frac{1}{2}||\alpha_{ij}^{r+1}-(\theta_{ij}^{r}-\theta_{ji}^{r})||_2^2.
\end{aligned} \qquad \text{(E.17)}$$

On the other hand, by using the relations (2.8), (2.9) and (2.10), we have $\tau_{ij}^{r+1}-\tau_{ij}^{r}=\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})$, $\xi_{ij}^{r+1}-\xi_{ij}^{r}=\beta_i^{r+1}-\theta_{ij}^{r+1}$ and $\xi_{ji}^{r+1}-\xi_{ji}^{r}=\beta_j^{r+1}-\theta_{ji}^{r+1}$. These results imply that

$$\begin{aligned}
\frac{1}{2}||\beta_i^{r+1}-\theta_{ij}^{r+1}||_2^2 &= \frac{1}{2}||\xi_{ij}^{r+1}-\xi_{ij}^{r}||_2^2, \\
\frac{1}{2}||\beta_j^{r+1}-\theta_{ji}^{r+1}||_2^2 &= \frac{1}{2}||\xi_{ji}^{r+1}-\xi_{ji}^{r}||_2^2, \\
\frac{1}{2}||\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})||_2^2 &= \frac{1}{2}||\tau_{ij}^{r+1}-\tau_{ij}^{r}||_2^2. \qquad \text{(E.18)}
\end{aligned}$$

By using the results in (E.16), (E.17) and (E.18), we obtain

$$\begin{aligned}
& A(\beta)^{r+1} + B(\alpha)^{r+1} - \frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||\tau_{ij}^{r+1} - \tau_{ij}^{r}||_2^2 - \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\xi_{ij}^{r+1} - \xi_{ij}^{r}||_2^2 \\
= & \sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} \Omega(\beta_i, \theta_{ij})^{r+1} - \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\beta_i^{r+1} - \theta_{ij}^{r}||_2^2 \\
& + \sum_{(i,j)\in\mathcal{H}} \Omega(\alpha_{ij}, \theta_{ij} - \theta_{ji})^{r+1} - \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^{r+1} - (\theta_{ij}^{r} - \theta_{ji}^{r})||_2^2. \qquad \text{(E.19)}
\end{aligned}$$

Next we turn our attention to $C(\theta)^{r+1}$ and $D(\theta)^{r+1}$ in (E.6). First note that we can express $C(\theta)^{r+1}$ and $D(\theta)^{r+1}$ as

$$C(\theta)^{r+1} = \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ij}^{r+1}, \theta_{ij} - \theta_{ij}^{r+1} \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1}, \theta_{ji} - \theta_{ji}^{r+1} \right\rangle, \qquad \text{(E.20)}$$

and

$$D(\theta)^{r+1} = \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \theta_{ij} - \theta_{ij}^{r+1} \right\rangle - \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \theta_{ji} - \theta_{ji}^{r+1} \right\rangle. \qquad \text{(E.21)}$$

By using (E.65) in Proposition E.6 and representations in (E.20) and (E.21), we can obtain

$$\begin{aligned}
& C(\theta)^{r+1} + D(\theta)^{r+1} \\
= & \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ij}^{r+1} + \tau_{ij}^{r+1}, \theta_{ij} - \theta_{ij}^{r+1} \right\rangle + \sum_{(i,j)\in\mathcal{H}} \left\langle \xi_{ji}^{r+1} - \tau_{ij}^{r+1}, \theta_{ji} - \theta_{ji}^{r+1} \right\rangle \\
= & \sum_{(i,j)\in\mathcal{H}} \left\langle (\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^{r} + \theta_{ji}^{r}), (\theta_{ij} + \theta_{ji}) - (\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) \right\rangle \\
= & \sum_{(i,j)\in\mathcal{H}} \Omega(\theta_{ij} + \theta_{ji}, \theta_{ij} + \theta_{ji})^{r+1} - \frac{\rho}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^{r} + \theta_{ji}^{r})||_2^2.
\end{aligned} \qquad \text{(E.22)}$$

where the derivation from the second line to the third line has used the result (E.70) in Lemma E.1 with $a = \theta_{ij} + \theta_{ji}$, $b = \theta_{ij}^{r+1} + \theta_{ji}^{r+1}$, $c = \theta_{ij}^{r+1} + \theta_{ji}^{r+1}$ and $d = \theta_{ij}^{r} + \theta_{ji}^{r}$.

By applying the results in (E.19) and (E.22) into the right hand side of the inequality (E.6), we can recover (E.15), which completes the proof. □

Proof of Theorem B.1. By summing up the inequality (E.15) in Lemma E.2 from $s = 0$ to $s = r - 1$ and using the result in Proposition E.2, we obtain

$$\sum_{s=0}^{r-1} L(\beta^{s+1}, \alpha^{s+1}, \theta^{s+1}, \tau, \xi) \le \sum_{s=0}^{r-1} L(\beta, \alpha, \theta, \tau^{s+1}, \xi^{s+1}) + \rho F(\beta, \alpha, \theta, \tau, \xi), \qquad \text{(E.23)}$$

where

$$
\begin{aligned}
F(\beta,\alpha,\theta,\tau,\xi) &= \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\beta_i-\theta_{ij}^0||_2^2+\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\alpha_{ij}-(\theta_{ij}^0-\theta_{ji}^0)||_2^2 \\
&+\sum_{(i,j)\in\mathcal{H}}||\theta_{ij}-\theta_{ji}^0||_2^2+\sum_{(i,j)\in\mathcal{H}}||\theta_{ji}-\theta_{ji}^0||^2 \\
&+\frac{1}{2}\Bigg(\sum_{(i,j)\in\mathcal{H}}||\tau_{ij}-\tau_{ij}^0||^2+\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\xi_{ij}-\xi_{ij}^0||^2\Bigg).
\end{aligned}
$$

Now let $(\beta,\alpha,\theta)=(\beta^*,\alpha^*,\theta^*)$, where $(\beta^*,\alpha^*,\theta^*)$ is the first three elements of the saddle point $(\beta^*,\alpha^*,\theta^*,\tau^*,\xi^*)$ defined in Assumption 2. We replace (β,α,θ) with $(\beta^*,\alpha^*,\theta^*)$ in the first term on the right hand side of (E.23). Note that by using the fact that $\alpha_{ij}^*=\theta_{ij}^*-\theta_{ji}^*$ and $\beta_i^*=\theta_{ij}^*$, we obtain

$$
\begin{aligned}
L(\beta^*,\alpha^*,\theta^*,\tau^{s+1},\xi^{s+1}) &= \sum_{i=1}^{m}l_i(\beta_i^*)+\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g_{ij}(\alpha_{ij}^*)+\rho\sum_{(i,j)\in\mathcal{H}}\big\langle\tau_{ij}^{s+1},\alpha_{ij}^*-(\theta_{ij}^*-\theta_{ji}^*)\big\rangle \\
&+\rho\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\big\langle\xi_{ij}^{s+1},\beta_i^*-\theta_{ij}^*\big\rangle \\
&= \sum_{i=1}^{m}l_i(\beta_i^*)+\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g_{ij}(\alpha_{ij}^*) \\
&= \Psi_{\text{primal}}^{\text{opt}} \qquad\qquad (\text{E.24})
\end{aligned}
$$

for all $s\in\{0,1,2,\cdots,r-1\}$, where $\Psi_{\text{primal}}^{\text{opt}}$ is defined in (B.4). Therefore for the first term on the right hand side of (E.23), we have

$$
\sum_{s=0}^{r-1}L(\beta^*,\alpha^*,\theta^*,\tau^{s+1},\xi^{s+1})=r\Psi_{\text{primal}}^{\text{opt}}. \qquad (\text{E.25})
$$

