PENALIZED REGRESSION FOR MULTIPLE TYPES OF MANY FEATURES WITH MISSING DATA

Kin Yau Wong[1], Donglin Zeng[2], and D. Y. Lin[2]

[1] The Hong Kong Polytechnic University and [2] The University of North Carolina at Chapel Hill

Supplementary Material

S1 Additional simulation studies

We investigated the sensitivity of the proposed methods under misspecified latent-factor models. We considered two types of features, $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$, both with dimension $p/2$. We generated the features as follows:

$$\boldsymbol{S}^{(k)} = \sum_{j=1}^{3} \boldsymbol{\psi}_j^{(k)} U_j + \boldsymbol{\epsilon}^{(k)} \quad \text{for } k = 1, 2,$$

where (U_1, U_2, U_3) are independent standard normal variables, $\boldsymbol{\epsilon}^{(1)}$ and $\boldsymbol{\epsilon}^{(2)}$ are independent $(p/2)$-variate standard normal variables, and

$$\boldsymbol{\psi}_1^{(1)} = \boldsymbol{\psi}_1^{(2)} = (\underbrace{0.2, \ldots, 0.2}_{20 \text{ terms}}, \underbrace{-0.2, \ldots, -0.2}_{(p/4-10) \text{ terms}}, \underbrace{0.2, \ldots, 0.2}_{(p/4-10) \text{ terms}})^{\mathrm{T}},$$

$$\boldsymbol{\psi}_2^{(1)} = \boldsymbol{\psi}_3^{(2)} = (\underbrace{0.3, \ldots, 0.3}_{20 \text{ terms}}, \underbrace{0.3, \ldots, 0.3}_{(p/4-10) \text{ terms}}, \underbrace{-0.15, \ldots, -0.15}_{(p/4-10) \text{ terms}})^{\mathrm{T}},$$

$$\boldsymbol{\psi}_2^{(2)} = \boldsymbol{\psi}_3^{(1)} = (\underbrace{0.3, \ldots, 0.3}_{20 \text{ terms}}, \underbrace{0.3, \ldots, 0.3}_{(p/4-10) \text{ terms}}, \underbrace{0.15, \ldots, 0.15}_{(p/4-10) \text{ terms}})^{\mathrm{T}}.$$

In this setting, every component of $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ is dependent on all three latent variables, and there are no type-specific latent variables. The continuous outcome variable was generated according to the description in Section 5, whereas the binary outcome variable was generated with $P(Y = 1 \mid \boldsymbol{S}^{(1)}, \boldsymbol{S}^{(2)}) = \text{logit}^{-1}\{-3.5 + \sum_{j=1}^{15} 0.15(S_j^{(1)} + S_j^{(2)})\}$, such that $P(Y = 1) \approx 0.1$. We considered the same missing-data mechanisms and estimation methods as in Section 5, but for the imputation method based on the factor model and the proposed method, we only consider $r_0 = r_1 = r_2 = 1$. We set $n = 500$ and $p = 100$ or 300. The results, which are based on 200 replicates, are summarized in Tables S1 and S2.

The pattern of results is very similar to that presented in Section 5. The proposed method clearly outperforms the complete-case analysis in terms of variable selection and prediction. Under MCAR, single imputation methods and the proposed method perform similarly, whereas under MAR, the proposed method yields smaller prediction error and similar or better false discovery and true positive rates than the single imputation methods. The results suggest that even under a misspecified latent-variable structure, the proposed method can still yield satisfactory performance by utilizing information of the missing variables in the observed data and by accounting for the missing mechanism.

S2 Additional lemmas

Lemma S1. *Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ be i.i.d. random vectors and $\boldsymbol{f}$ be a vector-valued function. If $\mathrm{E}\boldsymbol{f}(\boldsymbol{Z}_1) = \mathbf{0}$, then $\|n^{-1/2} \sum_{i=1}^{n} \boldsymbol{f}(\boldsymbol{Z}_i)\| = O_p[\{\mathrm{E}\|\boldsymbol{f}(\boldsymbol{Z}_1)\|^2\}^{1/2}]$.*

In the sequel, the distribution of $(Y, \boldsymbol{S})$ is always evaluated conditional on $\boldsymbol{X}$, and we suppress the argument $\boldsymbol{X}$ in the conditional probabilities, expectations, and density functions. Let $\boldsymbol{S}^{(O)}$ be an arbitrary subvector of $\boldsymbol{S}$ and M be a binary random variable that is conditionally independent of $\boldsymbol{S}$ given $(Y, \boldsymbol{X}, \boldsymbol{S}^{(O)})$ and satisfies $P(M=1 \mid Y, \boldsymbol{S}^{(O)}) > C^{-1}$ almost surely for some $C>0$. Let $\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}) \equiv \partial \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}} / \partial \boldsymbol{\theta}_{\mathcal{S}}^{\mathrm{T}}$ be the Hessian matrix of the log-likelihood function for a subject with complete data and

