Generalized degrees of freedom for Network Lasso and its influence on model selection

Koki Kirishima[1] Mineaki Ohishi[2] Ryoya Oda[3] Hirokazu Yanagihara[1,4]

February 19, 2026

Abstract

This paper considers fitting a linear regression model to each vertex of a graph with m vertices. For such models, Network Lasso has been proposed as a method to estimate the regression coefficients by minimizing the penalized residual sum of squares, where a penalty is added based on the norm of the difference in regression coefficients for adjacent vertices. In estimation using Network Lasso, it is possible to estimate the regression coefficients as equal for adjacent vertices. However, since the estimation results of the regression coefficients change when the weights between vertices or the tuning parameters vary, it is necessary to select the appropriate weights and tuning parameters. For this selection, methods based on minimizing model selection criteria, such as the C_p criterion, are often used. Since the C_p criterion is defined by adding twice the degrees of freedom to the minimum residual sum of squares, it is important to estimate the degrees of freedom. In this paper, the estimator of the degrees of freedom is derived for Network Lasso based on the concept of generalized degrees of freedom. Numerical simulations are used to describe how the estimator of generalized degrees of freedom influences model selection for Network Lasso.

1 Introduction

This paper deals with the generalized degrees of freedom for Network Lasso (Hallac *et al.* [5]), one of the estimation methods used for data with graphical structure. The underlying task is to fit the following linear regression model at each vertex of a graph with m vertices:

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{Z}_j\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_j \quad (j = 1, \ldots, m), \tag{1}$$

where $\boldsymbol{y}_j$ is the n_j-dimensional response variable vector, $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$ are the explanatory matrices of size $n_j \times k_L$ and $n_j \times k_G$, respectively, $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}$ are the regression coefficient vectors of dimensions k_L and k_G, respectively, $\boldsymbol{\varepsilon}_j$ is the n_j-dimensional error vector, and n_j is the number of samples. We assume that $\mathrm{rank}(\boldsymbol{X}_j) = k_L < n_j$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_m')' \sim N_n(\boldsymbol{0}_n, \sigma^2\boldsymbol{I}_n)$, where $n = \sum_{j=1}^m n_j$ is the sum of the sample size for each vertex, $\boldsymbol{0}_n$ is an n-dimensional vector of zeros, and $\boldsymbol{I}_n$ is the n-dimensional identity matrix. Additionally, for a matrix $\boldsymbol{A}$ and a vector $\boldsymbol{a}$, $\boldsymbol{A}'$ and $\boldsymbol{a}'$ denote the transpose of $\boldsymbol{A}$ and $\boldsymbol{a}$, respectively. In (1), $\boldsymbol{\beta}_j$ varies for each vertex, but $\boldsymbol{\gamma}$ is common to all vertices. The term for $\boldsymbol{\gamma}$ is not included in Hallac *et al.* [5]; however, in this paper, we consider a more general case that includes the term for $\boldsymbol{\gamma}$. In Network Lasso, $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}$ in (1) are estimated by minimizing the following penalized residual sum of squares:

$$\mathrm{PRSS}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{W}, \lambda) = \frac{1}{2}\sum_{j=1}^m \|\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}_j - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2 + \frac{\lambda}{2}\sum_{j=1}^m \sum_{\ell \in D_j} w_{j\ell}\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_\ell\|, \tag{2}$$

[1]Osaka Central Advanced Mathematical Institute, Osaka Metropolitan University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan
[2]Center for Data-driven Science and Artificial Intelligence, Tohoku University, Kawauchi 41, Aoba-ku, Sendai 980-8576, Japan
[3]Graduate School of Advanced Science and Engineering, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526, Japan
[4]Research & Development Center, Osaka Medical and Pharmaceutical University, 2-7 Daigaku-machi, Takatsuki, Osaka 569-8686, Japan

where $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_m)'$, $D_j \subseteq \{1, \ldots, m\} \backslash \{j\}$ is the set of vertices adjacent to the j-th vertex, $\boldsymbol{W} = (w_{j\ell})$ is the weight matrix, and λ is a non-negative tuning parameter. Assume that D_j is symmetric, that is, $\ell \in D_j \Leftrightarrow j \in D_\ell$. Also, the weight matrix $\boldsymbol{W}$ is an $m \times m$ symmetric matrix with all non-negative elements, and $w_{j\ell} = 0$ when $\ell \notin D_j$. For the weight $\boldsymbol{W}$, simple choices like $w_{j\ell} = 1$ when $\ell \in D_j$ and $w_{j\ell} = 0$ otherwise, or weights based on adaptive Lasso (Zou [14]) defined by (3) using the least squares (LS) estimates $\boldsymbol{\beta}_j$ and a non-negative tuning parameter δ, such as $\boldsymbol{W}_{\mathrm{AL}}(\delta)$, can be considered.

$$\boldsymbol{W}_{\mathrm{AL}}(\delta) = \left(w_{j\ell}^{\mathrm{AL}}(\delta)\right), \quad w_{j\ell}^{\mathrm{AL}}(\delta) = \begin{cases} \|\hat{\boldsymbol{\beta}}_{j,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-\delta} & (\ell \in D_j) \\ 0 & (\text{otherwise}) \end{cases}. \tag{3}$$

$\hat{\boldsymbol{\beta}}_{j,\mathrm{LS}}$ in (3) indicates the LS estimates $\boldsymbol{\beta}_j$ defined by the following:

$$\hat{\boldsymbol{\beta}}_{j,\mathrm{LS}} = (\boldsymbol{X}'_j \boldsymbol{X}_j)^{-1} \boldsymbol{X}'_j (\boldsymbol{y}_j - \boldsymbol{Z}_j \hat{\boldsymbol{\gamma}}_{\mathrm{LS}}), \quad \hat{\boldsymbol{\gamma}}_{\mathrm{LS}} = \{\boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1} \boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}, \tag{4}$$

where $\boldsymbol{y} = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_m)'$, $\boldsymbol{X} = \boldsymbol{X}_1 \oplus \cdots \oplus \boldsymbol{X}_m$, $\boldsymbol{Z} = (\boldsymbol{Z}'_1, \ldots, \boldsymbol{Z}'_m)'$, and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Moreover, for an $a_1 \times a_2$ matrix $\boldsymbol{A}$ and a $b_1 \times b_2$ matrix $\boldsymbol{B}$, the direct sum of matrices $\boldsymbol{A} \oplus \boldsymbol{B}$ is defined as follows (Horn and Johnson [6]):

$$\boldsymbol{A} \oplus \boldsymbol{B} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{O}_{a_1,b_2} \\ \boldsymbol{O}_{b_1,a_2} & \boldsymbol{B} \end{pmatrix},$$

where $\boldsymbol{O}_{p,q}$ is a $p \times q$ matrix of zeros. To calculate $\hat{\boldsymbol{\gamma}}_{\mathrm{LS}}$ in (4), we assume that $\boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}$ is invertible. Let $\hat{\boldsymbol{\beta}}_{j,(\boldsymbol{W},\lambda)}$ and $\hat{\boldsymbol{\gamma}}_{(\boldsymbol{W},\lambda)}$ be the values of $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}$ that minimize (2). In estimation using Network Lasso, it is possible to have estimates such that $\hat{\boldsymbol{\beta}}_{j,(\boldsymbol{W},\lambda)} = \hat{\boldsymbol{\beta}}_{\ell,(\boldsymbol{W},\lambda)}$.

Network Lasso is an estimation method that extends Fused Lasso (Tibshirani *et al.* [11]) and its extension, Group Fused Lasso (Bleakley and Vert [2]). By setting $k_G = 0$ and defining D_j and $w_{j\ell}$ as given in (5), Network Lasso coincides with Group Fused Lasso. In addition to the Group Fused Lasso setting, setting $k_L = 1$ makes Network Lasso coincide with Fused Lasso.

$$D_j = \begin{cases} \{2\} & (j = 1) \\ \{j-1, j+1\} & (j = 2, \ldots, m-1) \\ \{m-1\} & (j = m) \end{cases}, \quad w_{j\ell} = \begin{cases} 1 & (\ell \in D_j) \\ 0 & (\text{otherwise}) \end{cases}. \tag{5}$$

Fused Lasso and Group Fused Lasso can be optimized within the framework of ordinary Lasso (Tibshirani *et al.* [10]) or Group Lasso (Yuan and Lin [13]) through variable transformation. Similarly to Group Fused Lasso, Network Lasso can be optimized within the framework of Group Lasso. However, because the adjacency relationships of the vertices are one-to-many, a unique solution cannot be determined. An optimization method using ADMM (Boyd *et al.* [3]) has been proposed by Hallac *et al.* [5]. Additionally, a faster optimization method using coordinate descent has been proposed by Ohishi *et al.* [8].

In estimation using Network Lasso, $\hat{\boldsymbol{\beta}}_{j,(\boldsymbol{W},\lambda)}$ and $\hat{\boldsymbol{\gamma}}_{(\boldsymbol{W},\lambda)}$ change when $\boldsymbol{W}$ and λ are varied. Therefore, it is necessary to select $\boldsymbol{W}$ and λ (particularly, when $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{AL}}(\delta)$, choosing δ corresponds to selecting the weights). A common method for selecting $\boldsymbol{W}$ and λ is to choose them by minimizing a model selection criterion. As a model selection criterion, the C_p criterion (Mallows [7]) defined as (6) is often used.

$$C_p(\boldsymbol{W}, \lambda) = \frac{n}{s^2}\hat{\sigma}^2(\boldsymbol{W}, \lambda) + 2\sum_{j=1}^{m}\sum_{i=1}^{n_j} \frac{Cov(y_{ji}, \hat{y}_{ji,(\boldsymbol{W},\lambda)})}{\sigma^2}, \tag{6}$$

where y_{ji} and $\hat{y}_{ji,(\boldsymbol{W},\lambda)}$ are the i-th components of $\boldsymbol{y}_j$ and $\hat{\boldsymbol{y}}_{j,(\boldsymbol{W},\lambda)} = \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_{j,(\boldsymbol{W},\lambda)} + \boldsymbol{Z}_j\hat{\boldsymbol{\gamma}}_{(\boldsymbol{W},\lambda)}$, respectively, and $\hat{\sigma}^2(\boldsymbol{W}, \lambda)$ and s^2 are defined as follows.

$$\hat{\sigma}^2(\boldsymbol{W}, \lambda) = \frac{\sum_{j=1}^{m} \|\boldsymbol{y}_j - \hat{\boldsymbol{y}}_{j,(\boldsymbol{W},\lambda)}\|^2}{n}, \quad s^2 = \frac{\sum_{j=1}^{m} \|\boldsymbol{y}_j - \hat{\boldsymbol{y}}_{j,\mathrm{LS}}\|^2}{n - mk_L - k_G}.$$

Additionally, $\hat{\boldsymbol{y}}_{j,\mathrm{LS}} = \boldsymbol{X}_j \hat{\boldsymbol{\beta}}_{j,\mathrm{LS}} + \boldsymbol{Z}_j \hat{\boldsymbol{\gamma}}_{\mathrm{LS}}$. Stein [9] showed that when $\hat{\boldsymbol{y}}_{(\boldsymbol{W},\lambda)} = (\hat{\boldsymbol{y}}'_{1,(\boldsymbol{W},\lambda)}, \ldots, \hat{\boldsymbol{y}}'_{m,(\boldsymbol{W},\lambda)})'$ is almost everywhere differentiable with respect to $\boldsymbol{y}$, under $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n)$, the following holds:

$$\sum_{j=1}^{m} \sum_{i=1}^{n_j} \frac{Cov(y_{ji}, \hat{y}_{ji,(\boldsymbol{W},\lambda)})}{\sigma^2} = E\left[\sum_{j=1}^{m} \sum_{i=1}^{n_j} \frac{\partial \hat{y}_{ji,(\boldsymbol{W},\lambda)}}{\partial y_{ji}}\right] = E\left[\sum_{j=1}^{m} \mathrm{tr}\left(\frac{\partial \hat{\boldsymbol{y}}'_{j,(\boldsymbol{W},\lambda)}}{\partial \boldsymbol{y}_j}\right)\right]. \tag{7}$$

The rightmost side of (7) is called the generalized degrees of freedom (Efron [4], Ye [12]). Here, we denote this by $\mathrm{df}(\boldsymbol{W}, \lambda)$. Therefore, (6) can be expressed as

$$C_p(\boldsymbol{W}, \lambda) = \frac{n}{s^2} \hat{\sigma}^2(\boldsymbol{W}, \lambda) + 2\mathrm{df}(\boldsymbol{W}, \lambda). \tag{8}$$

Thus, it is important to obtain the estimator $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ of $\mathrm{df}(\boldsymbol{W}, \lambda)$ to calculate model selection criteria such as the C_p criterion. As estimators of the degrees of freedom, the number of explanatory variables in ordinary linear regression or the number of nonzero regression coefficients in standard Lasso (Zou *et al.* [15]) are often used. Moreover, the degrees of freedom for Group Lasso when using weights based on adaptive Lasso have been proposed in Bernardi *et al.* [1]. However, the degrees of freedom for Network Lasso has not yet been derived. This is because it becomes complicated since Network Lasso includes a penalty on the norms of the differences between regression coefficient vectors. Furthermore, since the weights are often defined as a function of $\boldsymbol{y}$, such as $\boldsymbol{W}_{\mathrm{AL}}(\delta)$, they affect the calculation of the generalized degrees of freedom. Therefore, the generalized degrees of freedom derived by treating the weights as constants with respect to $\boldsymbol{y}$ may be significantly biased. As a result, this may affect the outcomes of model selection. In this paper, we derive an estimator $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ of the generalized degrees of freedom $\mathrm{df}(\boldsymbol{W}, \lambda)$ for Network Lasso, assuming that $\hat{\boldsymbol{y}}_{(\boldsymbol{W},\lambda)}$ is almost everywhere differentiable with respect to $\boldsymbol{y}$.

The remainder of the paper is organized as follows: In Section 2, we prepare the notation necessary for deriving the generalized degrees of freedom. In Section 3, we derive an estimator of the generalized degrees of freedom for Network Lasso. In Subsection 3.1, we derive an estimator of the generalized degrees of freedom for a general $\boldsymbol{W}$, and in Subsection 3.2, we derive the generalized degrees of freedom in a more specific form when the weight is given as $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{AL}}(\delta)$. In Section 4, we conduct numerical experiments for the case when $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{AL}}(\delta)$. Proofs of the mathematical formulas are provided in the Appendix.

2 Preliminaries

This section gives the notation used for deriving the generalized degrees of freedom for Network Lasso and provides proof of several related properties. To simplify the notation, we omit $\boldsymbol{W}$ and λ except where emphasis is needed. That is, we write $\hat{\boldsymbol{\beta}}_{j,(\boldsymbol{W},\lambda)}$ and $\hat{\boldsymbol{\gamma}}_{(\boldsymbol{W},\lambda)}$ as $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\gamma}}$, respectively.

Let $\hat{\boldsymbol{\xi}}_1 = \hat{\boldsymbol{\xi}}_{1,(\boldsymbol{W},\lambda)}, \ldots, \hat{\boldsymbol{\xi}}_{\hat{t}} = \hat{\boldsymbol{\xi}}_{\hat{t},(\boldsymbol{W},\lambda)}$ be the distinct vector of $\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_m$, where $\hat{t} = \hat{t}_{(\boldsymbol{W},\lambda)}$ is an integer satisfying $1 \le \hat{t} \le m$. In addition, let $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_{(\boldsymbol{W},\lambda)} = (\hat{\boldsymbol{\xi}}'_1, \ldots, \hat{\boldsymbol{\xi}}'_{\hat{t}})'$. Then, for $s = 1, \ldots, \hat{t}$, we can define the sets $E_s = E_{s,(\boldsymbol{W},\lambda)}$ and $F_s = F_{s,(\boldsymbol{W},\lambda)}$ as follows:

$$E_s = \left\{ j \in \{1, \ldots, m\} \;\middle|\; \hat{\boldsymbol{\beta}}_j = \hat{\boldsymbol{\xi}}_s \right\}, \quad F_s = \left\{ u \in \{1, \ldots, \hat{t}\} \setminus \{s\} \;\middle|\; \bigcup_{j \in E_s} D_j \cap E_u \neq \emptyset \right\}.$$

E_s is the set of vertices for which the regression coefficients are estimated to be equal, and the vertices within the same E_s can be considered as being connected. Here, we refer to E_s as the *s-th connected vertex*. F_s is the set of indices of connected vertices adjacent to the s-th connected vertex. E_s satisfies the following properties: (1) $E_s \neq \emptyset$, (2) $\bigcup_{s=1}^{\hat{t}} E_s = \{1, \ldots, m\}$, (3) $E_s \cap E_u = \emptyset$ $(s \neq u)$. Furthermore, F_s satisfies $u \in F_s \Leftrightarrow s \in F_u$. Using E_s and F_s, a graph with m vertices can be replaced by a graph with $\hat{t}$ vertices. Therefore, under $\boldsymbol{\beta}_j = \boldsymbol{\xi}_s = (\boldsymbol{e}'_s \otimes \boldsymbol{I}_{k_L})\boldsymbol{\xi}$ $(j \in E_s)$, (2) can be rewritten to the following $f(\boldsymbol{\xi}, \boldsymbol{\gamma})$ (the proof is given in Appendix A.1):

$$f(\boldsymbol{\xi}, \boldsymbol{\gamma}) = \frac{1}{2} \sum_{s=1}^{\hat{t}} \sum_{j \in E_s} \|\boldsymbol{y}_j - (\boldsymbol{e}'_s \otimes \boldsymbol{X}_j)\boldsymbol{\xi} - \boldsymbol{Z}_j \boldsymbol{\gamma}\|^2 + \frac{\lambda}{2} \sum_{s=1}^{\hat{t}} \sum_{u \in F_s} v_{su} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|, \tag{9}$$

where $\boldsymbol{e}_s$ is a $\hat{t}$-dimensional vector with the s-th component being 1 and other components being 0, $v_{su} = \sum_{j \in E_s} \sum_{\ell \in E_u \cap D_j} w_{j\ell}$, and $\otimes$ is the Kronecker product. Note that $v_{su} = v_{us}$ (the proof is given in Appendix A.1). Moreover, note that since $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ minimize (2), $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\gamma}}$ also minimize $f(\boldsymbol{\xi}, \boldsymbol{\gamma})$.