Further note that by maximizing both sides of (E.24) with respect to $(\tau,\xi)\in\mathcal{D}$ (defined in (B.6)) and using result (E.25), we obtain

$$
\begin{aligned}
\max_{(\tau,\xi)\in\mathcal{D}}\Bigg\{\sum_{s=0}^{r-1}L(\beta^{s+1},\alpha^{s+1},\theta^{s+1},\tau,\xi)\Bigg\} &\leq \sum_{s=0}^{r-1}L(\beta^*,\alpha^*,\theta^*,\tau^{s+1},\xi^{s+1})+\rho\cdot\max_{(\tau,\xi)\in\mathcal{D}}\{F(\beta^*,\alpha^*,\theta^*,\tau,\xi)\} \\
&= r\Psi_{\text{primal}}^{\text{opt}}+\rho F_0^{**}, \qquad (\text{E.26})
\end{aligned}
$$

where F_0^{**} is defined in (B.9). On the other hand, for the left hand side of (E.26), we can bound it from below in a way such that

$$
\begin{aligned}
& \max_{\tau,\xi\in\mathcal{D}}\left\{\sum_{s=0}^{r-1} L(\beta^{s+1},\alpha^{s+1},\theta^{s+1},\tau,\xi)\right\} \\
= & \sum_{s=0}^{r-1}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s+1}) + \rho\cdot\max_{\tau,\xi\in\mathcal{D}}\left(\sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}, \alpha_{ij}^{s+1} - (\theta_{ij}^{s+1}-\theta_{ji}^{s+1})\right\rangle\right.\right. \\
& \left.\left. + \sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} \left\langle \xi_{ij}, \beta_i^{s+1} - \theta_{ij}^{s+1}\right\rangle\right)\right\} \\
= & \sum_{s=0}^{r-1}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s+1}) + \rho\sum_{(i,j)\in\mathcal{H}} v_{ij}||\alpha_{ij}^{s+1} - (\theta_{ij}^{s+1}-\theta_{ji}^{s+1})||\right. \\
& \left. + \rho\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} w_{ij}||\beta_i^{s+1} - \theta_{ij}^{s+1}||\right\} \\
\geq & \sum_{s=0}^{r-1}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s+1})\right\}. \qquad \text{(E.27)}
\end{aligned}
$$

By applying the result (E.27) to the inequality (E.26), rearranging the index s and the corresponding terms and dividing them by r, we obtain the inequality

$$
\frac{1}{r}\sum_{s=1}^{r}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s})\right\} - \Psi_{\text{primal}}^{\text{opt}} \leq \frac{\rho F_0^{**}}{r}.
$$

Further by Assumption 1, we have

$$
\begin{aligned}
& \frac{1}{r}\sum_{s=1}^{r}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s})\right\} \\
\geq & \sum_{i=1}^{m} l_i\left(\frac{1}{r}\sum_{s=1}^{r}\beta_i^{s}\right) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}\left(\frac{1}{r}\sum_{s=1}^{r}\alpha_{ij}^{s}\right) \\
= & \sum_{i=1}^{m} l_i(\beta_i^{\text{erg},r}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{\text{erg},r}). \qquad \text{(E.28)}
\end{aligned}
$$

The inequality (E.28) implies

$$
\left\{\sum_{i=1}^{m} l_i(\beta_i^{\text{erg},r}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{\text{erg},r})\right\} - \Psi_{\text{primal}}^{\text{opt}} \leq \frac{\rho F_0^{**}}{r},
$$

which completes the proof. □

Remark: Note that from the last line of (E.27) we also have

$$
\begin{aligned}
& \sum_{s=0}^{r-1}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s+1})\right\} \\
\geq & \; r \min_{s\in\{0,1,2,\cdots,r-1\}}\left\{\sum_{i=1}^{m} l_i(\beta_i^{s+1}) + \sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{s+1})\right\},
\end{aligned}
$$

which implies

$$\min_{s\in\{1,2,\cdots,r\}}\left\{\sum_{i=1}^{m} l_i(\beta_i^s)+\sum_{(i,j)\in\mathcal{H}}\lambda_{ij}g_{ij}(\alpha_{ij}^s)\right\}-\Psi_{\text{primal}}^{\text{opt}}\leq\frac{\rho F_0^{**}}{r}.$$

E.3 Proof of Theorem B.2

Below we provide a proof of Theorem B.2. We first give a result that is key for proving the theorem. The results is summarized in Lemma E.3.

Lemma E.3. *Define*

$$\begin{aligned}H^{r+1} \;=\;& \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\beta_i^{r+1}-\theta_{ij}^{r+1}||_2^2+\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\theta_{ij}^{r+1}-\theta_{ij}^{r}||_2^2\\&+\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})||_2^2+\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})-(\theta_{ij}^{r}-\theta_{ji}^{r})||_2^2\\&+\frac{1}{4}\sum_{(i,j)\in\mathcal{H}}||\xi_{ij}^{r+1}+\tau_{ij}^{r+1}||_2^2+\frac{1}{4}\sum_{(i,j)\in\mathcal{H}}||\xi_{ji}^{r+1}-\tau_{ij}^{r+1}||_2^2.\end{aligned}\tag{E.29}$$

Under Assumpton 1, we have

$$H^{r+1}\leq H^r,\tag{E.30}$$

i.e. $\{H^r\}_r$ is a non-increasing sequence.

Proof of Lemma E.3. Given that Assumption 1 is true, we can directly apply Lemma E.1. Now by letting $(\beta,\alpha,\theta,\tau,\xi)=(\beta^r,\alpha^r,\theta^r,\tau^r,\xi^r)$ in the inequality (E.6) in Lemma E.1 we obtain

$$\begin{aligned}L(\beta^{r+1},\alpha^{r+1},\theta^{r+1},\tau^r,\xi^r)\;\leq\;& L(\beta^r,\alpha^r,\theta^r,\tau^{r+1},\xi^{r+1})\\&+\rho\Bigg[A(\beta^r)^{r+1}+B(\alpha^r)^{r+1}+C(\theta^r)^{r+1}+D(\theta^r)^{r+1}\\&+\sum_{(i,j)\in\mathcal{H}}\Omega(\tau_{ij}^r,\tau_{ij})^{r+1}-\frac{1}{2}\sum_{(i,j\in\mathcal{H})}||\tau_{ij}^{r+1}-\tau_{ij}^r||_2^2\\&+\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\Omega(\xi_{ij}^r,\xi_{ij})^{r+1}-\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\xi_{ij}^{r+1}-\xi_{ij}^r||_2^2\Bigg].\end{aligned}\tag{E.31}$$