$$
\begin{aligned}
\boldsymbol{K}_1(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}) &= M\Big[\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}) - \mathrm{E}\Big\{\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}) \mid Y, \boldsymbol{S}^{(O)}; \boldsymbol{\theta}\Big\}\Big],\\
\boldsymbol{K}_2(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}) &= \mathrm{E}\Big\{\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}) + \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta})\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta})^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \boldsymbol{\theta}\Big\},\\
\boldsymbol{K}_3(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}) &= -M\mathrm{E}\Big\{\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta})\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta})^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \boldsymbol{\theta}\Big\},\\
\boldsymbol{K}(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}) &= \boldsymbol{K}_1(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}) + \boldsymbol{K}_2(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}) + \boldsymbol{K}_3(\boldsymbol{\theta}, \boldsymbol{S}^{(O)}).
\end{aligned}
$$

Let $\mathcal{N}_K = \{(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) : \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2 + \|\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{0\mathcal{S}}\|^2 + \|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|^2 + p_n^{-1}\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\|^2 + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\|^2 + p_n^{-1}\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|^2 + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|^2 + \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\|^2 \le K^2 n^{-1} p_n, \boldsymbol{\beta}_{\mathcal{N}} = \boldsymbol{0}\}$ for some positive constant K. Let $\boldsymbol{\theta}_{0\mathcal{S}}$ be the true value of $\boldsymbol{\theta}_{\mathcal{S}}$.

Lemma S2. *Under conditions (C1), (C5), and (C6) and for any $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in \mathcal{N}_K$,*

$$
(\mathbb{P}_n - \mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}} \boldsymbol{K}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{S}^{(O)})\boldsymbol{d}_\theta
$$

is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0; \boldsymbol{S}^{(O)})\boldsymbol{d}_\theta$, where $\boldsymbol{d}_\theta = \boldsymbol{\theta}_{\mathcal{S}} - \boldsymbol{\theta}_{0\mathcal{S}}$.

Lemma S3. *Under conditions (C1), (C5), and (C6) and for any $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in \mathcal{N}_K$,*

$$
\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\{\boldsymbol{K}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{S}^{(O)}) - \boldsymbol{K}(\boldsymbol{\theta}_0, \boldsymbol{S}^{(O)})\}\boldsymbol{d}_\theta
$$

is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0; \boldsymbol{S}^{(O)})\boldsymbol{d}_\theta$, where $\boldsymbol{d}_\theta = \boldsymbol{\theta}_{\mathcal{S}} - \boldsymbol{\theta}_{0\mathcal{S}}$.

S3 Proofs of lemmas

We now prove Lemmas 1 and S1–3. For any real number sequences a_n and b_n, $a_n \asymp b_n$ means that $\limsup |a_n/b_n| < \infty$ and $\limsup |b_n/a_n| < \infty$.

Proof of Lemma 1. For simplicity of presentation, we assume that $\boldsymbol{S}$ is either completely observed or is observed only for a fixed subset of $\boldsymbol{S}$, denoted by $\boldsymbol{S}^{(O)}$. Under this missing-data pattern, the missing-data indicator M can be defined as a univariate variable, with $M = 1$ if $\boldsymbol{S}$ is fully observed and $M = 0$ otherwise. We consider only the case that $p_n \to \infty$; the case for $p_n < \infty$ is relatively straightforward.

The main step of the proof is to show that for any fixed $\delta > 0$ and large enough K and n, $P\big\{\sup_{\boldsymbol{\theta}\in\partial\mathcal{N}_K} p\ell_n(\boldsymbol{\theta}) < p\ell_n(\boldsymbol{\theta}_0)\big\} \geq 1-\delta$, such that there exists a local maximum of $p\ell_n(\cdot)$ in $\mathcal{N}_K$. Let $\boldsymbol{\theta}$ be some value in $\partial\mathcal{N}_K$. By the Taylor series expansion,

$$\begin{aligned} & n^{-1}\{p\ell_n(\boldsymbol{\theta}) - p\ell_n(\boldsymbol{\theta}_0)\} \\ =& (\boldsymbol{\theta}_{\mathcal{S}} - \boldsymbol{\theta}_{0\mathcal{S}})^{\mathrm{T}}\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\theta}_{\mathcal{S}}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \frac{1}{2}(\boldsymbol{\theta}_{\mathcal{S}} - \boldsymbol{\theta}_{0\mathcal{S}})^{\mathrm{T}}\mathbb{P}_n \frac{\partial^2}{\partial \boldsymbol{\theta}_{\mathcal{S}}\partial \boldsymbol{\theta}_{\mathcal{S}}^{\mathrm{T}}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}(\boldsymbol{\theta}_{\mathcal{S}} - \boldsymbol{\theta}_{0\mathcal{S}}) \\ & + \lambda_n \sum_{j=1}^{p_{1n}} w_j(|\beta_j| - |\beta_{0j}|), \end{aligned} \tag{S1}$$