We define the vectors $\boldsymbol{g} = \boldsymbol{g}_{(\boldsymbol{W},\lambda)}$ and $\boldsymbol{h} = \boldsymbol{h}_{(\boldsymbol{W},\lambda)}$ as follows (the proof is given in Appendix A.2):

$$\begin{aligned} \boldsymbol{g} &= \frac{\partial}{\partial \boldsymbol{\xi}} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{A}\hat{\boldsymbol{\xi}} + \boldsymbol{B}\hat{\boldsymbol{\gamma}} - \boldsymbol{c} + \lambda \boldsymbol{\theta}, \\ \boldsymbol{h} &= \frac{\partial}{\partial \boldsymbol{\gamma}} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}'\hat{\boldsymbol{\xi}} + \boldsymbol{Z}'\boldsymbol{Z}\hat{\boldsymbol{\gamma}} - \boldsymbol{Z}'\boldsymbol{y}. \end{aligned} \tag{10}$$

Here, $\boldsymbol{A} = \boldsymbol{A}_{(\boldsymbol{W},\lambda)}$, $\boldsymbol{B} = \boldsymbol{B}_{(\boldsymbol{W},\lambda)}$, $\boldsymbol{c} = \boldsymbol{c}_{(\boldsymbol{W},\lambda)}$, and $\boldsymbol{\theta} = \boldsymbol{\theta}_{(\boldsymbol{W},\lambda)}$ are matrices or vectors defined as follows:

$$\begin{aligned} &\boldsymbol{A} = \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s \boldsymbol{e}_s' \otimes \sum_{j \in E_s} \boldsymbol{X}_j' \boldsymbol{X}_j \right), \quad \boldsymbol{B} = \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s \otimes \sum_{j \in E_s} \boldsymbol{X}_j' \boldsymbol{Z}_j \right), \quad \boldsymbol{c} = \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s \otimes \sum_{j \in E_s} \boldsymbol{X}_j' \boldsymbol{y}_j \right), \\ &\boldsymbol{\theta} = \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s \otimes \sum_{u \in F_s} v_{su} \boldsymbol{\theta}_{su} \right), \quad \boldsymbol{\theta}_{su} = \|\hat{\boldsymbol{\xi}}_s - \hat{\boldsymbol{\xi}}_u\|^{-1} (\hat{\boldsymbol{\xi}}_s - \hat{\boldsymbol{\xi}}_u). \end{aligned} \tag{11}$$

We note that $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\gamma}}$ minimize $f(\boldsymbol{\xi}, \boldsymbol{\gamma})$ and that $\hat{\boldsymbol{\xi}}_s \neq \hat{\boldsymbol{\xi}}_u$ $(s \neq u)$. Hence, $f(\boldsymbol{\xi}, \boldsymbol{\gamma})$ is minimized at a differentiable point and $\boldsymbol{g} = \boldsymbol{0}_{\hat{t}k_L}$, $\boldsymbol{h} = \boldsymbol{0}_{k_G}$.

Additionally, we define the matrices $\boldsymbol{G} = \boldsymbol{G}_{(\boldsymbol{W},\lambda)}$, $\boldsymbol{H} = \boldsymbol{H}_{(\boldsymbol{W},\lambda)}$, $\boldsymbol{J} = \boldsymbol{J}_{(\boldsymbol{W},\lambda)}$, and $\boldsymbol{K} = \boldsymbol{K}_{(\boldsymbol{W},\lambda)}$ as follows (the proof is given in Appendix A.3):

$$\begin{aligned} \boldsymbol{G} &= \frac{\partial^2}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}'} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{A} + \lambda \boldsymbol{L}, \quad & \boldsymbol{H} &= \frac{\partial^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{Z}'\boldsymbol{Z}, \\ \boldsymbol{J} &= \frac{\partial^2}{\partial \boldsymbol{\xi} \partial \boldsymbol{\gamma}'} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}, & \boldsymbol{K} &= \frac{\partial^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\xi}'} f(\boldsymbol{\xi}, \boldsymbol{\gamma}) \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}', \end{aligned} \tag{12}$$

where $\boldsymbol{L} = \boldsymbol{L}_{(\boldsymbol{W},\lambda)}$ is a matrix defined using $\boldsymbol{\Theta}_{pq} = \boldsymbol{I}_{k_L} - \boldsymbol{\theta}_{pq} \boldsymbol{\theta}_{pq}'$ as follows.

$$\boldsymbol{L} = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \left(\boldsymbol{e}_p \boldsymbol{e}_q' \otimes \boldsymbol{L}_{pq} \right), \quad \boldsymbol{L}_{pq} = \begin{cases} \sum_{u \in F_p} v_{pu} \|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_u\|^{-1} \boldsymbol{\Theta}_{pu} & (q = p) \\ -v_{pq} \|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_q\|^{-1} \boldsymbol{\Theta}_{pu} & (q \in F_p) \\ \boldsymbol{O}_{k_L, k_L} & (otherwise) \end{cases}. \tag{13}$$

Note that $\boldsymbol{K} = \boldsymbol{B}' = \boldsymbol{J}'$. $\boldsymbol{G}$ satisfies the following lemma (the proof is given in Appendix A.4):

Lemma 1. $\boldsymbol{A}$ is a positive definite matrix, and $\boldsymbol{L}$ is a positive semi-definite matrix. Therefore, $\boldsymbol{G}$ is a positive definite matrix and is an invertible matrix.

Let the matrix $\boldsymbol{M} = \boldsymbol{M}_{(\boldsymbol{W},\lambda)}$ be defined as follows:

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{G} & \boldsymbol{J} \\ \boldsymbol{K} & \boldsymbol{H} \end{pmatrix} = \begin{pmatrix} \boldsymbol{G} & \boldsymbol{J} \\ \boldsymbol{J}' & \boldsymbol{H} \end{pmatrix}.$$

Using Lemma 1, we can show the following lemma regarding $\boldsymbol{M}$ (the proof is given in Appendix A.5):

Lemma 2. $\boldsymbol{M}$ is a positive semi-definite matrix.

Since $\boldsymbol{G}$ is invertible, we can define the matrix $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$. This matrix is called the Schur complement of $\boldsymbol{M}$ with respect to $\boldsymbol{G}$. Using Lemma 2, we can show the following lemma (the proof is given in Appendix A.6):

Lemma 3. $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$ is a positive semi-definite matrix.

Since $\boldsymbol{G}$ is invertible, $\boldsymbol{M}$ is also invertible if $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$ is invertible. Then, the inverse matrix of $\boldsymbol{M}$, denoted by $\boldsymbol{M}^{-1}$, is obtained as follows:

$$\boldsymbol{M}^{-1} = \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{pmatrix}, \quad \begin{cases} \boldsymbol{M}_{11} = \boldsymbol{M}_{11,(\boldsymbol{W},\lambda)} = \boldsymbol{G}^{-1} + \boldsymbol{G}^{-1}\boldsymbol{J}(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{J}'\boldsymbol{G}^{-1} \\ \boldsymbol{M}_{12} = \boldsymbol{M}_{12,(\boldsymbol{W},\lambda)} = -\boldsymbol{G}^{-1}\boldsymbol{J}(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1} \\ \boldsymbol{M}_{21} = \boldsymbol{M}_{21,(\boldsymbol{W},\lambda)} = -(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{J}'\boldsymbol{G}^{-1} \\ \boldsymbol{M}_{22} = \boldsymbol{M}_{22,(\boldsymbol{W},\lambda)} = (\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1} \end{cases}.$$

From the positive definiteness of $\boldsymbol{G}$ and the semi-positive definiteness of $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$, the following lemma regarding $\boldsymbol{M}_{11}$ holds (the proof is given in Appendix A.7):

Lemma 4. $\boldsymbol{M}_{11}$ is a positive semi-definite matrix.

From here, we assume that $\boldsymbol{M}$ is invertible.

3 Main result

In this section, we derive the estimator of the generalized degrees of freedom for Network Lasso.

The generalized degrees of freedom value for Network Lasso is defined as follows using $\hat{\boldsymbol{y}}_j = \hat{\boldsymbol{y}}_{j,(\boldsymbol{W},\lambda)}$:

$$\mathrm{df}(\boldsymbol{W}, \lambda) = E\left[\sum_{j=1}^{m} \mathrm{tr}\left(\frac{\partial \hat{\boldsymbol{y}}_j'}{\partial \boldsymbol{y}_j}\right)\right].$$

Therefore, the estimator $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ of $\mathrm{df}(\boldsymbol{W}, \lambda)$ is obtained as follows:

$$\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda) = \sum_{j=1}^{m} \mathrm{tr}\left(\frac{\partial \hat{\boldsymbol{y}}_j'}{\partial \boldsymbol{y}_j}\right) = \sum_{j=1}^{m} \mathrm{tr}\left(\frac{\partial \hat{\boldsymbol{\beta}}_j'}{\partial \boldsymbol{y}_j}\boldsymbol{X}_j' + \frac{\partial \hat{\boldsymbol{\gamma}}'}{\partial \boldsymbol{y}_j}\boldsymbol{Z}_j'\right). \tag{14}$$

Since $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', \ldots, \hat{\boldsymbol{\beta}}_m')'$ and $\hat{\boldsymbol{\gamma}}$ minimize (2), if (2) is differentiable, then the following holds:

$$\frac{\partial}{\partial \boldsymbol{\beta}}\mathrm{PRSS}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{W}, \lambda)\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{0}_{mk_L}, \quad \frac{\partial}{\partial \boldsymbol{\gamma}}\mathrm{PRSS}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{W}, \lambda)\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{0}_{k_G}. \tag{15}$$

Using (15), we can obtain $\partial\hat{\boldsymbol{\beta}}'/\partial\boldsymbol{y}_j$ and $\partial\hat{\boldsymbol{\gamma}}'/\partial\boldsymbol{y}_j$, and thus obtain $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$. However, if there exist $j \in \{1, \ldots, m\}$ and $\ell \in D_j$ such that $\hat{\boldsymbol{\beta}}_j = \hat{\boldsymbol{\beta}}_\ell$, then (2) is minimized at a non-differentiable point. In this case, it is difficult to obtain $\partial\hat{\boldsymbol{\beta}}'/\partial\boldsymbol{y}_j$ and $\partial\hat{\boldsymbol{\gamma}}'/\partial\boldsymbol{y}_j$. To overcome this issue, we use the rewriting of (2) into (9) from Section 2. In Subsection 3.1, we derive $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ by obtaining $\partial\hat{\boldsymbol{\xi}}'/\partial\boldsymbol{y}_j$ and $\partial\hat{\boldsymbol{\gamma}}'/\partial\boldsymbol{y}_j$. Furthermore, in Subsection 3.2, we derive the more specific form of $\widehat{\mathrm{df}}(\boldsymbol{W}_{\mathrm{AL}}(\delta), \lambda)$ in the case of $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{AL}}(\delta)$.

3.1 Generalized degrees of freedom for Network Lasso

Noting that for $j \in E_s$, $\boldsymbol{X}_j\hat{\boldsymbol{\beta}}_j = (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j)\hat{\boldsymbol{\xi}}$, (14) can be rewritten as follows:

$$\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda) = \sum_{s=1}^{\hat{t}} \sum_{j \in E_s} \mathrm{tr}\left(\frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}(\boldsymbol{e}_s \otimes \boldsymbol{X}_j') + \frac{\partial \hat{\boldsymbol{\gamma}}'}{\partial \boldsymbol{y}_j}\boldsymbol{Z}_j'\right). \tag{16}$$

Therefore, $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ can be obtained by finding $\partial\hat{\boldsymbol{\xi}}'/\partial\boldsymbol{y}_j$ and $\partial\hat{\boldsymbol{\gamma}}/\partial\boldsymbol{y}_j$. The following lemma provides the results of differentiating $\boldsymbol{g}$ and $\boldsymbol{h}$ with respect to $\boldsymbol{y}_j$ (the proof is given in Appendix B.1):

Lemma 5. For $j \in E_s$ $(s = 1, \ldots, \hat{t})$, $\partial \boldsymbol{g}'/\partial \boldsymbol{y}_j$ and $\partial \boldsymbol{h}'/\partial \boldsymbol{y}_j$ are given by the following:

$$\begin{aligned}\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{y}_j} &= \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}\boldsymbol{G} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j}\boldsymbol{J}' - (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j) + \lambda \boldsymbol{\Phi}_j, \\ \frac{\partial \boldsymbol{h}'}{\partial \boldsymbol{y}_j} &= \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}\boldsymbol{J} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j}\boldsymbol{H} - \boldsymbol{Z}_j,\end{aligned}$$

where $\boldsymbol{\Phi}_j = \boldsymbol{\Phi}_{j,(\boldsymbol{W},\lambda)}$ is the matrix given by the following:

$$\boldsymbol{\Phi}_j = \sum_{p=1}^{\hat{t}} \left(\boldsymbol{e}_p' \otimes \sum_{u \in F_p} \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j} \boldsymbol{\theta}_{pu}' \right). \tag{17}$$

Since $\boldsymbol{g} = \boldsymbol{0}_{\hat{t}k_L}$ and $\boldsymbol{h} = \boldsymbol{0}_{k_G}$, we have $\partial \boldsymbol{g}'/\partial \boldsymbol{y}_j = \boldsymbol{O}_{n_j,\hat{t}k_L}$ and $\partial \boldsymbol{h}'/\partial \boldsymbol{y}_j = \boldsymbol{O}_{n_j,k_G}$. Therefore, from Lemma 5, we obtain the following lemma (the proof is given in Appendix B.2):

Lemma 6. For $j \in E_s$ $(s = 1, \ldots, \hat{t})$, $\partial \hat{\boldsymbol{\xi}}'/\partial \boldsymbol{y}_j$ and $\partial \hat{\gamma}'/\partial \boldsymbol{y}_j$ are given by the following:

$$\left(\frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}, \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \right) = (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j - \lambda \boldsymbol{\Phi}_j, \boldsymbol{Z}_j)\, \boldsymbol{M}^{-1}.$$

By substituting the result of Lemma 6 into (16), we obtain the following theorem (the proof is given in Appendix B.3):

Theorem 1. The estimator of the generalized degrees of freedom for Network Lasso, $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$, is given by

$$\begin{aligned}\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda) = \hat{t}_{(\boldsymbol{W},\lambda)} k_L + k_G &- \lambda \mathrm{tr}\left(\boldsymbol{M}_{11,(\boldsymbol{W},\lambda)} \boldsymbol{L}_{(\boldsymbol{W},\lambda)} \right) \\ &- \lambda \mathrm{tr}\left(\boldsymbol{M}_{11,(\boldsymbol{W},\lambda)} \boldsymbol{R}_{(\boldsymbol{W},\lambda)} \right) - \lambda \mathrm{tr}\left(\boldsymbol{M}_{12,(\boldsymbol{W},\lambda)} \boldsymbol{U}_{(\boldsymbol{W},\lambda)} \right),\end{aligned}$$

where $\boldsymbol{R} = \boldsymbol{R}_{(\boldsymbol{W},\lambda)}$ and $\boldsymbol{U} = \boldsymbol{U}_{(\boldsymbol{W},\lambda)}$ are matrices defined by the following:

$$\begin{aligned}\boldsymbol{R} &= \sum_{s=1}^{\hat{t}} \sum_{p=1}^{\hat{t}} \left(\boldsymbol{e}_s \boldsymbol{e}_p' \otimes \boldsymbol{R}_{sp} \right), \quad \boldsymbol{R}_{sp} = \boldsymbol{R}_{sp,(\boldsymbol{W},\lambda)} = \sum_{u \in F_p} \sum_{j \in E_s} \boldsymbol{X}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j} \boldsymbol{\theta}_{pu}', \\ \boldsymbol{U} &= \sum_{p=1}^{\hat{t}} \left(\boldsymbol{e}_p' \otimes \boldsymbol{U}_p \right), \qquad \boldsymbol{U}_p = \boldsymbol{U}_{p,(\boldsymbol{W},\lambda)} = \sum_{u \in F_p} \sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j} \boldsymbol{\theta}_{pu}'.\end{aligned} \tag{18}$$

From Theorem 1, $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ can be obtained by finding $\partial v_{pu}/\partial \boldsymbol{y}_j$, i.e., $\partial w_{r\ell}/\partial \boldsymbol{y}_j$. In the next section, we present the results for the case where $\boldsymbol{W}$ is the weight based on adaptive Lasso, $\boldsymbol{W}_{\mathrm{AL}}(\delta)$.

3.2 Generalized degrees of freedom when the weights are based on the adaptive Lasso

When $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{AL}}(\delta)$, then $\boldsymbol{W}$ is determined solely by δ. Therefore, let $\nu_{\mathrm{AL}}(\delta, \lambda) = \widehat{\mathrm{df}}(\boldsymbol{W}_{\mathrm{AL}}(\delta), \lambda)$, and derive $\nu_{\mathrm{AL}}(\delta, \lambda)$. By explicitly finding $\boldsymbol{R}$ and $\boldsymbol{U}$ in (18), we can obtain $\nu_{\mathrm{AL}}(\delta, \lambda)$. In other words, one needs to find $\partial w_{r\ell}/\partial \boldsymbol{y}_j$. The following lemma provides the result for $\partial w_{r\ell}^{\mathrm{AL}}(\delta)/\partial \boldsymbol{y}_j$ (the proof is given in Appendix B.4):

Lemma 7. For $r \in \{1, \ldots, m\}$ and $\ell \in D_r$, the following holds:

$$\frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j} = \begin{cases} -\delta w_{r\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}} \left(\dfrac{\partial \hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} - \dfrac{\partial \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} \right) \boldsymbol{a}_{r\ell} & (\delta > 0) \\ \boldsymbol{0}_{n_j} & (\delta = 0) \end{cases},$$

where $\boldsymbol{a}_{r\ell} = \|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-1} (\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}})$.