On the other hand, by replacing $r+1$ with r and letting $(\beta,\alpha,\theta,\tau,\xi)=(\beta^{r+1},\alpha^{r+1},\theta^{r+1},\tau^{r+1},\xi^{r+1})$

in the inequality (E.6) in Lemma E.1 we obtain

$$\begin{aligned}
L(\beta^r,\alpha^r,\theta^r,\tau^{r+1},\xi^{r+1}) \quad \leq \quad & L(\beta^{r+1},\alpha^{r+1},\theta^{r+1},\tau^r,\xi^r) \\
& +\rho\Bigg[A(\beta^{r+1})^r + B(\alpha^{r+1})^r + C(\theta^{r+1})^r + D(\theta^{r+1})^r \\
& + \sum_{(i,j)\in\mathcal{H}} \Omega(\tau_{ij}^{r+1},\tau_{ij})^r - \frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2 \\
& + \sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} \Omega(\xi_{ij}^{r+1},\xi_{ij})^r - \frac{1}{2}\sum_{i=1}^m\sum_{j\in\mathcal{N}(i)} ||\xi_{ij}^{r-1} - \xi_{ij}^{r-1}||_2^2\Bigg].
\end{aligned} \tag{E.32}$$

Adding (E.31) and (E.32) yields

$$\begin{aligned}
0 \quad \leq \quad & C(\theta^r)^{r+1} + D(\theta^r)^{r+1} + C(\theta^{r+1})^r + D(\theta^{r+1})^r \\
& + \sum_{(i,j)\in\mathcal{H}} [\Omega(\tau_{ij}^r,\tau_{ij})^{r+1} + \Omega(\tau_{ij}^{r+1},\tau)^r] - \frac{1}{2}\sum_{(i,j\in\mathcal{H})}\left(||\tau_{ij}^{r+1} - \tau_{ij}^r||_2^2 + ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2\right) \\
& + B(\alpha^r)^{r+1} + B(\alpha^{r+1})^r \\
& + \sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} [\Omega(\xi_{ij}^r,\xi_{ij})^{r+1} + \Omega(\xi^{r+1},\xi_{ij})^r] - \frac{1}{2}\sum_{i=1}^m\sum_{j\in\mathcal{N}(i)}\left(||\xi_{ij}^{r+1} - \xi_{ij}^r||_2^2 + ||\xi_{ij}^r - \xi_{ij}^{r-1}||_2^2\right) \\
& + [A(\beta^r)^{r+1} + A(\beta^{r+1})^r].
\end{aligned} \tag{E.33}$$

The first line of (E.33): By applying the expression (E.22) (see the proof of Lemma E.2) we have

$$\begin{aligned}
& C(\theta^r)^{r+1} + D(\theta^r)^{r+1} + C(\theta^{r+1})^r + D(\theta^{r+1})^r \\
= \quad & \sum_{(i,j)\in\mathcal{H}} \Omega(\theta_{ij}^r + \theta_{ji}^r, \theta_{ij} + \theta_{ji})^{r+1} + \sum_{(i,j)\in\mathcal{H}} \Omega(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}, \theta_{ij} + \theta_{ji})^r \\
& -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^r + \theta_{ji}^r)||_2^2 - \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^r + \theta_{ji}^r) - (\theta_{ij}^{r-1} + \theta_{ji}^{r-1})||_2^2.
\end{aligned} \tag{E.34}$$

Next by applying the result (E.75) in Proposition E.4 with $a = \theta_{ij} + \theta_{ji}$ to the right hand side of

(E.34) and using the result in Lemma E.6 we further obtain

$$
\begin{aligned}
& C(\theta^r)^{r+1} + D(\theta^r)^{r+1} + C(\theta^{r+1})^r + D(\theta^{r+1})^r \\
= \;& \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^r + \theta_{ji}^r) - (\theta_{ij}^{r-1} + \theta_{ji}^{r-1})||_2^2 - \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^r + \theta_{ji}^r)||_2^2 \\
& -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||[(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^r + \theta_{ji}^r)] - [(\theta_{ij}^r + \theta_{ji}^r) - (\theta_{ij}^{r-1} + \theta_{ji}^{r-1})]||_2^2 \\
= \;& \frac{1}{4}\sum_{(i,j)\in\mathcal{H}} ||\xi_{ij}^r + \tau_{ij}^r||_2^2 + \frac{1}{4}\sum_{(i,j)\in\mathcal{H}} ||\xi_{ji}^r - \tau_{ij}^r||_2^2 \\
& -\frac{1}{4}\sum_{(i,j)\in\mathcal{H}} ||\xi_{ij}^{r+1} + \tau_{ij}^{r+1}||_2^2 - \frac{1}{4}\sum_{(i,j)\in\mathcal{H}} ||\xi_{ji}^{r+1} - \tau_{ij}^{r+1}||_2^2 \\
& -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||[(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^r + \theta_{ji}^r)] - [(\theta_{ij}^r + \theta_{ji}^r) - (\theta_{ij}^{r-1} + \theta_{ji}^{r-1})]||_2^2. \qquad \text{(E.35)}
\end{aligned}
$$

The second and third lines of (E.33): Again by applying the result (E.75) in Proposition E.4 with $a = \tau_{ij}$, we have

$$
\begin{aligned}
& \sum_{(i,j)\in\mathcal{H}} [\Omega(\tau_{ij}^r, \tau_{ij})^{r+1} + \Omega(\tau_{ij}^{r+1}, \tau)^r] - \frac{1}{2}\sum_{(i,j\in\mathcal{H})} \left(||\tau_{ij}^{r+1} - \tau_{ij}^r||^2 + ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2 \right) \\
= \;& \frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2 - \frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||\tau_{ij}^{r+1} - \tau_{ij}^r||_2^2 \\
& -\frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||(\tau_{ij}^{r+1} - \tau_{ij}^r) - (\tau_{ij}^r - \tau_{ij}^{r-1})||_2^2. \qquad \text{(E.36)}
\end{aligned}
$$

Note that by applying the relation (2.8) we can express the third term on the right hand side of (E.36) as

$$
\frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||(\tau_{ij}^{r+1} - \tau_{ij}^r) - (\tau_{ij}^r - \tau_{ij}^{r-1})||_2^2 = \frac{1}{2}\sum_{(i,j\in\mathcal{H})} ||[(\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^r - \theta_{ji}^r)] - (\alpha_{ij}^{r+1} - \alpha_{ij}^r)]||_2^2. \qquad \text{(E.37)}
$$

On the other hand, we have

$$
\begin{aligned}
& B(\alpha^r)^{r+1} + B(\alpha^{r+1})^r \\
= \;& -\sum_{(i,j)\in\mathcal{H}} \left\langle [(\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^r - \theta_{ji}^r)] - [(\theta_{ij}^r - \theta_{ji}^r) - (\theta_{ij}^{r-1} - \theta_{ji}^{r-1})], \alpha_{ij}^{r+1} - \alpha_{ij}^r \right\rangle. \qquad \text{(E.38)}
\end{aligned}
$$