where $\widetilde{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. The first term on the right-hand side of (S1) is

$$\begin{aligned} & (\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{0\mathcal{S}})^{\mathrm{T}}\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\beta}_{\mathcal{S}}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} + \big((\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^{\mathrm{T}}, (\boldsymbol{\xi}-\boldsymbol{\xi}_0)^{\mathrm{T}}\big)\mathbb{P}_n \frac{\partial}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\xi}^{\mathrm{T}})^{\mathrm{T}}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ & + \mathrm{tr}\Big\{(\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0)\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\Sigma}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Big\} + \mathrm{tr}\Big\{(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)^{\mathrm{T}}\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\Psi}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Big\} \\ & + \mathrm{tr}\Big\{(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)^{\mathrm{T}}\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\Gamma}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Big\}. \end{aligned} \tag{S2}$$

For the first term of (S2), note that

$$\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\beta}_{\mathcal{S}}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbb{P}_n\Big\{M\dot{\boldsymbol{\ell}}_{\beta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0) + (1-M)\int \dot{\boldsymbol{\ell}}_{\beta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0) f(Y, \boldsymbol{S})\,\mathrm{d}\boldsymbol{S}^{(M)}\Big\},$$

where $f(Y, \boldsymbol{S})$ is the true density function of $(Y, \boldsymbol{S})$ (conditional on $\boldsymbol{X}$). By the properties of the score statistic and the missing-at-random assumption, the term in the curly brackets of the above equation is mean zero. By Jensen's inequality and condition (C1),

$$\mathrm{E}\left\|M\dot{\boldsymbol{\ell}}^{(C)}_{\beta_{\mathcal{S}}}(\boldsymbol{\theta}_0) + (1-M)\int \dot{\boldsymbol{\ell}}^{(C)}_{\beta_{\mathcal{S}}}(\boldsymbol{\theta}_0) f(Y, \boldsymbol{S})\,\mathrm{d}\boldsymbol{S}^{(M)}\right\|^2 \leq \mathrm{E}\|\dot{\boldsymbol{\ell}}^{(C)}_{\beta_{\mathcal{S}}}(\boldsymbol{\theta}_0)\|^2 = O(p_n).$$

By Lemma S1, we conclude that $\|\mathbb{P}_n \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\beta}_{\mathcal{S}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\| = O_p(n^{-1/2}p_n^{1/2})$. Likewise, by condition (C1), Lemma S1, Jensen's inequality, and the fact that $\partial\ell(\boldsymbol{\theta})/\partial(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\xi}^{\mathrm{T}})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is mean zero, we conclude that $\|\mathbb{P}_n \partial\ell(\boldsymbol{\theta})/\partial(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\xi}^{\mathrm{T}})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\| = O_p(n^{-1/2}p_n^{1/2})$.

For the third term of (S2), we can likewise show that $\mathrm{E}\{\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\Sigma}\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\} = \mathbf{0}$, such that Lemma S1 and Jensen's inequality imply that $\|\mathbb{P}_n\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\Sigma}\| = O_p[n^{-1/2}\{\mathrm{E}\|\dot{\boldsymbol{\ell}}^{(C)}_{\Sigma}(\boldsymbol{\theta})\|^2\}^{1/2}]$. Using the fact that $\boldsymbol{\Omega}_0^{-1} \equiv (\boldsymbol{\Psi}_0\boldsymbol{\Psi}_0^{\mathrm{T}} + \boldsymbol{\Sigma}_0)^{-1} = \boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}$, we can write $\dot{\boldsymbol{\ell}}^{(C)}_{\Sigma}(\boldsymbol{\theta}_0)$ as

$$\begin{aligned}
&\mathrm{diag}[\boldsymbol{\Omega}_0^{-1}\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\boldsymbol{\Omega}_0^{-1}] \\
=&\mathrm{diag}[\{\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\} \\
&\times\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\{\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\}] \\
=&\mathrm{diag}[\boldsymbol{\Sigma}_0^{-1}\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\boldsymbol{\Sigma}_0^{-1}] \\
&- 2\mathrm{diag}[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\boldsymbol{\Sigma}_0^{-1}] \\
&+ \mathrm{diag}[\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0 \\
&\times(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}].
\end{aligned}$$

The expected squared L_2-norm of the first term on the right-hand side of the second equality above is bounded by $\sum_{j=1}^{p_n}\mathrm{E}(S_j^4/\sigma_{0j}^8) = O(p_n)$, where σ_{0j}^2 is the jth diagonal element of $\boldsymbol{\Sigma}_0$. Similar arguments show that $\mathrm{E}\|\dot{\boldsymbol{\ell}}^{(C)}_{\Sigma}(\boldsymbol{\theta}_0)\|^2 = O(p_n)$, so $\|\mathbb{P}_n\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\Sigma}\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\| =$

$O_p(n^{-1/2}p_n^{1/2})$.