From Lemma 7, $\partial w_{r\ell}^{\mathrm{AL}}(\delta)/\partial \boldsymbol{y}_j$ is expressed using $\partial\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}/\partial \boldsymbol{y}_j$ and $\partial\hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}/\partial \boldsymbol{y}_j$. The following lemma provides the results for $\partial\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}/\partial \boldsymbol{y}_j$ and $\partial\hat{\gamma}_{\mathrm{LS}}/\partial \boldsymbol{y}_j$ (the proof is given in Appendix B.5):

Lemma 8. For $j, r \in \{1, \ldots, m\}$, the following holds:

$$\frac{\partial\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} = \begin{cases} \boldsymbol{X}_j(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1} - \dfrac{\partial\hat{\gamma}_{\mathrm{LS}}'}{\partial \boldsymbol{y}_j}\boldsymbol{Z}_j'\boldsymbol{X}_j(\boldsymbol{X}_j'\boldsymbol{X}_j)^{-1} & (r = j) \\ -\dfrac{\partial\hat{\gamma}_{\mathrm{LS}}'}{\partial \boldsymbol{y}_j}\boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} & (r \neq j) \end{cases},$$

$$\frac{\partial\hat{\gamma}_{\mathrm{LS}}'}{\partial \boldsymbol{y}_j} = (\boldsymbol{I}_{n_j} - \boldsymbol{P}_{\boldsymbol{X}_j})\boldsymbol{Z}_j\{\boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1}.$$

By substituting the result from Lemma 8 into Lemma 7, we can obtain $\partial w_{r\ell}^{\mathrm{AL}}(\delta)/\partial \boldsymbol{y}_j$. Furthermore, by substituting this result into (18), we have the following two lemmas (the proofs of Lemma 9 and Lemma 10 are given in Appendix B.6 and Appendix B.7, respectively):

Lemma 9. Let $\delta \geq 0$ and $\lambda \geq 0$. We define the matrix $\boldsymbol{R}_{\mathrm{AL}} = \boldsymbol{R}_{\mathrm{AL}}(\delta, \lambda)$ as follows:

$$\boldsymbol{R}_{\mathrm{AL}} = \sum_{s=1}^{\hat{t}}\sum_{p=1}^{\hat{t}}\left(\boldsymbol{e}_s\boldsymbol{e}_p' \otimes \boldsymbol{R}_{sp}^{\mathrm{AL}}\right),$$
$$\boldsymbol{R}_{sp}^{\mathrm{AL}} = \boldsymbol{R}_{sp}^{\mathrm{AL}}(\delta, \lambda) = \begin{cases} \displaystyle\sum_{u\in F_s}\sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{j\ell}\boldsymbol{\theta}_{su}' & ((p = s) \wedge (\delta \neq 0)) \\ \displaystyle -\sum_{j\in E_s}\sum_{\ell\in E_p\cap D_j} w_{j\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{j\ell}\boldsymbol{\theta}_{sp}' & ((p \in F_s) \wedge (\delta \neq 0)) \\ \boldsymbol{O}_{k_L,k_L} & (otherwise) \end{cases}. \tag{19}$$

Then, we have $\boldsymbol{R}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} = -\delta\boldsymbol{R}_{\mathrm{AL}}(\delta, \lambda)$.

Lemma 10. Let $\delta \geq 0$ and $\lambda \geq 0$. Then, we have $\boldsymbol{U}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} = \boldsymbol{O}_{k_G,\hat{t}k_L}$.

We have the following theorem by substituting the results of Lemma 9 and Lemma 10 into Theorem 1 (the proof is given in Appendix B.8):

Theorem 2. For $\delta \geq 0$ and $\lambda \geq 0$, $\nu_{\mathrm{AL}}(\delta, \lambda)$ is given by the following:

$$\begin{aligned}\nu_{\mathrm{AL}}(\delta, \lambda) = {}& \hat{t}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}k_L + k_G \\ & - \lambda\mathrm{tr}\left(\boldsymbol{M}_{11,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}\boldsymbol{L}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}\right) + \lambda\delta\mathrm{tr}\left(\boldsymbol{M}_{11,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}\boldsymbol{R}_{\mathrm{AL}}(\delta, \lambda)\right).\end{aligned}$$

From Theorem 2, we see that $\nu_{\mathrm{AL}}(\delta, \lambda)$ consists of four significant terms. Here, we define $\nu_{\mathrm{AL}}^{(a)}(\delta, \lambda)$ $(a = 1, 2, 3)$ as follows:

$$\begin{aligned}\nu_{\mathrm{AL}}^{(1)}(\delta, \lambda) &= \hat{t}k_L + k_G, \\ \nu_{\mathrm{AL}}^{(2)}(\delta, \lambda) &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}), \\ \nu_{\mathrm{AL}}^{(3)}(\delta, \lambda) &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) + \lambda\delta\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{R}_{\mathrm{AL}}).\end{aligned} \tag{20}$$

$\nu_{\mathrm{AL}}^{(3)}$ is the estimator which is found precisely by considering $\boldsymbol{W}_{\mathrm{AL}}(\delta)$ as a function of $\boldsymbol{y}_j$. $\nu_{\mathrm{AL}}^{(2)}$ corresponds to the estimator when $\boldsymbol{W}_{\mathrm{AL}}(\delta)$ is regarded as a constant with respect to $\boldsymbol{y}_j$. $\nu_{\mathrm{AL}}^{(1)}$ is the number of distinct regression coefficients and is often used as a formal estimator of generalized degrees of freedom. Regarding $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$ and $\nu_{\mathrm{AL}}^{(3)}$, the following inequality holds (the proof is given in Appendix B.9):

Proposition 1. The following properties hold:

(1) For any $\delta \geq 0$ and $\lambda \geq 0$, it holds that $\nu_{\mathrm{AL}}^{(1)}(\delta, \lambda) \geq \nu_{\mathrm{AL}}^{(2)}(\delta, \lambda)$. In particular, when $k_L = 1$, $\nu_{\mathrm{AL}}^{(1)}(\delta, \lambda) = \nu_{\mathrm{AL}}^{(2)}(\delta, \lambda)$.

(2) When $\delta = 0$, $\nu_{\mathrm{AL}}^{(2)}(0, \lambda) = \nu_{\mathrm{AL}}^{(3)}(0, \lambda)$.

4 Numerical studies

In this section, we compare the estimators of the generalized degrees of freedom for Network Lasso through numerical experiments. Specifically, we focus on the case where $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{AL}}(\delta)$, and compare $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$, and $\nu_{\mathrm{AL}}^{(3)}$ from (20). As k_G increases, the differences between $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$, and $\nu_{\mathrm{AL}}^{(3)}$ become relatively small. Therefore, we set $k_G = 0$ to make the differences between $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$, and $\nu_{\mathrm{AL}}^{(3)}$ more clear.

Here, we consider the graph as shown in Figure 1 with $m = 10$. In this situation, the set of vertices

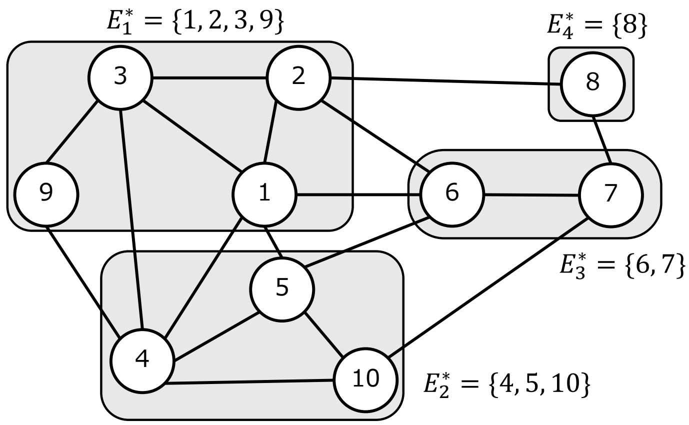

Figure 1: Adjacent relationships among vertices and true connected vertices

adjacent to the j-th vertex, denoted by D_j, and the true connected vertex, denoted by E_s^* ($s \in \{1,2,3,4\}$), are given by the following:

- D_j : the set of vertices adjacent to the j-th vertex ($j \in \{1,\ldots,10\}$)

$$D_1 = \{2,3,4,5,6\}, \quad D_2 = \{1,3,6,8\}, \quad D_3 = \{1,2,4,9\}, \quad D_4 = \{1,3,5,9,10\},$$
$$D_5 = \{1,4,6,10\}, \quad D_6 = \{1,2,5,7\}, \quad D_7 = \{6,8,10\}, \quad D_8 = \{2,7\},$$
$$D_9 = \{3,4\}, \quad D_{10} = \{4,5,7\}.$$

- E_s^* : s-th true connected vertex ($s \in \{1,2,3,4\}$)

$$E_1^* = \{1,2,3,9\}, \quad E_2^* = \{4,5,10\}, \quad E_3^* = \{6,7\}, \quad E_4^* = \{8\}.$$

For this graph, we consider fitting the following linear regression model to the j-th vertex:

$$\boldsymbol{y}_j = \boldsymbol{X}_j \boldsymbol{\beta}_j^* + \boldsymbol{\varepsilon}_j \quad (j = 1,\ldots,10),$$

where $(\boldsymbol{X}_1', \ldots, \boldsymbol{X}_{10}')' = (\boldsymbol{1}_n, \boldsymbol{X}_0 \boldsymbol{\Psi}(0.5)^{1/2})$, $(\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_{10}')' \sim N_n(\boldsymbol{0}_n, \sigma^2 \boldsymbol{I}_n)$, and $\boldsymbol{1}_n$ is an n-dimensional vector of ones. Moreover, $\boldsymbol{X}_0$ is an $n \times (k_L - 1)$ matrix where each component is independently drawn from $U(-1,1)$, and $\boldsymbol{\Psi}(\rho)$ is a $(k_L - 1) \times (k_L - 1)$ symmetric matrix whose (a,b) component is $\rho^{|a-b|}$. Let $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_1^{*\prime}, \boldsymbol{\xi}_2^{*\prime}, \boldsymbol{\xi}_3^{*\prime}, \boldsymbol{\xi}_4^{*\prime})'$ be a $4k_L$-dimensional vector where each component is independently generated from $U(-1,1)$, and $\boldsymbol{\beta}_j^* = \boldsymbol{\xi}_s^*$ for $j \in E_s^*$.

4.1 Comparison $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$ and $\nu_{\mathrm{AL}}^{(3)}$ with the true degrees of freedom

In this subsection, we compare $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$ and $\nu_{\mathrm{AL}}^{(3)}$ with the true degrees of freedom, denoted by $\mathrm{df}_{\mathrm{true}}$. Here, we calculate $\mathrm{df}_{\mathrm{true}}$ as follows:

$$\mathrm{df}_{\mathrm{true}}(\delta, \lambda) = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \frac{Cov(y_{ji}, \hat{y}_{ji,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)})}{\sigma^2}.$$

The settings for k_L, n, σ^2, and δ are as follows:

- $(k_L, n) \in \{(1, 20), (10, 200), (20, 400)\}$
- $\sigma^2 = 1$
- $\delta \in \{0, 0.5, 1, 2\}$

Under these settings, we calculate $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$, $\nu_{\mathrm{AL}}^{(3)}$, and $\mathrm{df}_{\mathrm{true}}$ through a Monte Carlo simulation with 1,000 iterations. In Figure 2, we plot $\mathrm{df}_{\mathrm{true}}$, $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$, and $\nu_{\mathrm{AL}}^{(3)}$, with the horizontal axis representing λ and the vertical axis representing degrees of freedom.

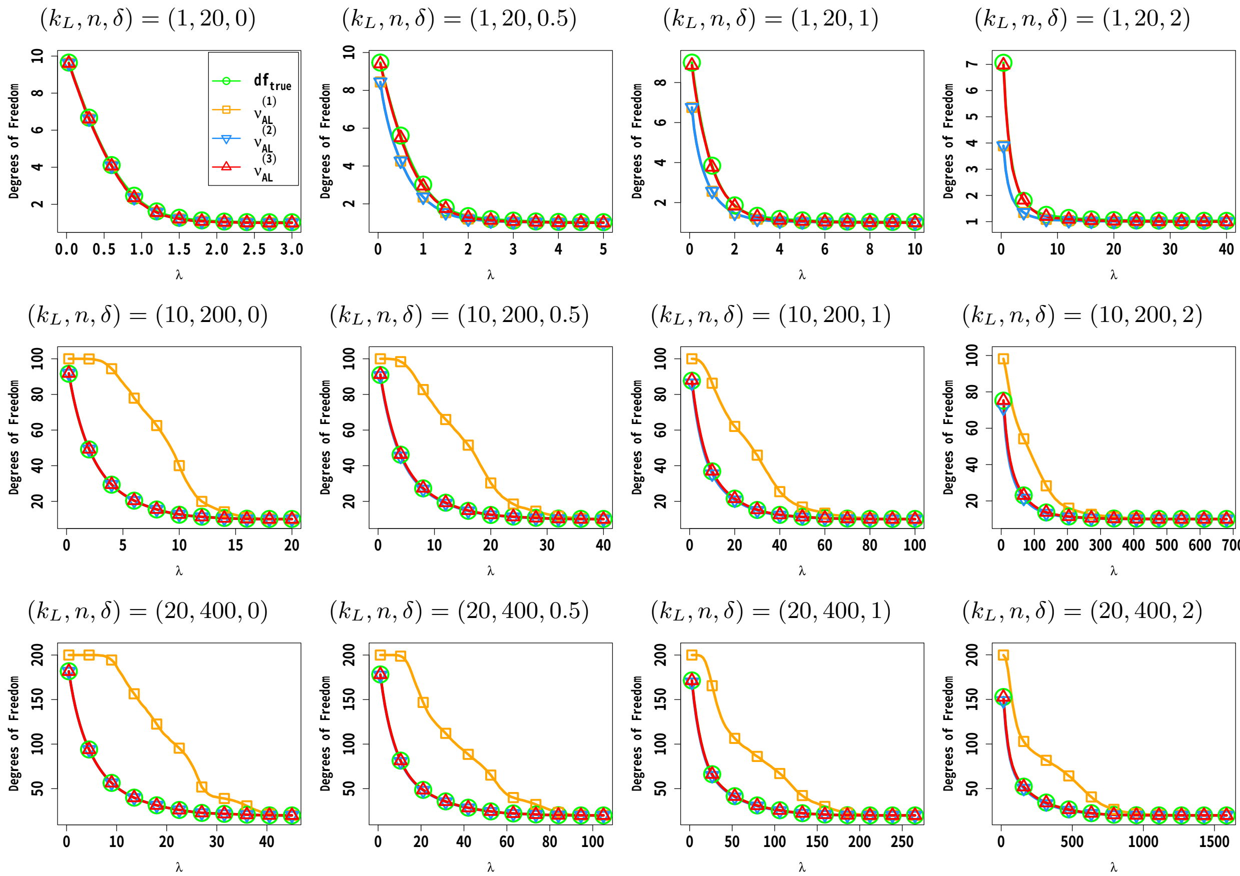

Figure 2: Comparison $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$ and $\nu_{\mathrm{AL}}^{(3)}$ with $\mathrm{df}_{\mathrm{true}}$

$\nu_{\mathrm{AL}}^{(1)}$ takes values different from $\mathrm{df}_{\mathrm{true}}$ except when $k_L = 1$ and $\delta = 0$. In particular, the difference between $\nu_{\mathrm{AL}}^{(1)}$ and $\mathrm{df}_{\mathrm{true}}$ is more clear when λ is small. From Proposition 1 (1), when $k_L = 1$, $\nu_{\mathrm{AL}}^{(2)}$ is equal to $\nu_{\mathrm{AL}}^{(1)}$. Therefore, for $k_L = 1$ and $\delta > 0$, $\nu_{\mathrm{AL}}^{(2)}$ differs from $\mathrm{df}_{\mathrm{true}}$ in the same way as $\nu_{\mathrm{AL}}^{(1)}$. On the other hand, when $k_L > 1$, $\nu_{\mathrm{AL}}^{(2)}$ takes values close to $\mathrm{df}_{\mathrm{true}}$. It can be seen that $\nu_{\mathrm{AL}}^{(3)}$ takes values close to $\mathrm{df}_{\mathrm{true}}$ in all cases. Therefore, $\nu_{\mathrm{AL}}^{(3)}$ is considered an appropriate estimator of the degrees of freedom for the Network Lasso.