Applying the results in (E.36), (E.37) and (E.38) we obtain

$$
\begin{aligned}
& \sum_{(i,j)\in\mathcal{H}} [\Omega(\tau_{ij}^r, \tau_{ij})^{r+1} + \Omega(\tau_{ij}^{r+1}, \tau)^r] - \frac{1}{2}\sum_{(i,j\in\mathcal{H})}\left(||\tau_{ij}^{r+1} - \tau_{ij}^r||_2^2 + ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2\right) \\
& +B(\alpha^r)^{r+1} + B(\alpha^{r+1})^r \\
= & \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2 - \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\tau_{ij}^r - \tau_{ij}^{r+1}||_2^2 \\
& -\Bigg\{ \sum_{(i,j)\in\mathcal{H}} \Bigg[\frac{1}{2}||[(\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^r - \theta_{ji}^r) - (\alpha_{ij}^{r+1} - \alpha_{ij}^r)]||_2^2 \\
& + \big\langle [(\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^r - \theta_{ji}^r)] - [(\theta_{ij}^r - \theta_{ji}^r) - (\theta_{ij}^{r-1} - \theta_{ji}^{r-1})], \alpha_{ij}^{r+1} - \alpha_{ij}^r \big\rangle \Bigg\}.
\end{aligned}
\tag{E.39}
$$

Now we need the inequality $(1/2)||a-b||_2^2 + \langle a-c, b\rangle = (1/2)||a||_2^2 + (1/2)||b-c||_2^2 - (1/2)||c||_2^2$. By letting $a = (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^r - \theta_{ji}^r)$, $b = \alpha_{ij}^{r+1} - \alpha_{ij}^r$ and $c = (\theta_{ij}^r - \theta_{ji}^r) - (\theta_{ij}^{r-1} - \theta_{ij}^{r-1})$, and using the relation (2.8), we can further simplify the right hand side of (E.39) and obtain

$$
\begin{aligned}
& \sum_{(i,j)\in\mathcal{H}} [\Omega(\tau_{ij}^r, \tau_{ij})^{r+1} + \Omega(\tau_{ij}^{r+1}, \tau)^r] - \frac{1}{2}\sum_{(i,j\in\mathcal{H})}\left(||\tau_{ij}^{r+1} - \tau_{ij}^r||_2^2 + ||\tau_{ij}^r - \tau_{ij}^{r-1}||_2^2\right) \\
& +B(\alpha^r)^{r+1} + B(\alpha^{r+1})^r \\
= & \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^r - (\theta_{ij}^r - \theta_{ji}^r)||_2^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^r - \theta_{ji}^r) - (\theta_{ij}^{r-1} - \theta_{ji}^{r-1})||_2^2 \\
& -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1})||_2^2 - \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^r - \theta_{ji}^r)||_2^2 \\
& -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\alpha_{ij}^{r+1} - \alpha_{ij}^r) - [(\theta_{ij}^r - \theta_{ji}^r) - (\theta_{ij}^{r-1} - \theta_{ij}^{r-1})]||_2^2.
\end{aligned}
\tag{E.40}
$$

The fourth and fifth lines of (E.33): Applying an argument similar to the one for dealing with the second and third lines of (E.33), we obtain

$$
\begin{aligned}
& +\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} [\Omega(\xi_{ij}^r, \xi_{ij})^{r+1} + \Omega(\xi^{r+1}, \xi_{ij})^r] - \frac{1}{2}\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)}\left(||\xi_{ij}^{r+1} - \xi_{ij}^r||_2^2 + ||\xi_{ij}^r - \xi_{ij}^{r-1}||_2^2\right) \\
& +A(\alpha^r)^{r+1} + A(\alpha^{r+1})^r \\
= & \frac{1}{2}\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} ||\beta_i^r - \theta_{ij}^r||_2^2 + \frac{1}{2}\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} ||\theta_{ij}^r - \theta_{ij}^{r-1}||_2^2 \\
& -\frac{1}{2}\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} ||\beta_i^{r+1} - \theta_{ij}^{r+1}||_2^2 - \frac{1}{2}\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} ||\theta_{ij}^{r+1} - \theta_{ij}^r||_2^2 \\
& -\frac{1}{2}\sum_{i=1}^m \sum_{j\in\mathcal{N}(i)} ||(\beta_i^{r+1} - \beta_i^r) - (\theta_{ij}^r - \theta_{ij}^{r-1})||_2^2.
\end{aligned}
\tag{E.41}
$$

Combining the results: Applying the results in (E.35), (E.40) and (E.41) to the right hand side of (E.33) and using the definition of H^{r+1} in (E.29), we obtain the following inequality:

$$\begin{aligned} 0 \leq\ & H^r - H^{r+1} \\ & -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||[(\theta_{ij}^{r+1}+\theta_{ji}^{r+1}) - (\theta_{ij}^{r}+\theta_{ji}^{r})] - [(\theta_{ij}^{r}+\theta_{ji}^{r}) - (\theta_{ij}^{r-1}+\theta_{ji}^{r-1})]||_2^2 \\ & -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\alpha_{ij}^{r+1}-\alpha_{ij}^{r}) - [(\theta_{ij}^{r}-\theta_{ji}^{r}) - (\theta_{ij}^{r-1}-\theta_{ij}^{r-1})]||_2^2 \\ & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||(\beta_i^{r+1}-\beta_i^{r}) - (\theta_{ij}^{r}-\theta_{ij}^{r-1})||_2^2, \end{aligned}$$

which implies $H^{r+1} \leq H^r$. The proof is complete. □

Proof of Theorem B.2. Under Assumption 2 there exists a saddle point $(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*)$ of $L(\beta, \alpha, \theta, \tau, \xi)$, and by definition of the saddle point we will have

$$L(\beta^*, \alpha^*, \theta^*, \tau^{r+1}, \xi^{r+1}) \leq L(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*) \leq L(\beta^{r+1}, \alpha^{r+1}, \theta^{r+1}, \tau^*, \xi^*) \tag{E.42}$$

for $r = 0, 1, 2, \cdots$. In addition, $(\beta^*, \alpha^*, \theta^*)$ solves the primal problem (2.1), and therefore β^*, α^* and θ^* will satisfy the relations $\beta_i^* - \theta_{ij}^* = 0$ and $\alpha_{ij}^* - (\theta_{ij}^* - \theta_{ji}^*) = 0$. As a result of that, we have

$$L(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*) = L(\beta^*, \alpha^*, \theta^*, \tau^{r+1}, \xi^{r+1}). \tag{E.43}$$

Now by letting $(\beta, \alpha, \theta, \tau, \xi) = (\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*)$ in the inequality (E.15) in Lemma E.2 and applying the results in (E.42) and (E.43), we obtain

$$\begin{aligned} 0 \leq\ & \sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} \Omega(\beta_i^*, \theta_{ij})^{r+1} + \sum_{(i,j)\in\mathcal{H}} \Omega(\alpha_{ij}^*, \theta_{ij}-\theta_{ji})^{r+1} \\ & + \sum_{(i,j)\in\mathcal{H}} \Omega(\tau_{ij}^*, \tau_{ij})^{r+1} + \sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} \Omega(\xi_{ij}^*, \xi_{ij})^{r+1} \\ & + \sum_{(i,j)\in\mathcal{H}} \Omega(\theta_{ij}^*+\theta_{ji}^*, \theta_{ij}+\theta_{ji})^{r+1} \\ & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\beta_i^{r+1}-\theta_{ij}^{r}||_2^2 - \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^{r+1} - (\theta_{ij}^{r}-\theta_{ji}^{r})||_2^2 \\ & -\frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^{r+1}+\theta_{ji}^{r+1}) - (\theta_{ij}^{r}+\theta_{ji}^{r})||_2^2. \end{aligned} \tag{E.44}$$