To evaluate the fourth term of (S2), we decompose $\boldsymbol{\Psi} - \boldsymbol{\Psi}_0 = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi} - \boldsymbol{\Psi}_0) + \boldsymbol{\Sigma}_0^{1/2}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi} - \boldsymbol{\Psi}_0)$ and show that the expected value of the fourth term of (S2) can be written as the sum of two terms that converge to zero at different rates as $n \to \infty$. Let $\boldsymbol{D}_\Psi = \boldsymbol{\Psi} - \boldsymbol{\Psi}_0$ and $\boldsymbol{D}_V = \boldsymbol{\Sigma}_0^{-1/2}\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\boldsymbol{\Sigma}_0^{-1/2}$. We find an orthonormal matrix $\boldsymbol{L}$ and a diagonal matrix $\boldsymbol{A}$ such that $\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0 = \boldsymbol{L}\boldsymbol{A}\boldsymbol{L}^{\mathrm{T}}$. Under condition (C5), $a_j \asymp p_n$, where a_j is the jth diagonal element of $\boldsymbol{A}$. Because $\boldsymbol{H}$ is a projection matrix onto the linear space of $\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0$, we can find a matrix $\boldsymbol{R} \in \mathbb{R}^{r\times r}$ with $\|\boldsymbol{R}\| = 1$ such that $\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi = d_\Psi\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{R}$ for some $d_\Psi = O(p_n^{-1/2}\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\|) = O(n^{-1/2}p_n^{1/2})$. It can be shown that

$$\begin{aligned}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi &= \boldsymbol{\Sigma}_0^{-1/2}\{\boldsymbol{I} - \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\}\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi \\ &= d_\Psi\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}, \qquad \text{(S3)}\end{aligned}$$

and

$$\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_0^{1/2}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi = \boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi. \qquad \text{(S4)}$$

Thus,

$$\begin{aligned}\mathrm{tr}\Big\{\dot{\boldsymbol{\ell}}_\Psi^{(C)}(\boldsymbol{\theta}_0)^{\mathrm{T}}\boldsymbol{D}_\Psi\Big\} &= \mathrm{tr}\Big[\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\{(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})(\boldsymbol{S} - \boldsymbol{\Gamma}_0\boldsymbol{X})^{\mathrm{T}} - \boldsymbol{\Omega}_0\}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\Big] \\ &= \mathrm{tr}\Big[\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\{\boldsymbol{I} - \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\}\boldsymbol{D}_V\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\Big] \\ &= \mathrm{tr}\Big[\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_V\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Sigma}_0^{1/2} \\ &\qquad \times \{\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi + (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\}\Big] \\ &= \mathrm{tr}\big\{d_\Psi\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_V\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R} \\ &\qquad + \boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_V(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\big\}.\end{aligned}$$

Note that $\mathrm{E}[\mathrm{tr}\{d_\Psi \boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_V\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}\}^2]$ is bounded above by $d_\Psi^2\|\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\|^4\|\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\|^4\mathrm{E}\|\boldsymbol{D}_V\|^2 = O(p_n^{-1}\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\|^2)$, and $\mathrm{E}[\mathrm{tr}\{\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_V(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\}^2]$ is bounded above by

$$
\begin{aligned}
&\|\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\|^2\|\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\|^2\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\|^2\mathrm{E}\|\boldsymbol{D}_V\|^2\\
&= O\{p_n\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\|^2\}.
\end{aligned}
$$

We conclude that

$$
\Big(\mathrm{E}[\mathrm{tr}\{\dot{\boldsymbol{\ell}}_\Psi^{(C)}(\boldsymbol{\theta}_0)^{\mathrm{T}}\boldsymbol{D}_\Psi\}^2]\Big)^{1/2} = O(p_n^{-1/2})\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\| + O(p_n^{1/2})\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\|.
$$

Similarly, we can decompose $\boldsymbol{D}_\Gamma \equiv \boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)+\boldsymbol{\Sigma}_0^{1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)$ and show that

$$
\Big(\mathrm{E}[\mathrm{tr}\{\dot{\boldsymbol{\ell}}_\Gamma^{(C)}(\boldsymbol{\theta}_0)^{\mathrm{T}}\boldsymbol{D}_\Gamma\}^2]\Big)^{1/2} = O(p_n^{-1/2})\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Gamma\| + O(p_n^{1/2})\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Gamma\|.
$$