4.2 Comparison of the MSE of the predicted values when δ and λ are selected simultaneously

In this subsection, we compare the MSE of the predicted values when δ and λ are selected simultaneously by minimizing the C_p criterion. Here, we replace $\mathrm{df}(\boldsymbol{W}, \lambda)$ in (8) with $\nu_{\mathrm{AL}}^{(1)}$, $\nu_{\mathrm{AL}}^{(2)}$, or $\nu_{\mathrm{AL}}^{(3)}$. The settings for k_L, n, σ^2, and δ are as follows:

- $(k_L, n) \in \{(5,100),(10,200),(20,400),(50,1000),(8,100),(16,200),(32,400),(80,1000)\}$
- $\sigma^2 \in \{1, 2\}$
- $\delta \in \{0, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

Furthermore, the MSE of the predicted values is calculated as follows:

$$\text{MSE} = E\left[\frac{1}{n\sigma^2}\sum_{j=1}^{10}\|\hat{\boldsymbol{y}}_{j,(\boldsymbol{W}_{\text{AL}}(\hat{\delta}),\hat{\lambda})} - \boldsymbol{X}_j\boldsymbol{\beta}_j^*\|^2\right],$$

where $\hat{\delta}$ and $\hat{\lambda}$ are the selected values of δ and λ, respectively. Under these settings, we calculate the MSE of the predicted values through 1,000 Monte Carlo simulations. Table 1 shows the results of the MSE of

k_L	n	$\sigma^2 = 1$			$\sigma^2 = 2$		
		$\nu_{\text{AL}}^{(1)}$	$\nu_{\text{AL}}^{(2)}$	$\nu_{\text{AL}}^{(3)}$	$\nu_{\text{AL}}^{(1)}$	$\nu_{\text{AL}}^{(2)}$	$\nu_{\text{AL}}^{(3)}$
5	100	0.37625	0.27871	**0.25678**	0.36008	0.25261	**0.22357**
10	200	0.31975	0.25136	**0.24803**	0.35945	0.23120	**0.21910**
20	400	0.23035	**0.21158**	0.21232	0.28315	0.21753	**0.21723**
50	1000	0.19222	0.18982	**0.18920**	0.19000	0.18420	**0.18388**
8	100	0.66342	0.38660	**0.37003**	0.54561	0.34245	**0.32240**
16	200	0.63085	0.39288	**0.38736**	0.52502	0.32716	**0.31853**
32	400	0.49339	0.37217	**0.37216**	0.60847	0.35601	**0.35467**
80	1000	0.31331	0.30485	**0.30479**	0.35206	**0.31939**	0.31956

Table 1: Results of the MSEs of the predicted values

the predicted values for each setting. Note that the bold values in the table indicate the minimum value for each setting. In this simulation, there are 16 settings. When using $\nu_{\text{AL}}^{(3)}$, the MSE of the predicted values is minimized in 14 out of the 16 settings. For $(k_L, n) = (20, 400), (80, 1000)$, using $\nu_{\text{AL}}^{(2)}$ minimizes the MSE of the predicted values. However, the results using $\nu_{\text{AL}}^{(3)}$ are nearly equivalent. Both $\nu_{\text{AL}}^{(2)}$ and $\nu_{\text{AL}}^{(3)}$ surpass $\nu_{\text{AL}}^{(1)}$ in all settings. Therefore, when $\boldsymbol{W} = \boldsymbol{W}_{\text{AL}}(\delta)$, using $\nu_{\text{AL}}^{(3)}$ allows for a well-chosen simultaneous selection of δ and λ.

5 Conclusion

In this paper, we dealt with the generalized degrees of freedom for Network Lasso. When selecting tuning parameters for Network Lasso, we often use methods that minimize model selection criteria such as the C_p criterion. However, in order to calculate the model selection criterion, we need to calculate the degrees of freedom. We derived a degrees of freedom estimator for Network Lasso based on the concept of generalized degrees of freedom (Efron, 2004 [4]). The generalized degrees of freedom for Network Lasso is calculated using the derivative of the predicted value $\hat{\boldsymbol{y}}_j = \boldsymbol{X}_j\hat{\boldsymbol{\beta}}_j + \boldsymbol{Z}_j\hat{\boldsymbol{\gamma}}$ with respect to $\boldsymbol{y}_j$, where $\hat{\boldsymbol{\beta}}_j$ and $\hat{\boldsymbol{\gamma}}$ are the values of $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}$ that minimize (2). However, when there are j, ℓ such that $\hat{\boldsymbol{\beta}}_j = \hat{\boldsymbol{\beta}}_\ell$, the expression (2) reaches a minimum at a non-differentiable point, making it impossible to calculate $\partial\hat{\boldsymbol{\beta}}_j/\partial\boldsymbol{y}_j$. We overcame this issue by rewriting the expression (2) into the form given by (9), allowing us to derive an estimator of the generalized degrees of freedom for Network Lasso. Additionally, through numerical studies, we confirmed that the derived estimator for the generalized degrees of freedom accurately estimates the true degrees of freedom. Furthermore, we verified that using the derived estimator as the degrees of freedom in the C_p criterion effectively selects the tuning parameters in terms of the MSE of the predicted values.

A Proofs of formulas in Section 2

A.1 Derivation of (9)

Note that when $j \in E_s$, $\boldsymbol{\beta}_j = \boldsymbol{\xi}_s = (\boldsymbol{e}_s' \otimes \boldsymbol{I}_{k_L})\boldsymbol{\xi}$. Then, (2) can be rewritten as follows:

$$\begin{aligned}
\mathrm{PRSS}(\boldsymbol{\beta}, \boldsymbol{\gamma} \,|\, \boldsymbol{W}, \lambda) &= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \|\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\beta}_j - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2 + \frac{\lambda}{2}\sum_{s=1}^{\hat{t}}\sum_{u=1}^{\hat{t}}\sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_\ell\| \\
&= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \|\boldsymbol{y}_j - \boldsymbol{X}_j(\boldsymbol{e}_s' \otimes \boldsymbol{I}_{k_L})\boldsymbol{\xi} - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2 + \frac{\lambda}{2}\sum_{s=1}^{\hat{t}}\sum_{u=1}^{\hat{t}}\sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}\|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\| \\
&= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \|\boldsymbol{y}_j - (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j)\boldsymbol{\xi} - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2 + \frac{\lambda}{2}\sum_{s=1}^{\hat{t}}\sum_{u=1}^{\hat{t}}\sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}\|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|.
\end{aligned}$$

Here, we note that when $u = s$, $\|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\| = 0$, and when $u \notin \{s\} \cup F_s$, $E_u \cap D_j = \emptyset$. Therefore, we have

$$\begin{aligned}
\mathrm{PRSS}(\boldsymbol{\beta}, \boldsymbol{\gamma} \,|\, \boldsymbol{W}, \lambda) &= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \|\boldsymbol{y}_j - (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j)\boldsymbol{\xi} - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2 + \frac{\lambda}{2}\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}\|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\| \\
&= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \|\boldsymbol{y}_j - (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j)\boldsymbol{\xi} - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2 + \frac{\lambda}{2}\sum_{s=1}^{\hat{t}}\sum_{u\in F_s} v_{su}\|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\| \\
&= f(\boldsymbol{\xi}, \boldsymbol{\gamma}),
\end{aligned}$$

where $v_{su} = \sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}$. Also, by noting that $w_{j\ell} = w_{\ell j}$ and $w_{j\ell} = 0$ ($\ell \notin D_j$), we obtain

$$\begin{aligned}
v_{su} &= \sum_{j\in E_s}\left(\sum_{\ell\in E_u\cap D_j} w_{j\ell} + \sum_{\ell\in E_u\setminus D_j} w_{j\ell}\right) = \sum_{j\in E_s}\sum_{\ell\in E_u} w_{j\ell} \\
&= \sum_{\ell\in E_u}\sum_{j\in E_s} w_{\ell j} = \sum_{\ell\in E_u}\left(\sum_{j\in E_s\cap D_\ell} w_{j\ell} + \sum_{j\in E_s\setminus D_\ell} w_{j\ell}\right) = \sum_{\ell\in E_u}\sum_{j\in E_s\cap D_\ell} w_{j\ell} = v_{us}.
\end{aligned}$$

A.2 Derivation of (10)

We define $f_1(\boldsymbol{\xi}, \boldsymbol{\gamma})$ and $f_2(\boldsymbol{\xi})$ as follows:

$$f_1(\boldsymbol{\xi}, \boldsymbol{\gamma}) = \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \|\boldsymbol{y}_j - (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j)\boldsymbol{\xi} - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2, \quad f_2(\boldsymbol{\xi}) = \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{u\in F_s} v_{su}\|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|. \tag{21}$$

Then, $f(\boldsymbol{\xi}, \boldsymbol{\gamma}) = f_1(\boldsymbol{\xi}, \boldsymbol{\gamma}) + \lambda f_2(\boldsymbol{\xi})$, and $\boldsymbol{g}$, $\boldsymbol{h}$ are obtained as follows:

$$\begin{aligned}
\boldsymbol{g} &= \frac{\partial}{\partial \boldsymbol{\xi}} f(\boldsymbol{\xi}, \boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \frac{\partial}{\partial \boldsymbol{\xi}} f_1(\boldsymbol{\xi}, \boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial}{\partial \boldsymbol{\xi}} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}, \\
\boldsymbol{h} &= \frac{\partial}{\partial \boldsymbol{\gamma}} f(\boldsymbol{\xi}, \boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \frac{\partial}{\partial \boldsymbol{\gamma}} f_1(\boldsymbol{\xi}, \boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial}{\partial \boldsymbol{\gamma}} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}.
\end{aligned}$$

$\partial f_1(\boldsymbol{\xi}, \boldsymbol{\gamma})/\partial \boldsymbol{\xi}$ is given by

$$\frac{\partial}{\partial \boldsymbol{\xi}} f_1(\boldsymbol{\xi}, \boldsymbol{\gamma}) = \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s} \frac{\partial}{\partial \boldsymbol{\xi}}\|\boldsymbol{y}_j - (\boldsymbol{e}_s' \otimes \boldsymbol{X}_j)\boldsymbol{\xi} - \boldsymbol{Z}_j\boldsymbol{\gamma}\|^2$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\{-2(\boldsymbol{e}_s'\otimes\boldsymbol{X}_j)'\}\{\boldsymbol{y}_j-(\boldsymbol{e}_s'\otimes\boldsymbol{X}_j)\boldsymbol{\xi}-\boldsymbol{Z}_j\boldsymbol{\gamma}\}\\
&= \sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\{(\boldsymbol{e}_s\otimes\boldsymbol{X}_j')(\boldsymbol{e}_s'\otimes\boldsymbol{X}_j)\boldsymbol{\xi}+(\boldsymbol{e}_s\otimes\boldsymbol{X}_j')\boldsymbol{Z}_j\boldsymbol{\gamma}-(\boldsymbol{e}_s\otimes\boldsymbol{X}_j')\boldsymbol{y}_j\}\\
&= \sum_{s=1}^{\hat{t}}\sum_{j\in E_s}(\boldsymbol{e}_s\boldsymbol{e}_s'\otimes\boldsymbol{X}_j'\boldsymbol{X}_j)\boldsymbol{\xi}+\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes\boldsymbol{X}_j'\boldsymbol{Z}_j)\boldsymbol{\gamma}-\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes\boldsymbol{X}_j'\boldsymbol{y}_j)\\
&= \boldsymbol{A}\boldsymbol{\xi}+\boldsymbol{B}\boldsymbol{\gamma}-\boldsymbol{c}.
\end{aligned}
$$

Therefore, we have

$$
\left.\frac{\partial}{\partial\boldsymbol{\xi}}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}=\boldsymbol{A}\hat{\boldsymbol{\xi}}+\boldsymbol{B}\hat{\boldsymbol{\gamma}}-\boldsymbol{c}.
$$

$\partial f_1(\boldsymbol{\xi},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$ is given by

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\gamma}}f_1(\boldsymbol{\xi},\boldsymbol{\gamma}) &= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\frac{\partial}{\partial\boldsymbol{\gamma}}\|\boldsymbol{y}_j-(\boldsymbol{e}_s'\otimes\boldsymbol{X}_j)\boldsymbol{\xi}-\boldsymbol{Z}_j\boldsymbol{\gamma}\|^2\\
&= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}(-2\boldsymbol{Z}_j')\{\boldsymbol{y}_j-(\boldsymbol{e}_s'\otimes\boldsymbol{X}_j)\boldsymbol{\xi}-\boldsymbol{Z}_j\boldsymbol{\gamma}\}\\
&= \sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\{\boldsymbol{Z}_j'(\boldsymbol{e}_s'\otimes\boldsymbol{X}_j)\boldsymbol{\xi}+\boldsymbol{Z}_j'\boldsymbol{Z}_j\boldsymbol{\gamma}-\boldsymbol{Z}_j'\boldsymbol{y}_j\}\\
&= \sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\{(\boldsymbol{e}_s\otimes\boldsymbol{X}_j')\boldsymbol{Z}_j\}'\boldsymbol{\xi}+\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\boldsymbol{Z}_j'\boldsymbol{Z}_j\boldsymbol{\gamma}-\sum_{s=1}^{\hat{t}}\sum_{j\in E_s}\boldsymbol{Z}_j'\boldsymbol{y}_j\\
&= \sum_{s=1}^{\hat{t}}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes\boldsymbol{X}_j'\boldsymbol{Z}_j)'\boldsymbol{\xi}+\sum_{j=1}^{m}\boldsymbol{Z}_j'\boldsymbol{Z}_j\boldsymbol{\gamma}-\sum_{j=1}^{m}\boldsymbol{Z}_j'\boldsymbol{y}_j\\
&= \boldsymbol{B}'\boldsymbol{\xi}+\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\gamma}-\boldsymbol{Z}'\boldsymbol{y}.
\end{aligned}
$$

Therefore, we have

$$
\left.\frac{\partial}{\partial\boldsymbol{\gamma}}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}=\boldsymbol{B}'\hat{\boldsymbol{\xi}}+\boldsymbol{Z}'\boldsymbol{Z}\hat{\boldsymbol{\gamma}}-\boldsymbol{Z}'\boldsymbol{y}.
$$

Note that $f_2(\boldsymbol{\xi})$ does not depend on $\boldsymbol{\gamma}$. Then, $\partial f_2(\boldsymbol{\xi})/\partial\boldsymbol{\gamma}=\boldsymbol{0}_{k_G}$, and we obtain

$$
\left.\frac{\partial}{\partial\boldsymbol{\gamma}}f_2(\boldsymbol{\xi})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}=\boldsymbol{0}_{k_G}.
$$

Let $\boldsymbol{\Omega}_{su}=\boldsymbol{e}_s'\otimes\boldsymbol{I}_{k_L}-\boldsymbol{e}_u'\otimes\boldsymbol{I}_{k_L}$ and $\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})=\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}\boldsymbol{\xi}$. Then, $\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})=\|\boldsymbol{\xi}_s-\boldsymbol{\xi}_u\|^{-1}(\boldsymbol{\xi}_s-\boldsymbol{\xi}_u)$, and $\partial f_2(\boldsymbol{\xi})/\partial\boldsymbol{\xi}$ can be transformed as follows:

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\xi}}f_2(\boldsymbol{\xi}) &= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}v_{su}\frac{\partial}{\partial\boldsymbol{\xi}}\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|=\frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}v_{su}\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}'\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\\
&= \frac{1}{2}\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\left\{v_{su}(\boldsymbol{e}_s\otimes\boldsymbol{I}_{k_L}-\boldsymbol{e}_u\otimes\boldsymbol{I}_{k_L})\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right\}
\end{aligned}
$$

$$= \frac{1}{2}\left\{\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\left(\boldsymbol{e}_s \otimes v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right) - \sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\left(\boldsymbol{e}_u \otimes v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right)\right\}.$$

Since $\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi}) = -\tilde{\boldsymbol{\theta}}_{us}(\boldsymbol{\xi})$ and $v_{su} = v_{us}$, we find

$$\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\left(\boldsymbol{e}_u \otimes v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right) = -\sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\left(\boldsymbol{e}_s \otimes v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right).$$

Hence, we have

$$\frac{\partial}{\partial \boldsymbol{\xi}} f_2(\boldsymbol{\xi}) = \sum_{s=1}^{\hat{t}}\sum_{u\in F_s}\left(\boldsymbol{e}_s \otimes v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right) = \sum_{s=1}^{\hat{t}}\left(\boldsymbol{e}_s \otimes \sum_{u\in F_s} v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right).$$

Therefore, noting that $\tilde{\boldsymbol{\theta}}_{su}(\hat{\boldsymbol{\xi}}) = \boldsymbol{\theta}_{su}$, the following holds:

$$\frac{\partial}{\partial \boldsymbol{\xi}} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \sum_{s=1}^{\hat{t}}\left(\boldsymbol{e}_s \otimes \sum_{u\in F_s} v_{su}\tilde{\boldsymbol{\theta}}_{su}(\hat{\boldsymbol{\xi}})\right) = \sum_{s=1}^{\hat{t}}\left(\boldsymbol{e}_s \otimes \sum_{u\in F_s} v_{su}\boldsymbol{\theta}_{su}\right) = \boldsymbol{\theta}.$$

Consequently, $\boldsymbol{g}$ and $\boldsymbol{h}$ are given by

$$\begin{aligned}
\boldsymbol{g} &= \frac{\partial}{\partial \boldsymbol{\xi}} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial}{\partial \boldsymbol{\xi}} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{A}\hat{\boldsymbol{\xi}} + \boldsymbol{B}\hat{\boldsymbol{\gamma}} - \boldsymbol{c} + \lambda\boldsymbol{\theta},\\
\boldsymbol{h} &= \frac{\partial}{\partial \boldsymbol{\gamma}} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial}{\partial \boldsymbol{\gamma}} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}'\hat{\boldsymbol{\xi}} + \boldsymbol{Z}'\boldsymbol{Z}\hat{\boldsymbol{\gamma}} - \boldsymbol{Z}'\boldsymbol{y}.
\end{aligned}$$

A.3 Deviation of (12)

From (21), $\boldsymbol{G}$, $\boldsymbol{H}$, $\boldsymbol{J}$, and $\boldsymbol{K}$ are obtained as follows:

$$\begin{aligned}
\boldsymbol{G} &= \frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'} f(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}},\\
\boldsymbol{H} &= \frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'} f(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}},\\
\boldsymbol{J} &= \frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\gamma}'} f(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\gamma}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\gamma}'} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}},\\
\boldsymbol{K} &= \frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\xi}'} f(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\xi}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\xi}'} f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}.
\end{aligned}$$