By applying the relation (2.8) we can express the sixth term of (E.44) as

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\beta_i^{r+1}-\theta_{ij}^r||_2^2 \quad = \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\beta_i^{r+1}-\theta_{ij}^{r+1}||_2^2+\frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\theta_{ij}^{r+1}-\theta_{ij}^r||_2^2 \\ & +\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\left\langle \xi_{ij}^{r+1}-\xi_{ij}^r, \theta_{ij}^{r+1}-\theta_{ij}^r\right\rangle. \end{aligned} \tag{E.45}$$

Following a similar argument we can express the seventh term of (E.44) as

$$\begin{aligned} & \frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\alpha_{ij}^{r+1}-(\theta_{ij}^r-\theta_{ji}^r)||_2^2 \\ = \quad & \frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||\alpha_{ij}^{r+1}-(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})||_2^2+\frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||(\theta_{ij}^{r+1}-\theta_{ji}^{r+1})-(\theta_{ij}^r-\theta_{ji}^r)||_2^2 \\ & +\sum_{(i,j)\in\mathcal{H}}\left\langle \tau_{ij}^{r+1}-\tau_{ij}^r, (\theta_{ij}^{r+1}-\theta_{ji}^{r+1})-(\theta_{ij}^r-\theta_{ji}^r)\right\rangle. \end{aligned} \tag{E.46}$$

On the other hand, by applying the result in Proposition E.6 we can express the eighth term of (E.44) as

$$\begin{aligned} & \frac{1}{2}\sum_{(i,j)\in\mathcal{H}}||(\theta_{ij}^{r+1}+\theta_{ji}^{r+1})-(\theta_{ij}^r+\theta_{ji}^r)||_2^2 \\ = \quad & \frac{1}{4}\sum_{(i,j)\in\mathcal{H}}||\xi_{ij}^{r+1}+\tau_{ij}^{r+1}||^2+\frac{1}{4}\sum_{(i,j)\in\mathcal{H}}||\xi_{ji}^{r+1}-\tau_{ij}^{r+1}||_2^2. \end{aligned} \tag{E.47}$$

Now we turn to the final terms of (E.45) and (E.46). Again by applying the result in Proposition E.6 we have

$$\begin{aligned} & \left\langle \xi_{ij}^{r+1}-\xi_{ij}^r, \theta_{ij}^{r+1}-\theta_{ij}^r\right\rangle+\left\langle \xi_{ji}^{r+1}-\xi_{ji}^r, \theta_{ji}^{r+1}-\theta_{ji}^r\right\rangle \\ & +\left\langle \tau_{ij}^{r+1}-\tau_{ij}^r, (\theta_{ij}^{r+1}-\theta_{ji}^{r+1})-(\theta_{ij}^r-\theta_{ji}^r)\right\rangle \\ = \quad & \left\langle (\xi_{ij}^{r+1}+\tau_{ij}^{r+1})-(\xi_{ij}^r+\tau_{ij}^r), \theta_{ij}^{r+1}-\theta_{ij}^r\right\rangle+\left\langle (\xi_{ji}^{r+1}-\tau_{ij}^{r+1})-(\xi_{ji}^r-\tau_{ij}^r), \theta_{ji}^{r+1}-\theta_{ji}^r\right\rangle \\ = \quad & ||(\theta_{ij}^{r+1}+\theta_{ji}^{r+1})-(\theta_{ij}^r+\theta_{ji}^r)||_2^2 \\ & -\left\langle (\theta_{ij}^r+\theta_{ji}^r)-(\theta_{ij}^{r-1}+\theta_{ji}^{r-1}), (\theta_{ij}^{r+1}+\theta_{ji}^{r+1})-(\theta_{ij}^r+\theta_{ji}^r)\right\rangle. \end{aligned} \tag{E.48}$$

By applying the identity $||a-b||_2^2-\langle a-b, b-c\rangle = (1/2)||(a-b)-(b-c)||_2^2+(1/2)||a-b||_2^2-(1/2)||b-c||_2^2$ with $a=\theta_{ij}^{r+1}+\theta_{ji}^{r+1}$, $b=\theta_{ij}^r+\theta_{ji}^r$ and $c=\theta_{ij}^{r-1}+\theta_{ji}^{r-1}$, we can express the right hand side of (E.48) as

$$
\begin{aligned}
& \langle \xi_{ij}^{r+1} - \xi_{ij}^{r}, \theta_{ij}^{r+1} - \theta_{ij}^{r} \rangle + \langle \xi_{ji}^{r+1} - \xi_{ji}^{r}, \theta_{ji}^{r+1} - \theta_{ji}^{r} \rangle + \langle \tau_{ij}^{r+1} - \tau_{ij}^{r}, (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^{r} - \theta_{ji}^{r}) \rangle \\
= & \frac{1}{2} ||[(\theta_{ij}^{r} + \theta_{ji}^{r}) - (\theta_{ij}^{r+1} + \theta_{ji}^{r+1})] - [(\theta_{ij}^{r-1} + \theta_{ji}^{r-1}) - (\theta_{ij}^{r} + \theta_{ji}^{r})]||_2^2 \\
& + \frac{1}{2} ||(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^{r} + \theta_{ji}^{r})||_2^2 - \frac{1}{2} ||(\theta_{ij}^{r} + \theta_{ji}^{r}) - (\theta_{ij}^{r-1} + \theta_{ji}^{r-1})||_2^2. \qquad \text{(E.49)}
\end{aligned}
$$

By applying the results in (E.45), (E.46), (E.47) and (E.49) to the inequality (E.44) and using the definition of H^{r+1} in (E.29), we can obtain the following inequality:

$$
\begin{aligned}
H^{r+1} \leq & \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} \Omega(\beta_i^*, \theta_{ij})^{r+1} + \sum_{(i,j) \in \mathcal{H}} \Omega(\alpha_{ij}^*, \theta_{ij} - \theta_{ji})^{r+1} \\
& + \sum_{(i,j) \in \mathcal{H}} \Omega(\tau_{ij}^*, \tau_{ij})^{r+1} + \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} \Omega(\xi_{ij}^*, \xi_{ij})^{r+1} \\
& + \sum_{(i,j) \in \mathcal{H}} \Omega(\theta_{ij}^* + \theta_{ji}^*, \theta_{ij} + \theta_{ji})^{r+1} \\
& + \frac{1}{2} \sum_{(i,j) \in \mathcal{H}} ||(\theta_{ij}^{r} + \theta_{ji}^{r}) - (\theta_{ij}^{r-1} + \theta_{ji}^{r-1})||_2^2 - \frac{1}{2} \sum_{(i,j) \in \mathcal{H}} ||(\theta_{ij}^{r+1} + \theta_{ji}^{r+1}) - (\theta_{ij}^{r} + \theta_{ji}^{r})||_2^2.
\end{aligned}
\qquad \text{(E.50)}
$$