Applying Lemma S1 and Jensen's inequality and combining the above results, we conclude that (S2) is bounded above by

$$
\begin{aligned}
&O_p(n^{-1/2}p_n^{1/2})\big\{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| + \|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\| + \|\boldsymbol{\xi}-\boldsymbol{\xi}_0\| + \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\|\\
&+\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\| + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|\big\}\\
&+O_p(n^{-1/2}p_n^{-1/2})\big\{\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\| + \|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|\big\}.
\end{aligned}
$$

Consider the second term of (S1). With $\boldsymbol{d}_\theta = \boldsymbol{\theta}_{\mathcal{S}} - \boldsymbol{\theta}_{0\mathcal{S}}$, we have

$$
\begin{aligned}
&\boldsymbol{d}_\theta^{\mathrm{T}}\mathbb{P}_n\frac{\partial^2}{\partial\boldsymbol{\theta}_{\mathcal{S}}\partial\boldsymbol{\theta}_{\mathcal{S}}^{\mathrm{T}}}\ell(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\boldsymbol{d}_\theta\\
&=\mathbb{P}_n\Big[\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta - (1-M)\mathrm{E}\big\{\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{d}_\theta \mid Y, \boldsymbol{S}^{(O)}; \widetilde{\boldsymbol{\theta}}\big\}^2\Big]\\
&\le(\mathbb{P}_n-\mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta + \mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\{\boldsymbol{K}(\widetilde{\boldsymbol{\theta}})-\boldsymbol{K}(\boldsymbol{\theta}_0)\}\boldsymbol{d}_\theta + \mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta\\
&\le\{1-o_p(1)\}\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta,
\end{aligned}
$$

where we have suppressed the argument $\boldsymbol{S}^{(O)}$ in $\boldsymbol{K}(\cdot,\cdot)$ and $\boldsymbol{K}_3(\cdot,\cdot)$, and the last inequality follows from Lemma S2, Lemma S3, and the fact that $\mathbb{P}\boldsymbol{K}_1(\boldsymbol{\theta}_0) = \mathbb{P}\boldsymbol{K}_2(\boldsymbol{\theta}_0) = \mathbf{0}$. Clearly,

$$\begin{aligned}\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta = & -\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\Big[M\mathrm{E}\Big\{\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0)\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0)^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \boldsymbol{\theta}_0\Big\}\Big]\boldsymbol{d}_\theta \\ \leq & -C^{-1}\mathbb{P}\Big\{\boldsymbol{d}_\theta^{\mathrm{T}}\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0)\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0)^{\mathrm{T}}\boldsymbol{d}_\theta\Big\} = C^{-1}\boldsymbol{d}_\theta^{\mathrm{T}}\mathbb{P}\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta,\end{aligned}$$

where the inequality follows from condition (C2). Note that for $(d_1 \times d_2)$-matrices $\boldsymbol{C}_1$ and $\boldsymbol{D}_1$, $(d_3 \times d_4)$-matrices $\boldsymbol{C}_2$ and $\boldsymbol{D}_2$, and any twice-differentiable function $f : \mathbb{R}^{d_1\times d_2} \times \mathbb{R}^{d_3\times d_4} \to \mathbb{R}$,

$$\mathrm{vec}(\boldsymbol{D}_1)^{\mathrm{T}}\frac{\partial^2}{\partial\mathrm{vec}(\boldsymbol{C}_1)\partial\mathrm{vec}(\boldsymbol{C}_2)^{\mathrm{T}}}f(\boldsymbol{C}_1,\boldsymbol{C}_2)\mathrm{vec}(\boldsymbol{D}_2) = \mathrm{tr}\Big[\boldsymbol{D}_1^{\mathrm{T}}\frac{\partial}{\partial\boldsymbol{C}_1}\mathrm{tr}\Big\{\frac{\partial}{\partial\boldsymbol{C}_2}f(\boldsymbol{C}_1,\boldsymbol{C}_2)^{\mathrm{T}}\boldsymbol{D}_2\Big\}\Big].$$

To simplify the presentation of the second derivatives of functions with respect to $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Sigma}$, we express the right-hand side of the above equation as

$$\mathrm{tr}\Big\{\boldsymbol{D}_1^{\mathrm{T}}\frac{\partial^2}{\partial\boldsymbol{C}_1\partial\boldsymbol{C}_2^{\mathrm{T}}}f(\boldsymbol{C}_1,\boldsymbol{C}_2)\boldsymbol{D}_2\Big\}.$$