$\partial^2 f_1(\boldsymbol{\xi},\boldsymbol{\gamma})/\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'$ is given by

$$\frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma}) = \frac{\partial}{\partial \boldsymbol{\xi}}\left(\frac{\partial}{\partial \boldsymbol{\xi}} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right)' = \frac{\partial}{\partial \boldsymbol{\xi}}(\boldsymbol{\xi}'\boldsymbol{A}' + \boldsymbol{\gamma}'\boldsymbol{B}' - \boldsymbol{c}') = \boldsymbol{A}'.$$

Here, noting that $\boldsymbol{A}$ is a symmetric matrix, we have

$$\frac{\partial^2}{\partial \boldsymbol{\xi}\partial \boldsymbol{\xi}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{A}.$$

$\partial^2 f_1(\boldsymbol{\xi},\boldsymbol{\gamma})/\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'$ is given by

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma}) = \frac{\partial}{\partial \boldsymbol{\gamma}}\left(\frac{\partial}{\partial \boldsymbol{\gamma}} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right)' = \frac{\partial}{\partial \boldsymbol{\gamma}}(\boldsymbol{\xi}'\boldsymbol{B} + \boldsymbol{\gamma}'\boldsymbol{Z}'\boldsymbol{Z} - \boldsymbol{y}'\boldsymbol{Z}) = \boldsymbol{Z}'\boldsymbol{Z}.$$

Hence, we have

$$\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{Z}'\boldsymbol{Z}.$$

$\partial^2 f_1(\boldsymbol{\xi},\boldsymbol{\gamma})/\partial\boldsymbol{\xi}\partial\boldsymbol{\gamma}'$ is given by

$$\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\gamma}'}f_1(\boldsymbol{\xi},\boldsymbol{\gamma}) = \frac{\partial}{\partial\boldsymbol{\xi}}\left(\frac{\partial}{\partial\boldsymbol{\gamma}}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right)' = \frac{\partial}{\partial\boldsymbol{\xi}}(\boldsymbol{\xi}'\boldsymbol{B}+\boldsymbol{\gamma}'\boldsymbol{Z}'\boldsymbol{Z}-\boldsymbol{y}'\boldsymbol{Z}) = \boldsymbol{B}.$$

Therefore, we have

$$\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\gamma}'}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}.$$

$\partial^2 f_1(\boldsymbol{\xi},\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\xi}'$ is given by

$$\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\xi}'}f_1(\boldsymbol{\xi},\boldsymbol{\gamma}) = \frac{\partial}{\partial\boldsymbol{\gamma}}\left(\frac{\partial}{\partial\boldsymbol{\xi}}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right)' = \frac{\partial}{\partial\boldsymbol{\gamma}}(\boldsymbol{\xi}'\boldsymbol{A}'+\boldsymbol{\gamma}'\boldsymbol{B}'-\boldsymbol{c}') = \boldsymbol{B}'.$$

Hence, we have

$$\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\xi}'}f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}'.$$

Noting that $f_2(\boldsymbol{\xi})$ does not depend on $\boldsymbol{\gamma}$, we obtain

$$\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\gamma}'}f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{O}_{\hat{t}k_L,k_G}, \quad \frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\xi}'}f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{O}_{k_G,\hat{t}k_L}, \quad \frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}f_2(\boldsymbol{\xi})\bigg|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{O}_{k_G,k_G}.$$

$\partial^2 f_2(\boldsymbol{\xi})/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'$ is given by

$$\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'}f_2(\boldsymbol{\xi}) = \frac{\partial}{\partial\boldsymbol{\xi}}\left(\frac{\partial}{\partial\boldsymbol{\xi}}f_2(\boldsymbol{\xi})\right)' = \sum_{s=1}^{\hat{t}}\frac{\partial}{\partial\boldsymbol{\xi}}\left(\boldsymbol{e}_s\otimes\sum_{u\in F_s}v_{su}\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\right)' = \sum_{s=1}^{\hat{t}}\left(\boldsymbol{e}_s'\otimes\sum_{u\in F_s}v_{su}\frac{\partial\tilde{\boldsymbol{\theta}}_{su}'(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}\right).$$

Here, we find

$$\frac{\partial\tilde{\boldsymbol{\theta}}_{su}'(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}} = \frac{\partial}{\partial\boldsymbol{\xi}}(\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}\boldsymbol{\xi})' = \left(\frac{\partial}{\partial\boldsymbol{\xi}}\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\right)(\boldsymbol{\Omega}_{su}\boldsymbol{\xi})' + \|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\left(\frac{\partial}{\partial\boldsymbol{\xi}}\boldsymbol{\xi}'\boldsymbol{\Omega}_{su}'\right).$$

Noting that

$$\frac{\partial}{\partial\boldsymbol{\xi}}\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1} = -\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-2}\boldsymbol{\Omega}_{su}'\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi}), \quad \frac{\partial}{\partial\boldsymbol{\xi}}\boldsymbol{\xi}'\boldsymbol{\Omega}_{su}' = \boldsymbol{\Omega}_{su}',$$

we have

$$\begin{aligned}\frac{\partial\tilde{\boldsymbol{\theta}}_{su}'(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}} &= -\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-2}\boldsymbol{\Omega}_{su}'\tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})(\boldsymbol{\Omega}_{su}\boldsymbol{\xi})' + \|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}' \\ &= \|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}'\left\{\boldsymbol{I}_{k_L} - \tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})(\|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}\boldsymbol{\xi})'\right\} \\ &= \|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1}\boldsymbol{\Omega}_{su}'\tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}),\end{aligned}$$

where $\tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) = \boldsymbol{I}_{k_L} - \tilde{\boldsymbol{\theta}}_{su}(\boldsymbol{\xi})\tilde{\boldsymbol{\theta}}'_{su}(\boldsymbol{\xi})$. Therefore,

$$\begin{aligned}
\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'} f_2(\boldsymbol{\xi}) &= \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}'_s \otimes \sum_{u\in F_s} v_{su} \|\boldsymbol{\Omega}_{su}\boldsymbol{\xi}\|^{-1} \boldsymbol{\Omega}'_{su} \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) \right) \\
&= \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}'_s \otimes \sum_{u\in F_s} v_{su} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|^{-1} (\boldsymbol{e}_s \otimes \boldsymbol{I}_{k_L} - \boldsymbol{e}_u \otimes \boldsymbol{I}_{k_L}) \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) \right) \\
&= \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}'_s \otimes \sum_{u\in F_s} v_{su} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|^{-1} (\boldsymbol{e}_s \otimes \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) - \boldsymbol{e}_u \otimes \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi})) \right) \\
&= \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}'_s \otimes \boldsymbol{e}_s \otimes \sum_{u\in F_s} v_{su} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|^{-1} \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) - \sum_{u\in F_s} \boldsymbol{e}'_s \otimes \boldsymbol{e}_u \otimes v_{su} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|^{-1} \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) \right) \\
&= \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s \boldsymbol{e}'_s \otimes \tilde{\boldsymbol{L}}'_{ss}(\boldsymbol{\xi}) + \sum_{u\in F_s} \boldsymbol{e}_u \boldsymbol{e}'_s \otimes \tilde{\boldsymbol{L}}'_{su}(\boldsymbol{\xi}) + \sum_{u\in\{1,\ldots,\hat{t}\}\setminus(\{s\}\cup F_s)} \boldsymbol{e}_u \boldsymbol{e}'_s \otimes \tilde{\boldsymbol{L}}'_{su}(\boldsymbol{\xi}) \right) \\
&= \sum_{s=1}^{\hat{t}} \sum_{u=1}^{\hat{t}} \boldsymbol{e}_u \boldsymbol{e}'_s \otimes \tilde{\boldsymbol{L}}'_{su}(\boldsymbol{\xi}) \\
&= \tilde{\boldsymbol{L}}'(\boldsymbol{\xi}),
\end{aligned} \tag{22}$$

where $\tilde{\boldsymbol{L}}(\boldsymbol{\xi})$ is a matrix defined as follows:

$$\tilde{\boldsymbol{L}}(\boldsymbol{\xi}) = \sum_{s=1}^{\hat{t}} \sum_{u=1}^{\hat{t}} \boldsymbol{e}_s \boldsymbol{e}'_u \otimes \tilde{\boldsymbol{L}}_{su}(\boldsymbol{\xi}), \quad \tilde{\boldsymbol{L}}_{su}(\boldsymbol{\xi}) = \begin{cases} \sum_{p\in F_s} v_{sp} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_p\|^{-1} \tilde{\boldsymbol{\Theta}}_{sp}(\boldsymbol{\xi}) & (u = p) \\ -v_{su} \|\boldsymbol{\xi}_s - \boldsymbol{\xi}_u\|^{-1} \tilde{\boldsymbol{\Theta}}_{su}(\boldsymbol{\xi}) & (u \in F_p) \\ \boldsymbol{O}_{k_L,k_L} & (\mathit{otherwise}) \end{cases}.$$

Noting that $\tilde{\boldsymbol{L}}'_{su}(\boldsymbol{\xi}) = \tilde{\boldsymbol{L}}_{su}(\boldsymbol{\xi})$, $\tilde{\boldsymbol{L}}_{us}(\boldsymbol{\xi}) = \tilde{\boldsymbol{L}}_{su}(\boldsymbol{\xi})$, we have $\tilde{\boldsymbol{L}}'(\boldsymbol{\xi}) = \tilde{\boldsymbol{L}}(\boldsymbol{\xi})$. Therefore, we obtain

$$\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'} f_2(\boldsymbol{\xi}) = \tilde{\boldsymbol{L}}(\boldsymbol{\xi}).$$

Since $\tilde{\boldsymbol{L}}(\hat{\boldsymbol{\xi}}) = \boldsymbol{L}$, we have

$$\left.\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'} f_2(\boldsymbol{\xi})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{L},$$

where $\boldsymbol{L}$ is a matrix defined in (13).

Consequently, we have

$$\begin{aligned}
\boldsymbol{G} &= \left.\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \left.\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}'} f_2(\boldsymbol{\xi})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{A} + \lambda\boldsymbol{L}, \\
\boldsymbol{H} &= \left.\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \left.\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} f_2(\boldsymbol{\xi})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{Z}'\boldsymbol{Z}, \\
\boldsymbol{J} &= \left.\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\gamma}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \left.\frac{\partial^2}{\partial\boldsymbol{\xi}\partial\boldsymbol{\gamma}'} f_2(\boldsymbol{\xi})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}, \\
\boldsymbol{K} &= \left.\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\xi}'} f_1(\boldsymbol{\xi},\boldsymbol{\gamma})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} + \lambda \left.\frac{\partial^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\xi}'} f_2(\boldsymbol{\xi})\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}},\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \boldsymbol{B}'.
\end{aligned}$$

A.4 Proof of Lemma 1

First, we show the positive definiteness of $\boldsymbol{A}$. From (11), $\boldsymbol{A}$ is symmetric matrix. Therefore, it suffices to show that $\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a} > 0$ for any $\boldsymbol{a} \in \mathbb{R}^{\hat{t}k_L} \setminus \{\boldsymbol{0}_{\hat{t}k_L}\}$. By assumption, $\mathrm{rank}(\boldsymbol{X}_j) = k_L < n_j$, which implies that $\boldsymbol{X}_j'\boldsymbol{X}_j$ is invertible. In other words, since $\boldsymbol{X}_j'\boldsymbol{X}_j$ is invertible and positive semi-definite, it is positive definite. Therefore, for $p = 1, \ldots, \hat{t}$, $\sum_{j \in E_p} \boldsymbol{X}_j'\boldsymbol{X}_j$ is also positive definite. This means that for any $\boldsymbol{a}_p \in \mathbb{R}^{k_L} \setminus \{\boldsymbol{0}_{k_L}\}$, the following holds:

$$\boldsymbol{a}_p' \left(\sum_{j \in E_p} \boldsymbol{X}_j'\boldsymbol{X}_j \right) \boldsymbol{a}_p > 0.$$

Let $\boldsymbol{a} = \sum_{p=1}^{\hat{t}} \boldsymbol{e}_p \otimes \boldsymbol{a}_p$ ($\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{\hat{t}} \in \mathbb{R}^{k_L}$). Then, since $\boldsymbol{a} \neq \boldsymbol{0}_{\hat{t}k_L}$, there exists $s \in \{1, \ldots, \hat{t}\}$ such that $\boldsymbol{a}_s \neq \boldsymbol{0}_{k_L}$. Therefore, we see that

$$\begin{aligned}
\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a} &= \left(\sum_{q=1}^{\hat{t}} \boldsymbol{e}_q \otimes \boldsymbol{a}_q \right)' \left\{ \sum_{p=1}^{\hat{t}} \left(\boldsymbol{e}_p\boldsymbol{e}_p' \otimes \sum_{j \in E_p} \boldsymbol{X}_j'\boldsymbol{X}_j \right) \right\} \left(\sum_{r=1}^{\hat{t}} \boldsymbol{e}_r \otimes \boldsymbol{a}_r \right) \\
&= \sum_{q=1}^{\hat{t}} \sum_{p=1}^{\hat{t}} \sum_{r=1}^{\hat{t}} \boldsymbol{e}_q'\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{e}_r \otimes \boldsymbol{a}_q' \left(\sum_{j \in E_p} \boldsymbol{X}_j'\boldsymbol{X}_j \right) \boldsymbol{a}_r \\
&= \sum_{p=1}^{\hat{t}} \boldsymbol{e}_p'\boldsymbol{e}_p\boldsymbol{e}_p'\boldsymbol{e}_p \otimes \boldsymbol{a}_p' \left(\sum_{j \in E_p} \boldsymbol{X}_j'\boldsymbol{X}_j \right) \boldsymbol{a}_p \\
&= \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \left(\sum_{j \in E_p} \boldsymbol{X}_j'\boldsymbol{X}_j \right) \boldsymbol{a}_p \geq \boldsymbol{a}_s' \left(\sum_{j \in E_s} \boldsymbol{X}_j'\boldsymbol{X}_j \right) \boldsymbol{a}_s > 0.
\end{aligned}$$

Thus, $\boldsymbol{A}$ is positive definite.

Next, we show the symmetry of $\boldsymbol{L}$. Let $p \in \{1, \ldots, \hat{t}\}$. When $q \notin \{p\} \cup F_p$, we see that $\boldsymbol{L}_{pq} = \boldsymbol{O}_{k_L,k_L} = \boldsymbol{L}_{qp}$ since $p \notin F_q$. When $q \in F_p$, we note that $\boldsymbol{\Theta}_{pq}' = \boldsymbol{\Theta}_{pq}$. Hence, we have

$$\boldsymbol{L}_{pq} = -v_{pq}\|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_q\|^{-1}\boldsymbol{\Theta}_{pq} = -v_{pq}\|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_q\|^{-1}\boldsymbol{\Theta}_{pq}' = \boldsymbol{L}_{pq}',$$

that is, $\boldsymbol{L}_{pq}$ is symmetric matrix. Moreover, since $v_{pq} = v_{qp}$ and $\boldsymbol{\Theta}_{pq} = \boldsymbol{\Theta}_{qp}$, we find

$$\boldsymbol{L}_{pq} = -v_{qp}\|\hat{\boldsymbol{\xi}}_q - \hat{\boldsymbol{\xi}}_p\|^{-1}\boldsymbol{\Theta}_{qp} = \boldsymbol{L}_{qp}.$$

When $q = p$, since $\boldsymbol{L}_{pp} = \sum_{u \in F_p}(-\boldsymbol{L}_{pu})$, $\boldsymbol{L}_{pp}$ is symmetric matrix. Therefore, we have

$$\begin{aligned}
\boldsymbol{L}' &= \left(\sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_p\boldsymbol{e}_q' \otimes \boldsymbol{L}_{pq} \right)' = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_q\boldsymbol{e}_p' \otimes \boldsymbol{L}_{pq}' = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_q\boldsymbol{e}_p' \otimes \boldsymbol{L}_{qp}' \\
&= \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_q\boldsymbol{e}_p' \otimes \boldsymbol{L}_{qp} = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_q\boldsymbol{e}_p' \otimes \boldsymbol{L}_{qp} = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_p\boldsymbol{e}_q' \otimes \boldsymbol{L}_{pq} = \boldsymbol{L},
\end{aligned}$$

i.e., $\boldsymbol{L}$ is symmetric matrix.