Further by summing up the inequality (E.50) from $r = 0$ to $r = s - 1$ and using the result in Proposition E.3, we obtain

$$
\begin{aligned}
\sum_{s=0}^{r-1} H^{s+1} \leq & \frac{1}{2} \Bigg(\sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} ||\beta^* - \theta_{ij}^{0}||_2^2 + \sum_{(i,j) \in \mathcal{H}} ||\alpha_{ij}^* - (\theta_{ij}^{0} - \theta_{ji}^{0})||_2^2 \\
& + \sum_{(i,j) \in \mathcal{H}} ||\tau_{ij}^* - \tau_{ij}^{0}||_2^2 + \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} ||\xi_{ij}^* - \xi_{ij}^{0}||_2^2 \\
& + \sum_{(i,j) \in \mathcal{H}} ||\theta_{ij}^* + \theta_{ji}^* - (\theta_{ij}^{0} + \theta_{ji}^{0})||_2^2 \Bigg) + \frac{1}{2} \sum_{(i,j) \in \mathcal{H}} ||(\theta_{ij}^{0} + \theta_{ji}^{0}) - (\theta_{ij}^{-1} + \theta_{ji}^{-1})||_2^2.
\end{aligned}
\qquad \text{(E.51)}
$$

Since θ_{ij}^{0} and θ_{ji}^{0} are initial values for θ_{ij} and θ_{ji}, and for convenience we let θ_{ij}^{-1} and θ_{ij}^{-1}, the values before the initial values, equal to the initial values. As a result of that, the last term of (E.51) is vanished. In addition, $||\theta_{ij}^* + \theta_{ji}^* - (\theta_{ij}^{0} + \theta_{ji}^{0})||_2^2 \leq 2||\theta_{ij}^* - \theta_{ij}^{0}||_2^2 + 2||\theta_{ji}^* - \theta_{ji}^{0}||_2^2$. Therefore inequality

(E.51) implies that

$$
\begin{aligned}
\sum_{s=0}^{r-1} H^{s+1} \quad \leq \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\beta^* - \theta_{ij}^0||_2^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^* - (\theta_{ij}^0 - \theta_{ji}^0)||_2^2 \\
& + \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\tau_{ij}^* - \tau_{ij}^0||_2^2 + \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\xi_{ij}^* - \xi_{ij}^0||_2^2 \\
& + \sum_{(i,j)\in\mathcal{H}} ||\theta_{ij}^* - \theta_{ij}^0||_2^2 + \sum_{(i,j)\in\mathcal{H}} ||\theta_{ji}^* - \theta_{ji}^0||_2^2 \\
= \quad & F_0(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*) \\
= \quad & F_0^*,
\end{aligned}
$$

where $F_0(\beta^*, \alpha^*, \theta^*, \tau^*, \xi^*)$ and F_0^* are the same as those given in (B.14). Note that from Lemma E.3 we know $H^{r+1} \leq H^r$. Now by applying (E.52) we further obtain

$$
\begin{aligned}
0 \quad \leq \quad & \sum_{s=0}^{r-1} s(H^s - H^{s+1}) \\
= \quad & \sum_{s=0}^{r-1} [sH^s - (s+1)H^{s+1} + H^{s+1}] \\
= \quad & -H^1 + H^1 - 2H^2 + 2H^2 - \cdots + (r-1)H^{r-1} - rH^r + \sum_{s=0}^{r-1} H^{s+1} \\
\leq \quad & -rH^r + F_0^*.
\end{aligned} \tag{E.52}
$$

The result in (E.52) implies

$$
\begin{aligned}
H^r \quad = \quad & \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\beta_i^r - \theta_{ij}^r||_2^2 + \frac{1}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)} ||\theta_{ij}^r - \theta_{ij}^{r-1}||_2^2 \\
& + \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||\alpha_{ij}^{r+1} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1})||_2^2 + \frac{1}{2}\sum_{(i,j)\in\mathcal{H}} ||(\theta_{ij}^r - \theta_{ji}^r) - (\theta_{ij}^{r-1} - \theta_{ji}^{r-1})||_2^2 \\
& + \frac{1}{4}\sum_{(i,j)\in\mathcal{H}} ||\xi_{ij}^r + \tau_{ij}^r||_2^2 + \frac{1}{4}\sum_{(i,j)\in\mathcal{H}} ||\xi_{ji}^r - \tau_{ij}^r||_2^2. \\
\leq \quad & F_0^* \cdot \frac{1}{r},
\end{aligned}
$$

which further implies

$$\sqrt{\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}||\beta_i^r-\theta_{ij}^r||_2^2} \leq \sqrt{\frac{2F_0^*}{r}}$$

$$\sqrt{\sum_{(i,j)\in\mathcal{H}}||\alpha_{ij}^r-(\theta_{ij}^r-\theta_{ji}^r)||_2^2} \leq \sqrt{\frac{2F_0^*}{r}}$$

$$\sqrt{\sum_{(i,j)\in\mathcal{H}}||\xi_{ij}^r+\tau_{ij}^r||_2^2} \leq \sqrt{\frac{4F_0^*}{r}},$$

$$\sqrt{\sum_{(i,j)\in\mathcal{H}}||\xi_{ji}^r-\tau_{ij}^r||_2^2} \leq \sqrt{\frac{4F_0^*}{r}}.$$

This completes the proof. □

E.4 Technical lemmas

Lemma E.4. *Under Assumption 1, we have*

$$\sum_{i=1}^{m} l_i(\beta_i^{r+1})+\rho\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\left\langle\xi_{ij}^{r+1},\beta_i^{r+1}\right\rangle \leq \sum_{i=1}^{m} l_i(\beta_i)+\rho\sum_{i=1}^{m}\sum_{j\in\mathcal{N}(i)}\left\langle\xi_{ij}^{r+1},\beta_i\right\rangle+\rho A(\beta)^{r+1}, \tag{E.53}$$

where $A(\beta)^{r+1}$ is defined in (E.1).

Proof of Lemma E.4. By definition (2.4), β_i^{r+1} is also a solution to the following system of equations:

$$0=\nabla_i l_i(\beta_i^{r+1})+\rho|\mathcal{N}(i)|\beta_i^{r+1}-\rho\sum_{j\in\mathcal{N}(i)}(\theta_{ij}^r-\xi_{ij}^r). \tag{E.54}$$

By carrying out the inner product multiplication on both sides of (E.54) with vector $\beta_i-\beta_i^{r+1}$, we obtain

$$0=\left\langle\nabla_i l_i(\beta_i^{r+1}),\beta_i-\beta_i^{r+1}\right\rangle+\rho\sum_{j\in\mathcal{N}(i)}\left\langle\beta_i^{r+1}-\theta_{ij}^r+\xi_{ij}^r,\beta_i-\beta_i^{r+1}\right\rangle. \tag{E.55}$$

Under Assumption 1 $l_i(\beta_i)$ is a convex function of β_i, and therefore we have

$$l_i(\beta_i^{r+1})\leq l_i(\beta_i)-\left\langle\nabla l_i(\beta_i^{r+1}),\beta_i-\beta_i^{r+1}\right\rangle. \tag{E.56}$$

By adding (E.55) and (E.56) and using the relation (2.9), we obtain

$$l_i(\beta_i^{r+1})\leq l_i(\beta_i)+\rho\sum_{j\in\mathcal{N}(i)}\left\langle\theta_{ij}^{r+1}-\theta_{ij}^r,\beta_i-\beta_i^{r+1}\right\rangle+\rho\sum_{j\in\mathcal{N}(i)}\left\langle\xi_{ij}^{r+1},\beta_i-\beta_i^{r+1}\right\rangle. \tag{E.57}$$