By simple matrix calculus,

$$\begin{aligned}&\mathrm{tr}\Big\{\boldsymbol{D}_\Psi^{\mathrm{T}}\frac{\partial^2}{\partial\boldsymbol{\Psi}\partial\boldsymbol{\Psi}^{\mathrm{T}}}\log f(\boldsymbol{S};\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{D}_\Psi\Big\} \\ &\quad = \mathrm{tr}\Big\{-\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi + (\boldsymbol{I} - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi})\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{B}\boldsymbol{D}_\Psi \\ &\qquad - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{D}_\Psi\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Psi}\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{D}_\Psi\Big\}, \qquad \text{(S5)}\end{aligned}$$

$$\begin{aligned}&\mathrm{tr}\Big\{\boldsymbol{D}_\Sigma\frac{\partial^2}{\partial\boldsymbol{\Sigma}\partial\boldsymbol{\Psi}^{\mathrm{T}}}\log f(\boldsymbol{S};\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{D}_\Psi\Big\} \\ &\quad = \mathrm{tr}\Big(-\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Psi - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Sigma\boldsymbol{B}\boldsymbol{D}_\Psi\Big), \qquad \text{(S6)}\end{aligned}$$

$$\begin{aligned}&\mathrm{tr}\Big\{\boldsymbol{D}_\Sigma\frac{\partial^2}{\partial\boldsymbol{\Sigma}\partial\boldsymbol{\Sigma}}\log f(\boldsymbol{S};\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{D}_\Sigma\Big\} \\ &\quad = \frac{1}{2}\mathrm{tr}\Big(-\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Sigma - \boldsymbol{B}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Sigma - \boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Sigma\boldsymbol{B}\boldsymbol{D}_\Sigma\Big), \qquad \text{(S7)}\end{aligned}$$

$$\mathrm{tr}\Big\{\boldsymbol{D}_\Gamma\frac{\partial^2}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}^{\mathrm{T}}}\log f(\boldsymbol{S};\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{D}_\Gamma\Big\} = -\mathrm{tr}\Big(\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Gamma\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\Big), \qquad \text{(S8)}$$

$$\mathrm{tr}\left\{\boldsymbol{D}_\Gamma \frac{\partial^2}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Psi}^{\mathrm{T}}}\log f(\boldsymbol{S};\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{D}_\Psi\right\} = -2\mathrm{tr}\left\{\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Gamma\boldsymbol{X}(\boldsymbol{S}-\boldsymbol{\Gamma}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\right\}, \quad \text{(S9)}$$

$$\mathrm{tr}\left\{\boldsymbol{D}_\Gamma \frac{\partial^2}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Sigma}}\log f(\boldsymbol{S};\boldsymbol{\Gamma},\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{D}_\Sigma\right\} = -\mathrm{tr}\left\{\boldsymbol{D}_\Sigma\boldsymbol{\Omega}^{-1}\boldsymbol{D}_\Gamma\boldsymbol{X}(\boldsymbol{S}-\boldsymbol{\Gamma}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\right\}, \quad \text{(S10)}$$

where $\boldsymbol{D}_\Sigma = \boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0$, and $\boldsymbol{B} = \boldsymbol{\Omega}^{-1}\{(\boldsymbol{S}-\boldsymbol{\Gamma}\boldsymbol{X})(\boldsymbol{S}-\boldsymbol{\Gamma}\boldsymbol{X})^{\mathrm{T}}-\boldsymbol{\Omega}\}\boldsymbol{\Omega}^{-1}$. With $\boldsymbol{d}_{(\alpha\beta\xi)}$ denoting the $(\boldsymbol{\alpha},\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\xi})$-component of $\boldsymbol{d}_\theta$, we can write

$$\begin{aligned}
&\boldsymbol{d}_\theta^{\mathrm{T}}\mathbb{P}\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta \\
&= \boldsymbol{d}_{(\alpha\beta\xi)}^{\mathrm{T}}\mathbb{P}\boldsymbol{V}_{\mathcal{S}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0)\boldsymbol{d}_{(\alpha\beta\xi)} - \mathrm{tr}\Big(\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi \\
&\quad + 2\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi + \frac{1}{2}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma + \boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\mathbb{P}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\Big), \qquad \text{(S11)}
\end{aligned}$$

where $\boldsymbol{V}_{\mathcal{S}}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi})$ is the submatrix of $\boldsymbol{V}(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi})$ that corresponds to $(\boldsymbol{\alpha},\boldsymbol{\beta}_{\mathcal{S}},\boldsymbol{\xi})$. By condition (C1),

$$\boldsymbol{d}_{(\alpha\beta\xi)}^{\mathrm{T}}\mathbb{P}\boldsymbol{V}_{\mathcal{S}}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\xi}_0)\boldsymbol{d}_{(\alpha\beta\xi)} < -c_1\Big(\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2 + \|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\|^2 + \|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|^2\Big)$$

for some positive constant c_1.