Lastly, we show that $\boldsymbol{a}'\boldsymbol{L}\boldsymbol{a} \geq 0$ for any $\boldsymbol{a} \in \mathbb{R}^{\hat{t}k_L}$. Let $\boldsymbol{a} = \sum_{p=1}^{\hat{t}} \boldsymbol{e}_p \otimes \boldsymbol{a}_p$ ($\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{\hat{t}} \in \mathbb{R}^{k_L}$). Then, we have

$$\boldsymbol{a}'\boldsymbol{L}\boldsymbol{a} = \left(\sum_{r=1}^{\hat{t}} \boldsymbol{e}_r \otimes \boldsymbol{a}_r \right)' \left(\sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{e}_p\boldsymbol{e}_q \otimes \boldsymbol{L}_{pq} \right) \left(\sum_{s=1}^{\hat{t}} \boldsymbol{e}_s \otimes \boldsymbol{a}_s \right)$$

$$= \sum_{r=1}^{\hat{t}} \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \sum_{s=1}^{\hat{t}} \boldsymbol{e}_r' \boldsymbol{e}_p \boldsymbol{e}_q' \boldsymbol{e}_s \otimes \boldsymbol{a}_r' \boldsymbol{L}_{pq} \boldsymbol{a}_s$$

$$= \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q.$$

Since $\boldsymbol{L}_{pq}$ is a symmetric matrix, we can see the following:

$$\sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q = \frac{1}{2} \left(\sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} (\boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q)' \right) = \frac{1}{2} \left(\sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_q' \boldsymbol{L}_{pq}' \boldsymbol{a}_p \right)$$

$$= \frac{1}{2} \left(\sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_p \right) = \frac{1}{2} \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} (\boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_p).$$

Note that

$$\begin{aligned} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_p &= (\boldsymbol{a}_p - \boldsymbol{a}_q)' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q + (\boldsymbol{a}_q - \boldsymbol{a}_p)' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p \\ &= (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) \boldsymbol{a}_p - (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) \boldsymbol{a}_q + \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \\ &= (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) (\boldsymbol{a}_p - \boldsymbol{a}_q) + \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q &= \frac{1}{2} \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \{ (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) (\boldsymbol{a}_p - \boldsymbol{a}_q) + \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \} \\ &= \frac{1}{2} \sum_{p=1}^{\hat{t}} \left[2 \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p + \sum_{q \in F_p} \{ (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) (\boldsymbol{a}_p - \boldsymbol{a}_q) + \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \} \right] \\ &= \frac{1}{2} \sum_{p=1}^{\hat{t}} \left[-2 \sum_{q \in F_p} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \sum_{q \in F_p} \{ (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) (\boldsymbol{a}_p - \boldsymbol{a}_q) + \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \} \right] \\ &= \frac{1}{2} \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} \{ (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) (\boldsymbol{a}_p - \boldsymbol{a}_q) - \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \} \\ &= \frac{1}{2} \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} (\boldsymbol{a}_p - \boldsymbol{a}_q)' (-\boldsymbol{L}_{pq}) (\boldsymbol{a}_p - \boldsymbol{a}_q) - \frac{1}{2} \left(\sum_{p=1}^{\hat{t}} \sum_{q \in F_p} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p - \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \right). \end{aligned}$$

Since $\boldsymbol{L}_{pq} = \boldsymbol{L}_{qp}$ and for $q \in \{1, \ldots, \hat{t}\} \setminus (\{p\} \cup F_p)$, $\boldsymbol{L}_{pq} = \boldsymbol{O}_{k_L, k_L}$, we have

$$\begin{aligned} \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p &= \sum_{p=1}^{\hat{t}} \left(\boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p + \sum_{q \in F_p} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p + \sum_{q \in \{1, \ldots, \hat{t}\} \setminus (\{p\} \cup F_p)} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p - \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p \right) \\ &= \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_p - \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p = \sum_{q=1}^{\hat{t}} \sum_{p=1}^{\hat{t}} \boldsymbol{a}_q' \boldsymbol{L}_{qp} \boldsymbol{a}_q - \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p \\ &= \sum_{q=1}^{\hat{t}} \sum_{p=1}^{\hat{t}} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q - \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q - \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p \\ &= \sum_{p=1}^{\hat{t}} \left(\boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p' + \sum_{q \in F_p} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \sum_{q \in \{1, \ldots, \hat{t}\} \setminus (\{p\} \cup F_p)} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q \right) - \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p \end{aligned}$$

$$
\begin{aligned}
&= \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p' + \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q + \sum_{p=1}^{\hat{t}} \sum_{q \in \{1,\dots,\hat{t}\} \setminus (\{p\} \cup F_p)} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q - \sum_{p=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pp} \boldsymbol{a}_p \\
&= \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} \boldsymbol{a}_q' \boldsymbol{L}_{pq} \boldsymbol{a}_q.
\end{aligned}
$$

$\boldsymbol{a}'\boldsymbol{L}\boldsymbol{a}$ is obtained by

$$
\boldsymbol{a}'\boldsymbol{L}\boldsymbol{a} = \sum_{p=1}^{\hat{t}} \sum_{q=1}^{\hat{t}} \boldsymbol{a}_p' \boldsymbol{L}_{pq} \boldsymbol{a}_q = \frac{1}{2} \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} (\boldsymbol{a}_p - \boldsymbol{a}_q)'(-\boldsymbol{L}_{pq})(\boldsymbol{a}_p - \boldsymbol{a}_q).
$$

Here, for $p \in \{1,\dots,\hat{t}\}$, $q \in F_p$ and $\boldsymbol{b} \in \mathbb{R}^{k_L}$, we have

$$
\begin{aligned}
\boldsymbol{b}'(-\boldsymbol{L}_{pq})\boldsymbol{b} &= v_{pq} \|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_q\|^{-1} \boldsymbol{b}' \boldsymbol{\Theta}_{pq} \boldsymbol{b} \\
&= v_{pq} \|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_q\|^{-1} \boldsymbol{b}' (\boldsymbol{I}_{k_L} - \boldsymbol{\theta}_{pq} \boldsymbol{\theta}_{pq}') \boldsymbol{b} \\
&= v_{pq} \|\hat{\boldsymbol{\xi}}_p - \hat{\boldsymbol{\xi}}_q\|^{-1} (\|\boldsymbol{b}\|^2 - \|\boldsymbol{\theta}_{pq}' \boldsymbol{b}\|^2).
\end{aligned}
$$

From $\|\boldsymbol{\theta}_{pq}\|^2 = 1$ and the Cauchy-Schwarz inequality, we obtain $\|\boldsymbol{\theta}_{pq}'\boldsymbol{b}\|^2 \leq \|\boldsymbol{\theta}_{pq}\|^2 \|\boldsymbol{b}\|^2 = \|\boldsymbol{b}\|^2$. This implies

$$
\boldsymbol{b}'(-\boldsymbol{L}_{pq})\boldsymbol{b} \geq 0.
$$

Hence, we obtain

$$
\boldsymbol{a}'\boldsymbol{L}\boldsymbol{a} = \frac{1}{2} \sum_{p=1}^{\hat{t}} \sum_{q \in F_p} (\boldsymbol{a}_p - \boldsymbol{a}_q)'(-\boldsymbol{L}_{pq})(\boldsymbol{a}_p - \boldsymbol{a}_q) \geq 0,
$$

and we find that $\boldsymbol{L}$ is a positive semi-definite matrix.

Consequently, since $\boldsymbol{A}$ is positive definite, $\boldsymbol{L}$ is positive semi-definite, and $\lambda \geq 0$, $\boldsymbol{G} = \boldsymbol{A} + \lambda \boldsymbol{L}$ is positive definite and non-singular.

A.5 Proof of Lemma 2

Let $\boldsymbol{a} = \sum_{p=1}^{\hat{t}} \boldsymbol{e}_p \otimes \boldsymbol{a}_p$ ($\boldsymbol{a}_1, \dots, \boldsymbol{a}_{\hat{t}} \in \mathbb{R}^{k_L}$), $\boldsymbol{b} \in \mathbb{R}^{k_G}$, and $\boldsymbol{x} = (\boldsymbol{a}', \boldsymbol{b}')' \in \mathbb{R}^{\hat{t}k_L + k_G}$. Then, $\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x}$ can be expressed as follows:

$$
\begin{aligned}
\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x} &= (\boldsymbol{a}', \boldsymbol{b}') \begin{pmatrix} \boldsymbol{G} & \boldsymbol{J} \\ \boldsymbol{J}' & \boldsymbol{H} \end{pmatrix} \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \\
&= \boldsymbol{a}'\boldsymbol{G}\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{J}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{J}'\boldsymbol{a} + \boldsymbol{b}\boldsymbol{H}\boldsymbol{b} \\
&= \boldsymbol{a}'\boldsymbol{A}\boldsymbol{a} + \lambda \boldsymbol{a}'\boldsymbol{L}\boldsymbol{a} + 2\boldsymbol{a}'\boldsymbol{J}\boldsymbol{b} + \boldsymbol{b}\boldsymbol{H}\boldsymbol{b} \\
&\geq \boldsymbol{a}'\boldsymbol{A}\boldsymbol{a} + 2\boldsymbol{a}'\boldsymbol{J}\boldsymbol{b} + \boldsymbol{b}\boldsymbol{H}\boldsymbol{b}.
\end{aligned}
$$

In the last inequality, the positive semi-definiteness of $\boldsymbol{L}$ was used. Here, noting that

$$
\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a} = \sum_{p=1}^{\hat{t}} \sum_{j \in E_p} \boldsymbol{a}_p' \boldsymbol{X}_j' \boldsymbol{X}_j \boldsymbol{a}_p, \quad \boldsymbol{a}'\boldsymbol{J}\boldsymbol{b} = \sum_{p=1}^{\hat{t}} \sum_{j \in E_p} \boldsymbol{a}_p' \boldsymbol{X}_j' \boldsymbol{Z}_j \boldsymbol{b}, \quad \boldsymbol{b}'\boldsymbol{H}\boldsymbol{b} = \sum_{p=1}^{\hat{t}} \sum_{j \in E_p} \boldsymbol{b}' \boldsymbol{Z}_j' \boldsymbol{Z}_j \boldsymbol{b},
$$

we have

$$
\begin{aligned}
\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x} &= \sum_{p=1}^{\hat{t}} \sum_{j \in E_p} (\boldsymbol{a}_p' \boldsymbol{X}_j' \boldsymbol{X}_j \boldsymbol{a}_p + 2\boldsymbol{a}_p' \boldsymbol{X}_j' \boldsymbol{Z}_j \boldsymbol{b} + \boldsymbol{b}' \boldsymbol{Z}_j' \boldsymbol{Z}_j \boldsymbol{b}) \\
&= \sum_{p=1}^{\hat{t}} \sum_{j \in E_p} \|\boldsymbol{X}_j \boldsymbol{a}_p + \boldsymbol{Z}_j \boldsymbol{b}\|^2 \geq 0.
\end{aligned}
$$

That is, $\boldsymbol{M}$ is a positive semi-definite matrix.

A.6 Proof of Lemma 3

Let $\boldsymbol{a} \in \mathbb{R}^{\hat{t}k_L}$, $\boldsymbol{b} \in \mathbb{R}^{k_G}$, and $\boldsymbol{x} = (\boldsymbol{a}', \boldsymbol{b}')' \in \mathbb{R}^{\hat{t}k_L+k_G}$. Since $\boldsymbol{M}$ is positive semi-definite from Lemma 2, we find that

$$\begin{aligned}
\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x} &= (\boldsymbol{a}', \boldsymbol{b}') \begin{pmatrix} \boldsymbol{G} & \boldsymbol{J} \\ \boldsymbol{J}' & \boldsymbol{H} \end{pmatrix} \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{pmatrix} \\
&= \boldsymbol{a}'\boldsymbol{G}\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{J}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{J}'\boldsymbol{a} + \boldsymbol{b}\boldsymbol{H}\boldsymbol{b} \\
&= \boldsymbol{a}'\boldsymbol{G}^{1/2}\boldsymbol{G}^{1/2}\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{G}^{1/2}\boldsymbol{G}^{-1/2}\boldsymbol{J}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{J}'\boldsymbol{G}^{-1/2}\boldsymbol{G}^{1/2}\boldsymbol{a} + \boldsymbol{b}'\boldsymbol{H}\boldsymbol{b} \\
&= \|\boldsymbol{G}^{1/2}\boldsymbol{a} + \boldsymbol{G}^{-1/2}\boldsymbol{J}\boldsymbol{b}\|^2 - \boldsymbol{b}'\boldsymbol{J}'\boldsymbol{G}^{-1/2}\boldsymbol{G}^{-1/2}\boldsymbol{J}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{H}\boldsymbol{b} \\
&= \|\boldsymbol{G}^{1/2}\boldsymbol{a} + \boldsymbol{G}^{-1/2}\boldsymbol{J}\boldsymbol{b}\|^2 + \boldsymbol{b}'(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})\boldsymbol{b} \geq 0.
\end{aligned}$$

This inequality must hold for any $\boldsymbol{a}$ and $\boldsymbol{b}$. Setting $\boldsymbol{a} = \boldsymbol{G}^{-1}\boldsymbol{J}\boldsymbol{b}$ gives $\|\boldsymbol{G}^{1/2}\boldsymbol{a} + \boldsymbol{G}^{-1/2}\boldsymbol{J}\boldsymbol{b}\|^2 = 0$. This means that $\boldsymbol{b}'(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})\boldsymbol{b} \geq 0$ must hold. Therefore, $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$ is positive semi-definite.

A.7 Proof of Lemma 4

From Lemma 1, $\boldsymbol{G}$ is positive definite. Hence, $\boldsymbol{G}^{-1}$ is also positive definite. From Lemma 3, we know that $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$ is positive semi-definite. In particular, if $\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J}$ is regular, then it is positive definite. Therefore, $(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}$ is also positive definite. Thus, the following holds for $\boldsymbol{a} \in \mathbb{R}^{\hat{t}k_L} \setminus \{\boldsymbol{0}_{\hat{t}k_L}\}$.

$$\begin{aligned}
\boldsymbol{a}'\boldsymbol{M}_{11}\boldsymbol{a} &= \boldsymbol{a}'\{\boldsymbol{G}^{-1} + \boldsymbol{G}^{-1}\boldsymbol{J}(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{J}'\boldsymbol{G}^{-1}\}\boldsymbol{a} \\
&= \boldsymbol{a}'\boldsymbol{G}^{-1}\boldsymbol{a} + \boldsymbol{a}'\boldsymbol{G}^{-1}\boldsymbol{J}(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{a} \\
&> \boldsymbol{a}'\boldsymbol{G}^{-1}\boldsymbol{J}(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{a}.
\end{aligned}$$

Letting $\boldsymbol{b} = \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{a}$, we can write $\boldsymbol{a}'\boldsymbol{M}_{11}\boldsymbol{a} > \boldsymbol{b}'(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{b}$. When $\boldsymbol{b} \neq \boldsymbol{0}_{k_G}$, due to the positive definiteness of $(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}$, we have $\boldsymbol{b}'(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{b} > 0$. Also, when $\boldsymbol{b} = \boldsymbol{0}_{k_G}$, we have $\boldsymbol{b}'(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{b} = 0$. Therefore, we can conclude that $\boldsymbol{a}'\boldsymbol{M}_{11}\boldsymbol{a} > \boldsymbol{b}'(\boldsymbol{H} - \boldsymbol{J}'\boldsymbol{G}^{-1}\boldsymbol{J})^{-1}\boldsymbol{b} \geq 0$, and thus $\boldsymbol{M}_{11}$ is positive definite.

B Proofs of formulas in Section 3

B.1 Proof of Lemma 5

First, we will derive $\partial\boldsymbol{g}'/\partial\boldsymbol{y}_j$. Since $\boldsymbol{A}$ is symmetric matrix, we have

$$\frac{\partial\boldsymbol{g}'}{\partial\boldsymbol{y}_j} = \frac{\partial\hat{\boldsymbol{\xi}}'}{\partial\boldsymbol{y}_j}\boldsymbol{A} + \frac{\partial\hat{\boldsymbol{\gamma}}'}{\partial\boldsymbol{y}_j}\boldsymbol{B}' - \frac{\partial\boldsymbol{c}'}{\partial\boldsymbol{y}_j} + \lambda\frac{\partial\boldsymbol{\theta}'}{\partial\boldsymbol{y}_j}.$$

Note that $j \in E_s$. Then, $\partial\boldsymbol{c}'/\partial\boldsymbol{y}_j$ is given by the following:

$$\frac{\partial\boldsymbol{c}'}{\partial\boldsymbol{y}_j} = \sum_{p=1}^{\hat{t}}\left(\boldsymbol{e}_p' \otimes \sum_{r\in E_p}\frac{\partial\boldsymbol{y}_r'}{\partial\boldsymbol{y}_j}\boldsymbol{X}_r\right) = \boldsymbol{e}_s' \otimes \boldsymbol{X}_j.$$

Additionally, $\partial\boldsymbol{\theta}'/\partial\boldsymbol{y}_j$ is given by the following:

$$\begin{aligned}
\frac{\partial\boldsymbol{\theta}'}{\partial\boldsymbol{y}_j} &= \sum_{s=1}^{\hat{t}}\left(\boldsymbol{e}_s' \otimes \sum_{u\in F_s}\frac{\partial v_{su}\boldsymbol{\theta}_{su}'}{\partial\boldsymbol{y}_j}\right) \\
&= \sum_{s=1}^{\hat{t}}\left\{\boldsymbol{e}_s' \otimes \sum_{u\in F_s}\left(\frac{\partial v_{su}}{\partial\boldsymbol{y}_j}\boldsymbol{\theta}_{su}' + v_{su}\frac{\partial\boldsymbol{\theta}_{su}'}{\partial\boldsymbol{y}_j}\right)\right\}
\end{aligned}$$

$$= \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s' \otimes \sum_{u \in F_s} \frac{\partial v_{su}}{\partial \boldsymbol{y}_j} \boldsymbol{\theta}_{su}' \right) + \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s' \otimes \sum_{u \in F_s} v_{su} \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \frac{\partial \boldsymbol{\theta}_{su}'}{\partial \hat{\boldsymbol{\xi}}} \right)$$
$$= \boldsymbol{\Phi}_j + \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s' \otimes \sum_{u \in F_s} v_{su} \frac{\partial \boldsymbol{\theta}_{su}'}{\partial \hat{\boldsymbol{\xi}}} \right),$$

where $\boldsymbol{\Phi}_j$ is a matrix defined by (17). Moreover, using similar calculation to (22), we obtain

$$\sum_{s=1}^{\hat{t}} \left(\boldsymbol{e}_s' \otimes \sum_{u \in F_s} v_{su} \frac{\partial \boldsymbol{\theta}_{su}'}{\partial \hat{\boldsymbol{\xi}}} \right) = \boldsymbol{L}.$$

Therefore, the following holds:

$$\frac{\partial \boldsymbol{\theta}'}{\partial \boldsymbol{y}_j} = \boldsymbol{\Phi}_j + \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{L}.$$