By rearranging (E.57) we obtain

$$\begin{aligned} l_i(\beta_i^{r+1}) + \rho \sum_{j\in\mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \beta_i^{r+1} \right\rangle &\leq l_i(\beta_i) + \rho \sum_{j\in\mathcal{N}(i)} \left\langle \xi_{ij}^{r+1}, \beta_i \right\rangle \\ &\quad + \rho \sum_{j\in\mathcal{N}(i)} \left\langle \theta_{ij}^{r+1} - \theta_{ij}^{r}, \beta_i - \beta_i^{r+1} \right\rangle . \end{aligned} \tag{E.58}$$

By summing up (E.58) from $i = 1$ to $i = m$ and using the definition of $A(\beta)^{r+1}$ in (E.1), we obtain (E.53), which completes the proof. □

Lemma E.5. *Under Assumption 1, we have*

$$\begin{aligned} &\sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}^{r+1}) + \rho \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \alpha_{ij}^{r+1} \right\rangle \\ \leq &\sum_{(i,j)\in\mathcal{H}} \lambda_{ij} g_{ij}(\alpha_{ij}) + \rho \sum_{(i,j)\in\mathcal{H}} \left\langle \tau_{ij}^{r+1}, \alpha_{ij} \right\rangle + \rho B(\alpha)^{r+1}, \end{aligned} \tag{E.59}$$

where $B(\alpha)^{r+1}$ is define in (E.2)

Proof of Lemma E.5. By definition (2.5) α_{ij}^{r+1} is also a solution to the following system of equations:

$$0 = u_{ij}^{r+1} + \frac{\rho}{\lambda_{ij}} [\alpha_{ij}^{r+1} - (\theta_{ij}^{r} - \theta_{ji}^{r}) + \tau_{ij}^{r}], \tag{E.60}$$

where $u_{ij}^{r+1} \in \partial g_{ij}(\alpha_{ij}^{r+1})$ is a subgradient of $g_{ij}(x)$ evaluated at α_{ij}^{r+1}. Now by carrying out the inner product multiplication on both sides of (E.60) with $\alpha_{ij} - \alpha_{ij}^{r+1}$ we obtain

$$0 = \left\langle a_{ij}^{r+1}, \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle + \frac{\rho}{\lambda_{ij}} \left\langle \alpha_{ij}^{r+1} - (\theta_{ij}^{r} - \theta_{ji}^{r}) + \tau_{ij}^{r}, \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle . \tag{E.61}$$

Under Assumption 1 $g_{ij}(\alpha_{ij})$ is a convex function of α_{ij}, and therefore we have

$$g_{ij}(\alpha_{ij}^{r+1}) \leq g_{ij}(\alpha_{ij}) - \left\langle a_{ij}^{r+1}, \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle . \tag{E.62}$$

Adding (E.61) and (E.62), multiplying both sides with λ_{ij}, and using the relation (2.8) we obtain

$$\begin{aligned} \lambda_{ij} g_{ij}(\alpha_{ij}^{r+1}) &\leq \lambda_{ij} g_{ij}(\alpha_{ij}) + \rho \left\langle \tau_{ij}^{r+1}, \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle \\ &\quad + \rho \left\langle (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^{r} - \theta_{ji}^{r}), \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle . \end{aligned} \tag{E.63}$$

By rearranging (E.63) we obtain

$$\lambda_{ij} g_{ij}(\alpha_{ij}^{r+1}) + \rho \left\langle \tau_{ij}^{r+1}, \alpha_{ij}^{r+1} \right\rangle \leq \lambda_{ij} g_{ij}(\alpha_{ij}) + \rho \left\langle \tau_{ij}^{r+1}, \alpha_{ij} \right\rangle + \rho \left\langle (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - (\theta_{ij}^{r} - \theta_{ji}^{r}), \alpha_{ij} - \alpha_{ij}^{r+1} \right\rangle . \tag{E.64}$$

By summing up (E.64) for all $(i,j) \in \mathcal{H}$ and using the definition of $B(\alpha)^{r+1}$ in (E.2), we obtain (E.59), which completes the proof. □

Lemma E.6. *From the relations (2.6), (2.7), (2.8), (2.9) and (2.10), we have*

$$\xi_{ij}^{r+1} + \tau_{ij}^{r+1} = \xi_{ji}^{r+1} - \tau_{ij}^{r+1} = \theta_{ij}^{r+1} - \theta_{ij}^{r} + \theta_{ji}^{r+1} - \theta_{ji}^{r}, \tag{E.65}$$

Proof of Lemma E.6. From the relation (2.6) we have

$$\begin{aligned} & 3\theta_{ij}^{r+1} = \theta_{ji}^{r} + \alpha_{ij}^{r+1} + \tau_{ij}^{r} + \beta_{i}^{r+1} + \xi_{ij}^{r} + \theta_{ij}^{r} \\ \Leftrightarrow \quad & \xi_{ij}^{r} + \tau_{ij}^{r} = 3\theta_{ij}^{r+1} - \theta_{ji}^{r} - \alpha_{ij}^{r+1} - \beta_{i}^{r+1} - \theta_{ij}^{r}. \end{aligned} \tag{E.66}$$

By using the relations (2.8) and (2.9) and the result in (E.66), we obtain

$$\begin{aligned} \xi_{ij}^{r+1} + \tau_{ij}^{r+1} &= \tau_{ij}^{r} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) + \alpha_{ij}^{r+1} + \xi_{ij}^{r} - \theta_{ij}^{r+1} + \beta_{i}^{r+1} \\ &= 3\theta_{ij}^{r+1} - \theta_{ji}^{r} - \alpha_{ij}^{r+1} - \beta_{i}^{r+1} - \theta_{ij}^{r} - (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) + \alpha_{ij}^{r+1} - \theta_{ij}^{r+1} + \beta_{i}^{r+1} \\ &= \theta_{ij}^{r+1} - \theta_{ij}^{r} + \theta_{ji}^{r+1} - \theta_{ji}^{r}. \end{aligned} \tag{E.67}$$

On the other hand, from the relation (2.7), we have

$$\begin{aligned} & 3\theta_{ji}^{r+1} = \theta_{ij}^{r} - \alpha_{ij}^{r+1} - \tau_{ij}^{r} + \beta_{j}^{r+1} + \xi_{ji}^{r} + \theta_{ji}^{r} \\ \Leftrightarrow \quad & \xi_{ji}^{r} - \tau_{ij}^{r} = 3\theta_{ji}^{r+1} - \theta_{ij}^{r} + \alpha_{ij}^{r+1} - \beta_{j}^{r+1} - \theta_{ji}^{r}. \end{aligned} \tag{E.68}$$

By using the result in (E.68) and the relations (2.8) and (2.10), we obtain

$$\begin{aligned} \xi_{ji}^{r+1} - \tau_{ij}^{r+1} &= \xi_{ji}^{r} - \theta_{ji}^{r+1} + \beta_{j}^{r+1} - \tau_{ij}^{r} + (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - \alpha_{ij}^{r+1} \\ &= 3\theta_{ji}^{r+1} - \theta_{ij}^{r} + \alpha_{ij}^{r+1} - \beta_{j}^{r+1} - \theta_{ji}^{r} - \theta_{ji}^{r+1} + \beta_{j}^{r+1} + (\theta_{ij}^{r+1} - \theta_{ji}^{r+1}) - \alpha_{ij}^{r+1} \\ &= \theta_{ji}^{r+1} - \theta_{ji}^{r} + \theta_{ij}^{r+1} - \theta_{ij}^{r}. \end{aligned} \tag{E.69}$$

The results in (E.67) and (E.69) imply (E.65), which completes the proof. □

Proposition E.1. *For $a, b, c, d \in \mathbb{R}^p$, we have*

$$\langle a - b, c - d\rangle \quad = \quad \frac{1}{2}\bigg(||a-d||^2 - ||a-c||^2 + ||b-c||^2 - ||b-d||^2\bigg). \tag{E.70}$$

In particular, when $b = c$, we have

$$\langle a - c, c - d\rangle \quad = \quad \frac{1}{2}\bigg(||a-d||^2 - ||a-c||^2 - ||c-d||^2\bigg). \tag{E.71}$$

Remark on proofs: The results in Proposition E.1 can be verified by directly computing left hand sides of (E.70) and (E.71) using elementary linear algebra techniques.