To bound the first two terms in the trace of the right-hand side of (S11), we note that

$$\begin{aligned}
\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0 &= \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\{\boldsymbol{I}-\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}\}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0 \\
&= \boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}. \qquad \text{(S12)}
\end{aligned}$$

By (S3), (S4), and (S12), we have

$$\begin{aligned}
&\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi \\
&= d_\Psi\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi \\
&= d_\Psi^2\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{R}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi \\
&= d_\Psi^2\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi.
\end{aligned}$$

Likewise, $\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi = d_\Psi^2\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}$. Thus, with $\widetilde{\boldsymbol{R}} \equiv \boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{L}$,

$$\begin{aligned}&\mathrm{tr}(\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi)\\ &= d_\Psi^2\mathrm{tr}\{\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\widetilde{\boldsymbol{R}}^{\mathrm{T}}+\widetilde{\boldsymbol{R}})\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\widetilde{\boldsymbol{R}}\} + \mathrm{tr}\{\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\}.\end{aligned}$$

The first term on the right-hand side of the above equation is nonnegative, because

$$\begin{aligned}&\mathrm{tr}\{\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\widetilde{\boldsymbol{R}}^{\mathrm{T}}+\widetilde{\boldsymbol{R}})\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\widetilde{\boldsymbol{R}}\}\\ &= \sum_{j=1}^r\sum_{k=1}^r \frac{a_j a_k}{(1+a_j)(1+a_k)}(\widetilde{r}_{jk}+\widetilde{r}_{kj})\widetilde{r}_{kj}\\ &= 2\sum_{j=1}^r \frac{a_j^2}{(1+a_j)^2}\widetilde{r}_{jj}^2 + \sum_{j<k}\frac{a_j a_k}{(1+a_j)(1+a_k)}(\widetilde{r}_{jk}+\widetilde{r}_{kj})^2, \qquad \text{(S13)}\end{aligned}$$

where $\widetilde{r}_{jk}$ is the (j,k)th element of $\widetilde{\boldsymbol{R}}$. For the third term in the trace on the right-hand side of (S11), (S3) and (S4) yield

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\Psi}_0{}^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi) &= \mathrm{tr}\{\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\}\\ &= d_\Psi\mathrm{tr}\{\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{R}\}\\ &\quad + \mathrm{tr}\{\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\}\\ &\le O(p_n^{-1})\|\boldsymbol{d}_\Sigma\|\big\{O(d_\Psi)+O(p_n^{1/2})\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi\|\big\}\\ &= o(n^{-1}p_n).\end{aligned}$$

For the fourth term in the trace on the right-hand side of (S11), we have

$$\begin{aligned}&\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)\\ &= \mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma) - 2\mathrm{tr}\{(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\}\\ &\quad + \mathrm{tr}\{(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\}.\end{aligned}$$

The third term of the expression on the right-hand side of the above equation is clearly nonnegative. To show that the second term of the expression is dominated by the first term, let $\boldsymbol{d}_\Sigma = \mathrm{vecd}(\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1/2})$. Clearly, $\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma) = \boldsymbol{d}_\Sigma^{\mathrm{T}}\boldsymbol{d}_\Sigma$, and

$$\begin{aligned}&\mathrm{tr}\{(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\}\\ &=\sum_{j=1}^{r}[\{\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\}_j\circ\boldsymbol{d}_\Sigma]^{\mathrm{T}}[\{\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\}_j\circ\boldsymbol{d}_\Sigma],\end{aligned}$$

where $\{\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}\}_j$ is the jth column of $\boldsymbol{\Psi}_0\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{A})^{-1/2}$ $(j=1,\dots,r)$. Because $a_j\to\infty$ $(j=1,\dots,r)$ and each element of $\boldsymbol{\Psi}_0$ is bounded, the right-hand side of the above equation is dominated by $\boldsymbol{d}_\Sigma^{\mathrm{T}}\boldsymbol{d}_\Sigma$, so $\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma) > \{1-o(1)\}\boldsymbol{d}_\Sigma^{\mathrm{T}}\boldsymbol{d}_\Sigma$.

To bound the fifth term in the trace on the right-hand side of (S11), let $\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Gamma = d_\Gamma\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{Q}$ for some $d_\Gamma = O(n^{-1/2}p_n^{1/2})$ and finite matrix $\boldsymbol{Q}$ of appropriate dimensions. By condition (C1),

$$\begin{aligned}\mathrm{tr}\Big(\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\mathbb{P}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\Big) &> c_2\mathrm{tr}\Big(\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\Big)\\ &= c_2\mathrm{tr}\Big\{d_\Gamma^2\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{Q}+\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Gamma\Big\}\end{aligned}$$

for some positive constant c_2. Combining the above results, we conclude that the right-hand side of (S11) is bounded above by

$$\begin{aligned}&-c_3\big[\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2+\|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\|^2+\|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|^2+\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\|^2\\ &\quad+\|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|^2+\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\|^2+d_\Gamma^2\mathrm{tr}\{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{Q}\}\big]\\ &\quad-d_\Psi^2\mathrm{tr}\{\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\widetilde{\boldsymbol{R}}^{\mathrm{T}}+\widetilde{\boldsymbol{R}})\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\widetilde{\boldsymbol{R}}\}+o(n^{-1}p_n)\end{aligned}$$

for some positive constant c_3 and large enough n.