Consequently, we have

$$\frac{\partial \boldsymbol{g}'}{\partial \boldsymbol{y}_j} = \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{A} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{B}' - \boldsymbol{e}_s' \otimes \boldsymbol{X}_j + \lambda \left(\boldsymbol{\Phi}_j + \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{L} \right)$$
$$= \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} (\boldsymbol{A} + \lambda \boldsymbol{L}) + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{B}' - \boldsymbol{e}_s' \otimes \boldsymbol{X}_j + \lambda \boldsymbol{\Phi}_j$$
$$= \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{G} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{J}' - \boldsymbol{e}_s' \otimes \boldsymbol{X}_j + \lambda \boldsymbol{\Phi}_j.$$

Next, we will derive $\partial \boldsymbol{h}'/\partial \boldsymbol{y}_j$. We note that $\boldsymbol{Z}'\boldsymbol{y} = \sum_{r=1}^m \boldsymbol{Z}_r' \boldsymbol{y}_r$, $\partial \boldsymbol{y}_j'/\partial \boldsymbol{y}_j = \boldsymbol{I}_{n_j}$, and $\partial \boldsymbol{y}_r'/\partial \boldsymbol{y}_j = \boldsymbol{O}_{n_j,n_r}$ $(r \neq j)$. Then, we have

$$\frac{\partial \boldsymbol{h}'}{\partial \boldsymbol{y}_j} = \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{B} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{Z}'\boldsymbol{Z} - \sum_{r=1}^m \frac{\partial \boldsymbol{y}_r'}{\partial \boldsymbol{y}_j} \boldsymbol{Z}_r$$
$$= \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{J} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{H} - \boldsymbol{Z}_j.$$

B.2 Proof of Lemma 6

From $\partial \boldsymbol{g}'/\partial \boldsymbol{y}_j = \boldsymbol{O}_{n_j,\hat{t}k_L}$, $\partial \boldsymbol{h}'/\partial \boldsymbol{y}_j = \boldsymbol{O}_{n_j,k_G}$, and Lemma 5, we have

$$\left(\frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}, \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \right) \begin{pmatrix} \boldsymbol{G} \\ \boldsymbol{J}' \end{pmatrix} = \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{G} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{J}' = \boldsymbol{e}_s' \otimes \boldsymbol{X}_j - \lambda \boldsymbol{\Phi}_j,$$
$$\left(\frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}, \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \right) \begin{pmatrix} \boldsymbol{J} \\ \boldsymbol{H} \end{pmatrix} = \frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j} \boldsymbol{J} + \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \boldsymbol{H} = \boldsymbol{Z}_j.$$

Therefore,

$$\left(\frac{\partial \hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}, \frac{\partial \hat{\gamma}'}{\partial \boldsymbol{y}_j} \right) \begin{pmatrix} \boldsymbol{G} & \boldsymbol{J} \\ \boldsymbol{J}' & \boldsymbol{H} \end{pmatrix} = \left(\boldsymbol{e}_s' \otimes \boldsymbol{X}_j - \lambda \boldsymbol{\Phi}_j, \, \boldsymbol{Z}_j \right).$$

Multiplying both sides from the left by $\boldsymbol{M}^{-1}$, we obtain the desired result.

B.3 Proof of Theorem 1

From (16) and Lemma 6, we have

$$\begin{aligned}
\widehat{\mathrm{df}}(\boldsymbol{W},\lambda) &= \sum_{s=1}^{\hat t}\sum_{j\in E_s}\mathrm{tr}\left(\frac{\partial\hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j') + \frac{\partial\hat{\boldsymbol{\gamma}}'}{\partial \boldsymbol{y}_j}\boldsymbol{Z}_j'\right)\\
&= \sum_{s=1}^{\hat t}\sum_{j\in E_s}\mathrm{tr}\left(\left(\frac{\partial\hat{\boldsymbol{\xi}}'}{\partial \boldsymbol{y}_j},\, \frac{\partial\hat{\boldsymbol{\gamma}}'}{\partial \boldsymbol{y}_j}\right)\begin{pmatrix}\boldsymbol{e}_s\otimes \boldsymbol{X}_j'\\ \boldsymbol{Z}_j'\end{pmatrix}\right)\\
&= \sum_{s=1}^{\hat t}\sum_{j\in E_s}\mathrm{tr}\left((\boldsymbol{e}_s'\otimes \boldsymbol{X}_j - \lambda\boldsymbol{\Phi}_j,\, \boldsymbol{Z}_j)\,\boldsymbol{M}^{-1}\begin{pmatrix}\boldsymbol{e}_s\otimes \boldsymbol{X}_j'\\ \boldsymbol{Z}_j'\end{pmatrix}\right)\\
&= \sum_{s=1}^{\hat t}\sum_{j\in E_s}\mathrm{tr}\left(\boldsymbol{M}^{-1}\begin{pmatrix}\boldsymbol{e}_s\otimes \boldsymbol{X}_j'\\ \boldsymbol{Z}_j'\end{pmatrix}(\boldsymbol{e}_s'\otimes \boldsymbol{X}_j - \lambda\boldsymbol{\Phi}_j,\, \boldsymbol{Z}_j)\right)\\
&= \mathrm{tr}\left(\boldsymbol{M}^{-1}\sum_{s=1}^{\hat t}\sum_{j\in E_s}\begin{pmatrix}\boldsymbol{e}_s\otimes \boldsymbol{X}_j'\\ \boldsymbol{Z}_j'\end{pmatrix}(\boldsymbol{e}_s'\otimes \boldsymbol{X}_j - \lambda\boldsymbol{\Phi}_j,\, \boldsymbol{Z}_j)\right).
\end{aligned}$$

Since $\boldsymbol{A} = \boldsymbol{G} - \lambda\boldsymbol{L}$, we find that

$$\begin{aligned}
&\sum_{s=1}^{\hat t}\sum_{j\in E_s}\begin{pmatrix}\boldsymbol{e}_s\otimes \boldsymbol{X}_j'\\ \boldsymbol{Z}_j'\end{pmatrix}(\boldsymbol{e}_s'\otimes \boldsymbol{X}_j - \lambda\boldsymbol{\Phi}_j,\, \boldsymbol{Z}_j)\\
&= \sum_{s=1}^{\hat t}\sum_{j\in E_s}\begin{pmatrix}\boldsymbol{e}_s\boldsymbol{e}_s'\otimes \boldsymbol{X}_j'\boldsymbol{X}_j & \boldsymbol{e}_s\otimes \boldsymbol{X}_j'\boldsymbol{Z}_j\\ \boldsymbol{e}_s'\otimes \boldsymbol{Z}_j'\boldsymbol{X}_j & \boldsymbol{Z}_j'\boldsymbol{Z}_j\end{pmatrix} - \lambda\sum_{s=1}^{\hat t}\sum_{j\in E_s}\begin{pmatrix}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\boldsymbol{\Phi}_j & \boldsymbol{O}_{\hat t k_L, k_G}\\ \boldsymbol{Z}_j'\boldsymbol{\Phi}_j & \boldsymbol{O}_{k_G,k_G}\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{B}' & \boldsymbol{Z}'\boldsymbol{Z}\end{pmatrix} - \lambda\begin{pmatrix}\sum_{s=1}^{\hat t}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\boldsymbol{\Phi}_j & \boldsymbol{O}_{\hat t k_L, k_G}\\ \sum_{s=1}^{\hat t}\sum_{j\in E_s}\boldsymbol{Z}_j\boldsymbol{\Phi}_j & \boldsymbol{O}_{k_G,k_G}\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{G} & \boldsymbol{J}\\ \boldsymbol{J}' & \boldsymbol{H}\end{pmatrix} - \lambda\begin{pmatrix}\boldsymbol{L} & \boldsymbol{O}_{\hat t k_L, k_G}\\ \boldsymbol{O}_{k_G,\hat t k_L} & \boldsymbol{O}_{k_G,k_G}\end{pmatrix} - \lambda\begin{pmatrix}\sum_{s=1}^{\hat t}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\boldsymbol{\Phi}_j & \boldsymbol{O}_{\hat t k_L, k_G}\\ \sum_{s=1}^{\hat t}\sum_{j\in E_s}\boldsymbol{Z}_j\boldsymbol{\Phi}_j & \boldsymbol{O}_{k_G,k_G}\end{pmatrix}\\
&= \boldsymbol{M} - \lambda\begin{pmatrix}\boldsymbol{L} & \boldsymbol{O}_{\hat t k_L, k_G}\\ \boldsymbol{O}_{k_G,\hat t k_L} & \boldsymbol{O}_{k_G,k_G}\end{pmatrix} - \lambda\begin{pmatrix}\sum_{s=1}^{\hat t}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\boldsymbol{\Phi}_j & \boldsymbol{O}_{\hat t k_L, k_G}\\ \sum_{j=1}^{m}\boldsymbol{Z}_j\boldsymbol{\Phi}_j & \boldsymbol{O}_{k_G,k_G}\end{pmatrix}.
\end{aligned}$$

Here, we note that

$$\begin{aligned}
\sum_{s=1}^{\hat t}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\boldsymbol{\Phi}_j &= \sum_{s=1}^{\hat t}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\left(\sum_{p=1}^{\hat t}\boldsymbol{e}_p'\otimes\sum_{u\in F_p}\frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}'\right)\\
&= \sum_{s=1}^{\hat t}\sum_{p=1}^{\hat t}\sum_{j\in E_s}(\boldsymbol{e}_s\otimes \boldsymbol{X}_j')\left(\boldsymbol{e}_p'\otimes\sum_{u\in F_p}\frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}'\right)\\
&= \sum_{s=1}^{\hat t}\sum_{p=1}^{\hat t}\sum_{j\in E_s}\left(\boldsymbol{e}_s\boldsymbol{e}_p'\otimes\sum_{u\in F_p}\boldsymbol{X}_j'\frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}'\right)\\
&= \sum_{s=1}^{\hat t}\sum_{p=1}^{\hat t}\left(\boldsymbol{e}_s\boldsymbol{e}_p'\otimes\sum_{u\in F_p}\sum_{j\in E_s}\boldsymbol{X}_j'\frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}'\right) = \boldsymbol{R},
\end{aligned}$$

and

$$\begin{aligned}\sum_{j=1}^{m} \boldsymbol{Z}_j \boldsymbol{\Phi}_j &= \sum_{j=1}^{m} \boldsymbol{Z}_j \left(\sum_{p=1}^{\hat{t}} \boldsymbol{e}'_p \otimes \sum_{u \in F_p} \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j} \boldsymbol{\theta}'_{pu} \right) \\ &= \sum_{p=1}^{\hat{t}} \left(\boldsymbol{e}'_p \otimes \sum_{u \in F_p} \sum_{j=1}^{m} \boldsymbol{Z}_j \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j} \boldsymbol{\theta}'_{pu} \right) = \boldsymbol{U},\end{aligned}$$

where $\boldsymbol{R}$ and $\boldsymbol{U}$ are the matrices defined by (18). Hence, we have

$$\sum_{s=1}^{\hat{t}} \sum_{j \in E_s} \begin{pmatrix} \boldsymbol{e}_s \otimes \boldsymbol{X}'_j \\ \boldsymbol{Z}'_j \end{pmatrix} \begin{pmatrix} \boldsymbol{e}'_s \otimes \boldsymbol{X}_j - \lambda \boldsymbol{\Phi}_j & \boldsymbol{Z}_j \end{pmatrix} = \boldsymbol{M} - \lambda \begin{pmatrix} \boldsymbol{L} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{O}_{k_G, \hat{t}k_L} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} - \lambda \begin{pmatrix} \boldsymbol{R} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{U} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix}.$$

Consequently, $\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda)$ is given by the following:

$$\begin{aligned}\widehat{\mathrm{df}}(\boldsymbol{W}, \lambda) &= \mathrm{tr} \left(\boldsymbol{M}^{-1} \left\{ \boldsymbol{M} - \lambda \begin{pmatrix} \boldsymbol{L} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{O}_{k_G, \hat{t}k_L} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} - \lambda \begin{pmatrix} \boldsymbol{R} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{U} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} \right\} \right) \\ &= \mathrm{tr} \left(\boldsymbol{I}_{\hat{t}k_L + k_G} - \lambda \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{L} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{O}_{k_G, \hat{t}k_L} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} - \lambda \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{R} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{U} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} \right) \\ &= \mathrm{tr}(\boldsymbol{I}_{\hat{t}k_L + k_G}) - \lambda \mathrm{tr} \begin{pmatrix} \boldsymbol{M}_{11}\boldsymbol{L} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{M}_{21}\boldsymbol{L} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} - \lambda \mathrm{tr} \begin{pmatrix} \boldsymbol{M}_{11}\boldsymbol{R} + \boldsymbol{M}_{12}\boldsymbol{U} & \boldsymbol{O}_{\hat{t}k_L, k_G} \\ \boldsymbol{M}_{21}\boldsymbol{R} + \boldsymbol{M}_{22}\boldsymbol{U} & \boldsymbol{O}_{k_G, k_G} \end{pmatrix} \\ &= \hat{t}k_L + k_G - \lambda \mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) - \lambda \mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{R} + \boldsymbol{M}_{12}\boldsymbol{U}) \\ &= \hat{t}k_L + k_G - \lambda \mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) - \lambda \mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{R}) - \lambda \mathrm{tr}(\boldsymbol{M}_{12}\boldsymbol{U}).\end{aligned}$$

B.4 Proof of Lemma 7

When $\delta = 0$, $w_{r\ell}^{\mathrm{AL}}(0) = \|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^0 = 1$. Hence, we have

$$\frac{\partial w_{r\ell}^{\mathrm{AL}}(0)}{\partial \boldsymbol{y}_j} = \boldsymbol{0}_{n_j}.$$

When $\delta > 0$, we have

$$\begin{aligned}\frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j} &= \frac{\partial}{\partial \boldsymbol{y}_j} (\|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^2)^{-\delta/2} \\ &= -\frac{\delta}{2} (\|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^2)^{-\delta/2-1} \frac{\partial}{\partial \boldsymbol{y}_j} \|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^2 \\ &= -\frac{\delta}{2} \|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-\delta-2} \left\{ 2 \left(\frac{\partial}{\partial \boldsymbol{y}_j} (\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}})' \right) (\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}) \right\} \\ &= -\delta \|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-\delta-1} \left(\frac{\partial \hat{\boldsymbol{\beta}}'_{r,\mathrm{LS}}}{\partial \boldsymbol{y}_j} - \frac{\partial \hat{\boldsymbol{\beta}}'_{\ell,\mathrm{LS}}}{\partial \boldsymbol{y}_j} \right) \frac{\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}}{\|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|} \\ &= -\delta (\|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-\delta})^{1+\delta^{-1}} \left(\frac{\partial \hat{\boldsymbol{\beta}}'_{r,\mathrm{LS}}}{\partial \boldsymbol{y}_j} - \frac{\partial \hat{\boldsymbol{\beta}}'_{\ell,\mathrm{LS}}}{\partial \boldsymbol{y}_j} \right) \boldsymbol{a}_{r\ell} \\ &= -\delta (\|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-\delta})^{1+\delta^{-1}} \left(\frac{\partial \hat{\boldsymbol{\beta}}'_{r,\mathrm{LS}}}{\partial \boldsymbol{y}_j} - \frac{\partial \hat{\boldsymbol{\beta}}'_{\ell,\mathrm{LS}}}{\partial \boldsymbol{y}_j} \right) \boldsymbol{a}_{r\ell} \\ &= -\delta w_{r\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}} \left(\frac{\partial \hat{\boldsymbol{\beta}}'_{r,\mathrm{LS}}}{\partial \boldsymbol{y}_j} - \frac{\partial \hat{\boldsymbol{\beta}}'_{\ell,\mathrm{LS}}}{\partial \boldsymbol{y}_j} \right) \boldsymbol{a}_{r\ell},\end{aligned}$$

where $\boldsymbol{a}_{r\ell} = \|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-1}(\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}} - \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}})$.