Proposition E.2. *For $\Omega(a, b)^{r+1}$ defined in (E.5), we have*

$$\sum_{r=0}^{s-1} \Omega(a, b)^{r+1} \leq \frac{1}{2}||a - b^0||^2.$$

Proof of Proposition E.2. By direct calculation we have

$$\begin{aligned}
\sum_{r=0}^{s-1} \Omega(a, b)^{r+1} \quad &= \quad \frac{1}{2}\bigg(||a-b^0||^2 - ||a-b^1||^2\bigg) + \frac{1}{2}\bigg(||a-b^1||^2 - ||a-b^2||^2\bigg) + \cdots \\
&\quad + \frac{1}{2}\bigg(||a-b^{s-1}||^2 - ||a-b^s||^2\bigg) \\
&= \quad \frac{1}{2}||a-b^0||^2 - \frac{1}{2}||a-b^s||^2 \\
&\leq \quad \frac{1}{2}||a-b^0||^2,
\end{aligned}$$

which completes the proof. □

Proposition E.3. *For $\Omega(a, a)^{r+1}$ defined in (E.5), we have*

$$\begin{aligned}
&\Omega(a^r, a)^{r+1} + \Omega(a^{r+1}, a)^r - \frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||a^r - a^{r-1}||^2 \\
\leq \quad &\frac{1}{2}\bigg(||a^r - a^{r-1}||^2 - ||a^{r+1} - a^r||^2\bigg).
\end{aligned} \tag{E.72}$$

Proof of Proposition E.3. By definition in (E.5) we have

$$\begin{aligned}
\Omega(a^r, a)^{r+1} \quad &= \quad -\frac{1}{2}||a^r - a^{r+1}||^2, \\
\Omega(a^{r+1}, a)^r \quad &= \quad \frac{1}{2}||a^{r+1} - a^{r-1}||^2 - \frac{1}{2}||a^{r+1} - a^r||^2.
\end{aligned} \tag{E.73}$$

Further note that

$$\begin{aligned}
||a^{r+1} - a^{r-1}||^2 \quad &= \quad ||a^{r+1} - a^r + a^r - a^{r-1}||^2 \\
&\leq \quad 2||a^{r+1} - a^r||^2 + 2||a^r - a^{r-1}||^2.
\end{aligned} \tag{E.74}$$

By applying the results in (E.73) and (E.74) we obtain

$$\begin{aligned}
& \Omega(a^r, a)^{r+1} + \Omega(a^{r+1}, a)^r - \frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||a^r - a^{r-1}||^2 \\
\leq \;& -\frac{1}{2}||a^r - a^{r+1}||^2 + \left[\frac{1}{2}\bigg(2||a^{r+1} - a^r||^2 + 2||a^r - a^{r-1}||^2\bigg) - \frac{1}{2}||a^{r+1} - a^r||^2\right] \\
& -\frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||a^r - a^{r-1}||^2 \\
= \;& \frac{1}{2}||a^r - a^{r-1}||^2 - \frac{1}{2}||a^{r+1} - a^r||^2,
\end{aligned}$$

which completes the proof. □

Proposition E.4. *For $\Omega(a, a)^{r+1}$ defined in (E.5), we have*

$$\begin{aligned}
& \Omega(a^r, a)^{r+1} + \Omega(a^{r+1}, a)^r - \frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||a^r - a^{r-1}||^2 \\
= \;& \frac{1}{2}||a^r - a^{r-1}||^2 - \frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2.
\end{aligned} \tag{E.75}$$

Proof of Proposition E.4. By definition in (E.5) we have

$$\begin{aligned}
\Omega(a^r, a)^{r+1} &= -\frac{1}{2}||a^r - a^{r+1}||^2, \\
\Omega(a^{r+1}, a)^r &= \frac{1}{2}||a^{r+1} - a^{r-1}||^2 - \frac{1}{2}||a^{r+1} - a^r||^2.
\end{aligned} \tag{E.76}$$

Note that

$$\begin{aligned}
||a^{r+1} - a^{r-1}||^2 &= ||(a^{r+1} - a^r) + (a^r - a^{r-1})||^2 \\
&= ||a^{r+1} - a^r||^2 + ||a^r - a^{r-1}||^2 + \left\langle a^{r+1} - a^r, a^r - a^{r-1}\right\rangle \\
&\quad + ||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2 - ||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2 \\
&= -||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2 + 2||a^{r+1} - a^r||^2 + 2||a^r - a^{r-1}||^2.
\end{aligned} \tag{E.77}$$

By applying the result in (E.77) we obtain

$$\begin{aligned}
& \Omega(a^{r+1}, a)^r \\
= \;& \frac{1}{2}\bigg(- ||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2 + 2||a^{r+1} - a^r||^2 + 2||a^r - a^{r-1}||^2\bigg) - \frac{1}{2}||a^{r+1} - a^r||^2 \\
= \;& -\frac{1}{2}||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2 + \frac{1}{2}||a^{r+1} - a^r||^2 + ||a^r - a^{r-1}||^2.
\end{aligned} \tag{E.78}$$

By applying the results in (E.76) and (E.78) we further obtain

$$
\begin{aligned}
& \Omega(a^r, a)^{r+1} + \Omega(a^{r+1}, a)^r - \frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||a^r - a^{r-1}||^2 \\
= \quad & -\frac{1}{2}||a^r - a^{r+1}||^2 + \Big(-\frac{1}{2}||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2 + \frac{1}{2}||a^{r+1} - a^r||^2 + ||a^r - a^{r-1}||^2 \Big) \\
& -\frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||a^r - a^{r-1}||^2 \\
= \quad & \frac{1}{2}||a^r - a^{r-1}||^2 - \frac{1}{2}||a^{r+1} - a^r||^2 - \frac{1}{2}||(a^{r+1} - a^r) - (a^r - a^{r-1})||^2,
\end{aligned}
$$

which completes the proof. □

References

Chen, M.-l. (2012). Open Government Data Project. `https://data.gov.tw/en/about`. Accessed: 2018-11-26.

Open Government Data (2012). Statistics on death causes (in chinese). `http://data.gov.tw/node/5965`. Accessed: 2018-11-26.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics 20*, 53–65.

Shefi, R. and M. Teboulle (2014). Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM Journal of Optimization 24*, 269–297.