B.5 Proof of Lemma 8

First, we derive $\partial\hat{\boldsymbol{\beta}}'_{r,\mathrm{LS}}/\partial\boldsymbol{y}_j$. Noting that

$$\frac{\partial\boldsymbol{y}'_r}{\partial\boldsymbol{y}_j}=\begin{cases}\boldsymbol{I}_{n_j} & (r=j)\\ \boldsymbol{O}_{n_j,n_r} & (r\neq j)\end{cases},$$

we have

$$\begin{aligned}\frac{\partial\hat{\boldsymbol{\beta}}'_{r,\mathrm{LS}}}{\partial\boldsymbol{y}_j}&=\frac{\partial}{\partial\boldsymbol{y}_j}\left\{(\boldsymbol{y}_r-\boldsymbol{Z}_r\hat{\gamma}_{\mathrm{LS}})'\boldsymbol{X}_r(\boldsymbol{X}'_r\boldsymbol{X}_r)^{-1}\right\}\\&=\frac{\partial\boldsymbol{y}'_r}{\partial\boldsymbol{y}_j}\boldsymbol{X}_r(\boldsymbol{X}'_r\boldsymbol{X}_r)^{-1}-\frac{\partial\hat{\gamma}'_{\mathrm{LS}}}{\partial\boldsymbol{y}_j}\boldsymbol{Z}'_r\boldsymbol{X}_r(\boldsymbol{X}'_r\boldsymbol{X}_r)^{-1}\\&=\begin{cases}\boldsymbol{X}_j(\boldsymbol{X}'_j\boldsymbol{X}_j)^{-1}-\dfrac{\partial\hat{\gamma}'_{\mathrm{LS}}}{\partial\boldsymbol{y}_j}\boldsymbol{Z}'_j\boldsymbol{X}_j(\boldsymbol{X}'_j\boldsymbol{X}_j)^{-1} & (r=j)\\ -\dfrac{\partial\hat{\gamma}'_{\mathrm{LS}}}{\partial\boldsymbol{y}_j}\boldsymbol{Z}'_r\boldsymbol{X}_r(\boldsymbol{X}'_r\boldsymbol{X}_r)^{-1} & (r\neq j)\end{cases}.\end{aligned}$$

Next, we derive $\partial\hat{\gamma}'_{\mathrm{LS}}/\partial\boldsymbol{y}_j$. Noting that $\boldsymbol{Z}'(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}=\sum_{r=1}^m\boldsymbol{Z}'_r(\boldsymbol{I}_{n_r}-\boldsymbol{P}_{\boldsymbol{X}_r})\boldsymbol{y}_r$, we have

$$\begin{aligned}\frac{\partial\hat{\gamma}'_{\mathrm{LS}}}{\partial\boldsymbol{y}_j}&=\frac{\partial}{\partial\boldsymbol{y}_j}\left[\boldsymbol{y}'(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\{\boldsymbol{Z}'(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1}\right]\\&=\left\{\sum_{r=1}^m\frac{\partial\boldsymbol{y}'_r}{\partial\boldsymbol{y}_j}(\boldsymbol{I}_{n_r}-\boldsymbol{P}_{\boldsymbol{X}_r})\boldsymbol{Z}_r\right\}\{\boldsymbol{Z}'(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1}\\&=(\boldsymbol{I}_{n_j}-\boldsymbol{P}_{\boldsymbol{X}_j})\boldsymbol{Z}_j\{\boldsymbol{Z}'(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1}.\end{aligned}$$

B.6 Proof of Lemma 9

First, we show the case of $\delta=0$. According to Lemma 7, we find that $\partial w^{\mathrm{AL}}_{r\ell}(0)/\partial\boldsymbol{y}_j=\boldsymbol{0}_{n_j}$. Therefore, for any $s,p\in\{1,\dots,\hat{t}\}$, we obtain

$$\boldsymbol{R}_{sp,(\boldsymbol{W}_{\mathrm{AL}}(0),\lambda)}=\boldsymbol{O}_{k_L,k_L}=-0\times\boldsymbol{R}^{\mathrm{AL}}_{sp}(0,\lambda),$$

where $\boldsymbol{R}^{\mathrm{AL}}_{sp}$ is a matrix defined by (19). Consequently, $\boldsymbol{R}_{(\boldsymbol{W}_{\mathrm{AL}}(0),\lambda)}=-0\times\boldsymbol{R}_{\mathrm{AL}}(0,\lambda)$, where $\boldsymbol{R}_{\mathrm{AL}}$ is a matrix defined by (19).

Next, we show the case of $\delta>0$. Let $j,r\in\{1,\dots,m\}$ and $\ell\in D_r$. Since $\ell\in D_r\subseteq\{1,\dots,m\}\setminus\{r\}$, $\ell\neq r$. Thus, note that $r=j$ and $\ell=j$ cannot occur at the same time. We also note that $\boldsymbol{X}'_j(\boldsymbol{I}_{n_j}-\boldsymbol{P}_{\boldsymbol{X}_j})=\boldsymbol{O}_{k_L,n_j}$ and that

$$\begin{aligned}\boldsymbol{a}_{r\ell}&=\|\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}-\hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}\|^{-1}(\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}-\hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}})\\&=-\|\hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}-\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}\|^{-1}(\hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}-\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}})=-\boldsymbol{a}_{\ell r}.\end{aligned}$$

According to Lemma 8, we have

$$\boldsymbol{X}'_j\frac{\partial\hat{\gamma}'_{\mathrm{LS}}}{\partial\boldsymbol{y}_j}=\boldsymbol{O}_{k_L,k_L},\quad\boldsymbol{X}'_j\frac{\partial\hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}}{\partial\boldsymbol{y}_j}=\begin{cases}\boldsymbol{I}_{k_L} & (r=j)\\ \boldsymbol{O}_{k_L,k_L} & (r\neq j)\end{cases}.$$

Hence, we obtain the following:

$$\boldsymbol{X}'_j\frac{\partial w^{\mathrm{AL}}_{r\ell}(\delta)}{\partial\boldsymbol{y}_j}=\begin{cases}-\delta w^{\mathrm{AL}}_{j\ell}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{j\ell} & (r=j\wedge\ell\neq j)\\ -\delta w^{\mathrm{AL}}_{jr}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{jr} & (r\neq j\wedge\ell=j)\\ \boldsymbol{0}_{k_L} & (r\neq j\wedge\ell\neq j)\end{cases}.$$

Here, to find $\boldsymbol{R}_{sp,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}$, we consider cases with respect to p.

Let $p = s$. If $u \in F_s$, then $u \neq s$ and $E_s \cap E_u = \emptyset$ hold. That is, if $j \in E_s$, then $j \notin E_u$ for all $u \in F_s$. Therefore, we have

$$\begin{aligned}
\boldsymbol{R}_{ss,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} &= \sum_{u\in F_s}\sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{su}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{su}' \\
&= \sum_{u\in F_s}\sum_{j\in E_s} \boldsymbol{X}_j' \sum_{r\in E_s}\sum_{\ell\in E_u\cap D_r} \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{su}' \\
&= \sum_{u\in F_s}\sum_{j\in E_s} \left\{ \sum_{\ell\in E_u\cap D_j} \boldsymbol{X}_j' \frac{\partial w_{j\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j} + \sum_{r\in E_s\setminus\{j\}}\sum_{\ell\in E_u\cap D_r} \boldsymbol{X}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j} \right\} \boldsymbol{\theta}_{su}' \\
&= -\delta \sum_{u\in F_s}\sum_{j\in E_s}\sum_{\ell\in E_u\cap D_j} w_{j\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}} \boldsymbol{a}_{j\ell}\boldsymbol{\theta}_{su}' = -\delta \boldsymbol{R}_{ss}^{\mathrm{AL}}(\delta,\lambda),
\end{aligned}$$

where $\boldsymbol{R}_{ss}^{\mathrm{AL}}$ is a matrix defined by (19).

Let $p \in F_s$. Since $p \neq s$, if $j \in E_s$, then $j \notin E_p$. Noting that $s \in F_p$, $v_{ps} = v_{sp}$, and $\boldsymbol{\theta}_{ps} = -\boldsymbol{\theta}_{sp}$, we have

$$\begin{aligned}
\boldsymbol{R}_{sp,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} &= \sum_{u\in F_p}\sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= \sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{ps}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{ps}' + \sum_{u\in F_p\setminus\{s\}}\sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= -\sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{sp}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{sp}' + \sum_{u\in F_p\setminus\{s\}}\sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= -\sum_{j\in E_s}\sum_{r\in E_s}\sum_{\ell\in E_p\cap D_r} \boldsymbol{X}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{sp}' + \sum_{u\in F_p\setminus\{s\}}\sum_{j\in E_s}\sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r} \boldsymbol{X}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= -\sum_{j\in E_s} \left\{ \sum_{\ell\in E_p\cap D_j} \boldsymbol{X}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{sp}' + \sum_{r\in E_s\setminus\{j\}}\sum_{\ell\in E_p\cap D_r} \boldsymbol{X}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{sp}' \right\} + \boldsymbol{O}_{k_L,k_L} \\
&= -\sum_{j\in E_s} \left\{ -\delta \sum_{\ell\in E_p\cap D_j} w_{j\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}} \boldsymbol{a}_{j\ell}\boldsymbol{\theta}_{sp}' + \boldsymbol{O}_{k_L,k_L} \right\} \\
&= -\delta \left\{ -\sum_{j\in E_s}\sum_{\ell\in E_p\cap D_j} w_{j\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}} \boldsymbol{a}_{j\ell}\boldsymbol{\theta}_{sp}' \right\} = -\delta \boldsymbol{R}_{sp}^{\mathrm{AL}}(\delta,\lambda),
\end{aligned}$$

where $\boldsymbol{R}_{sp}^{\mathrm{AL}}$ is a matrix defined by (19).

Let $p \in \{1,\ldots,\hat{t}\} \setminus (\{s\} \cup F_s)$. Since $E_p \cap E_s = \emptyset$, if $j \in E_s$, then $j \notin E_p$. Additionally, since $s \notin F_p$, if $u \in F_p$, then $u \neq s$. That is, $j \notin E_u$ for all $u \in F_p$. Therefore, we have

$$\begin{aligned}
\boldsymbol{R}_{sp,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} &= \sum_{u\in F_p}\sum_{j\in E_s} \boldsymbol{X}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= \sum_{u\in F_p}\sum_{j\in E_s} \boldsymbol{X}_j' \sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r} \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= \sum_{u\in F_p}\sum_{j\in E_s}\sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r} \boldsymbol{X}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j}\boldsymbol{\theta}_{pu}' \\
&= \boldsymbol{O}_{k_L,k_L} = -\delta \boldsymbol{R}_{sp}^{\mathrm{AL}}(\delta,\lambda),
\end{aligned}$$

where $\boldsymbol{R}_{sp}^{\mathrm{AL}}$ is a matrix defined by (19).

Consequently, when $\delta > 0$, we have

$$\begin{aligned}\boldsymbol{R}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} &= \sum_{s=1}^{\hat{t}}\sum_{p=1}^{\hat{t}} \boldsymbol{e}_s\boldsymbol{e}_p' \otimes \boldsymbol{R}_{sp,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} \\ &= \sum_{s=1}^{\hat{t}}\sum_{p=1}^{\hat{t}} \boldsymbol{e}_s\boldsymbol{e}_p' \otimes \left\{-\delta \boldsymbol{R}_{sp}^{\mathrm{AL}}(\delta,\lambda)\right\} \\ &= -\delta \sum_{s=1}^{\hat{t}}\sum_{p=1}^{\hat{t}} \boldsymbol{e}_s\boldsymbol{e}_p' \otimes \boldsymbol{R}_{sp}^{\mathrm{AL}}(\delta,\lambda) = -\delta \boldsymbol{R}_{\mathrm{AL}}(\delta,\lambda),\end{aligned}$$

where $\boldsymbol{R}_{\mathrm{AL}}$ is a matrix defined by (19).

B.7 Proof of Lemma 10

According to Lemma 7, when $\delta = 0$, $\partial w_{r\ell}^{\mathrm{AL}}(0)/\partial \boldsymbol{y}_j = \boldsymbol{0}_{n_j}$. Therefore, we have $\boldsymbol{U}_{p,(\boldsymbol{W}_{\mathrm{AL}}(0),\lambda)} = \boldsymbol{O}_{k_G,k_L}$ and $\boldsymbol{U}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} = \boldsymbol{O}_{k_G,\hat{t}k_L}$.

According to Lemma 8, when $\delta > 0$, for all $r \in \{1,\dots,m\}$, we have

$$\begin{aligned}\sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial \hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} &= \boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} - \sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial \hat{\boldsymbol{\gamma}}_{\mathrm{LS}}'}{\partial \boldsymbol{y}_j} \boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} \\ &= \left(\boldsymbol{I}_{k_G} - \sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial \hat{\boldsymbol{\gamma}}_{\mathrm{LS}}'}{\partial \boldsymbol{y}_j}\right) \boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} \\ &= \left[\boldsymbol{I}_{k_G} - \sum_{j=1}^{m} \boldsymbol{Z}_j'(\boldsymbol{I}_{n_j} - \boldsymbol{P}_{\boldsymbol{X}_j})\boldsymbol{Z}_j\{\boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1}\right] \boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} \\ &= \left[\boldsymbol{I}_{k_G} - \boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\{\boldsymbol{Z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Z}\}^{-1}\right] \boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} \\ &= (\boldsymbol{I}_{k_G} - \boldsymbol{I}_{k_G})\boldsymbol{Z}_r'\boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1} \\ &= \boldsymbol{O}_{k_G,k_L}.\end{aligned}$$

Hence, for $p \in \{1,\dots,\hat{t}\}$ and $u \in F_p$, we obtain

$$\begin{aligned}\sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial v_{pu}}{\partial \boldsymbol{y}_j} &= \sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r}\sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial w_{r\ell}^{\mathrm{AL}}(\delta)}{\partial \boldsymbol{y}_j} \\ &= \sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r}\sum_{j=1}^{m} \boldsymbol{Z}_j' \left\{-\delta w_{r\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}} \left(\frac{\partial \hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} - \frac{\partial \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}'}{\partial \boldsymbol{y}_j}\right)\boldsymbol{a}_{r\ell}\right\} \\ &= \sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r}\sum_{j=1}^{m} \boldsymbol{Z}_j' \left(\frac{\partial \hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} - \frac{\partial \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}'}{\partial \boldsymbol{y}_j}\right)\left(-\delta w_{r\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{r\ell}\right) \\ &= \sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r} \left(\sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial \hat{\boldsymbol{\beta}}_{r,\mathrm{LS}}'}{\partial \boldsymbol{y}_j} - \sum_{j=1}^{m} \boldsymbol{Z}_j' \frac{\partial \hat{\boldsymbol{\beta}}_{\ell,\mathrm{LS}}'}{\partial \boldsymbol{y}_j}\right)\left(-\delta w_{r\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{r\ell}\right) \\ &= \sum_{r\in E_p}\sum_{\ell\in E_u\cap D_r} \boldsymbol{O}_{k_G,k_L}\left(-\delta w_{r\ell}^{\mathrm{AL}}(\delta)^{1+\delta^{-1}}\boldsymbol{a}_{r\ell}\right) \\ &= \boldsymbol{0}_{k_G}.\end{aligned}$$

Consequently, we have $\boldsymbol{U}_{p,(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} = \boldsymbol{O}_{k_G,k_L}$ and $\boldsymbol{U}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)} = \boldsymbol{O}_{k_G,\hat{t}k_L}$.

B.8 Proof of Theorem 2

By substituting the results of Lemma 9 and Lemma 10 into Theorem 1, we obtain

$$\begin{aligned}\nu_{\mathrm{AL}}(\delta,\lambda) &= \widehat{\mathrm{df}}(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda) \\ &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) - \lambda\mathrm{tr}\left(\boldsymbol{M}_{11}\boldsymbol{R}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}\right) - \lambda\mathrm{tr}\left(\boldsymbol{M}_{12}\boldsymbol{U}_{(\boldsymbol{W}_{\mathrm{AL}}(\delta),\lambda)}\right) \\ &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) + \lambda\delta\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{R}_{\mathrm{AL}}).\end{aligned}$$

B.9 Proof of Proposition 1

(1). According to Lemma 1 and Lemma 4, respectively, $\boldsymbol{L}$ is a positive semi-definite matrix and $\boldsymbol{M}$ is a positive definite matrix. Therefore, we find that

$$\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) = \mathrm{tr}\left(\boldsymbol{L}^{1/2}\boldsymbol{M}_{11}^{1/2}\boldsymbol{M}_{11}^{1/2}\boldsymbol{L}^{1/2}\right) = \mathrm{tr}\left((\boldsymbol{M}_{11}^{1/2}\boldsymbol{L}^{1/2})'(\boldsymbol{M}_{11}^{1/2}\boldsymbol{L}^{1/2})\right) \geq 0,$$

which implies $\nu_{\mathrm{AL}}^{(1)}(\delta,\lambda) \geq \nu_{\mathrm{AL}}^{(2)}(\delta,\lambda)$. Furthermore, when $k_L = 1$, $\boldsymbol{L} = \boldsymbol{O}_{\hat{t},\hat{t}}$ since $\boldsymbol{\Theta}_{pq} = \boldsymbol{I}_{k_L} - \boldsymbol{\theta}_{pq}\boldsymbol{\theta}_{pq}' = 1 - 1 = 0$. Hence, we have $\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) = 0$, and $\nu_{\mathrm{AL}}^{(1)}(\delta,\lambda) = \nu_{\mathrm{AL}}^{(2)}(\delta,\lambda)$.

(2). According to Lemma 9, when $\delta = 0$, $\boldsymbol{R}_{\mathrm{AL}}(0,\lambda) = \boldsymbol{O}_{\hat{t}k_L,\hat{t}k_L}$. Hence, we have

$$\begin{aligned}\nu_{\mathrm{AL}}^{(3)}(0,\lambda) &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) + \lambda \times 0 \times \mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{R}_{\mathrm{AL}}) \\ &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) + 0 \times 0 \\ &= \hat{t}k_L + k_G - \lambda\mathrm{tr}(\boldsymbol{M}_{11}\boldsymbol{L}) \\ &= \nu_{\mathrm{AL}}^{(2)}(0,\lambda).\end{aligned}$$

Acknowledgment

The authors thank FORTE Science Communications (`https://www.forte-science.co.jp/`) for English language editing. This work was partially supported by JSPS KAKENHI Grant Numbers 23H00809, 25K17296, and 25K21159.

References

[1] Bernardi, M., Canale, A. and Stefanucci, M. (2025). On the Degrees of Freedom of some Lasso procedures. arXiv:2511.21595.

[2] Bleakley, K. and Vert, J.P. (2011). The group fused lasso for multiple change-point detection. arXiv:1106.4199.

[3] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, **3** (1), 1–122.

[4] Efron, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association*, **99** (467), 619–632.

[5] Hallac, D., Leskovec, J. and Boyd, S. (2015). Network lasso: clustering and optimization in large graphs. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 387–396.

[6] Horn, R.A. and Johnson, C.R. (1990). *Matrix Analysis*. Cambridge University Press.

[7] Mallows, C.L. (1973). Some comment on C_p. *Technometrics*, **15**, 661–675.

[8] Ohishi, M., Okamura, K., Itoh, Y., Wakaki, H. and Yanagihara, H. (2025). Coordinate descent algorithm for generalized group fused lasso. *Behaviormetrika*, **52** (1), 105–137.

[9] Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9** (6), 1135-–1151.

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, **58** (1), 267–288.

[11] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **67** (1), 91–108.

[12] Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, **93** (441), 120–131.

[13] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **68** (1), 49–67.

[14] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101** (476), 1418–1429.

[15] Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *The Annals of Statistics*, **35** (5), 2173–2192.