The third term on the right-hand side of (S1) is bounded above by

$$\lambda_n\sum_{j=1}^{p_{1n}}w_j|\beta_j-\beta_{0j}|\le\lambda_n\|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\|\|\boldsymbol{w}_{\mathcal{S}}\| = O_p(\lambda_n n^\rho p_n^{1/2}\|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\|) = o_p(n^{-1}p_n)$$

under conditions (C3) and (C4). Thus, for some positive constant c_4,

$$
\begin{aligned}
& n^{-1}\{p\ell_n(\boldsymbol{\theta}) - p\ell_n(\boldsymbol{\theta}_0)\} \\
< & \, O_p(n^{-1/2}p_n^{1/2})\big\{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| + \|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\| + \|\boldsymbol{\xi}-\boldsymbol{\xi}_0\| + \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\| \\
& + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\| + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|\big\} \\
& - c_4\big\{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2 + \|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\|^2 + \|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|^2 + \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\|^2 \\
& + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\|^2 + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|^2\big\} \\
& + O_p(n^{-1/2}p_n^{-1/2})\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\| - c_4 d_\Gamma^2 \mathrm{tr}\big\{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{Q}\big\} \\
& + O_p(n^{-1/2}p_n^{-1/2})\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\| - c_4 d_\Psi^2 \mathrm{tr}\{\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\widetilde{\boldsymbol{R}}^{\mathrm{T}} + \widetilde{\boldsymbol{R}})\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\widetilde{\boldsymbol{R}}\} \\
& + o_p(n^{-1}p_n).
\end{aligned}
$$

Let τ_1 and τ_2 be such that $\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2 + \|\boldsymbol{\beta}_{\mathcal{S}}-\boldsymbol{\beta}_{0\mathcal{S}}\|^2 + \|\boldsymbol{\xi}-\boldsymbol{\xi}_0\|^2 + \|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0\|^2 + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\|^2 + \|(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|^2 = \tau_1 K^2 n^{-1}p_n$ and $\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0)\|^2 = \tau_2 K^2 n^{-1}p_n^2$, so $\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_0)\|^2 = (1-\tau_1-\tau_2)K^2n^{-1}p_n^2$. If τ_1 does not vanish as $n \to \infty$, then the right-hand side of the above inequality is bounded by

$$
O_p(Kn^{-1}p_n) - c_5K^2n^{-1}p_n + O_p(Kn^{-1}p_n^{1/2}) + o_p(n^{-1}p_n)
$$

for some positive c_5. Because the above expression is negative for large enough K and n, the desired result follows. Alternatively, if $\tau_1 \to 0$ and τ_2 is bounded away from zero, then $d_\Psi \asymp n^{-1/2}p_n^{1/2}$. In this case, we show by contradiction that $\mathrm{tr}\{\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\widetilde{\boldsymbol{R}}^{\mathrm{T}}+\widetilde{\boldsymbol{R}})\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\widetilde{\boldsymbol{R}}\}$ is bounded away from zero or equivalently, that at least one element of $(\widetilde{\boldsymbol{R}}+\widetilde{\boldsymbol{R}}^{\mathrm{T}})$ is bounded away from zero. Suppose that $\widetilde{\boldsymbol{R}}+\widetilde{\boldsymbol{R}}^{\mathrm{T}} = o(1)$, so that $\boldsymbol{R}+\boldsymbol{R}^{\mathrm{T}} = o(1)$; here, the $o(1)$ term refers to a matrix that converges to 0 elementwise. Note that

$$
\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi}-\boldsymbol{\Psi}_0) = \boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi + (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi = d_\Psi\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Psi}_0\boldsymbol{R} + (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi,
$$

so

$$\mathbf{\Psi} = \mathbf{\Psi}_0(\boldsymbol{I} + d_\Psi \boldsymbol{R}) + \mathbf{\Sigma}_0^{-1/2}(\boldsymbol{I} - \boldsymbol{H})\mathbf{\Sigma}_0^{-1/2}\boldsymbol{D}_\Psi = \mathbf{\Psi}_0 + d_\Psi\{\mathbf{\Psi}_0\boldsymbol{R} + o(1)\};$$

the second equality holds because each element of the second term on the right-hand side of the first equality is $o(n^{-1/2}p_n^{1/2})$ if $\tau_1 \to 0$. Write $\boldsymbol{R}$ as a $(K+1)\times(K+1)$ block matrix and let $\boldsymbol{R}^{(j,k)} \in \mathbb{R}^{r_j \times r_k}$ be its $(j+1,k+1)$th block element for $j,k = 0,\ldots,K$. Because the upper-right corner of the block matrix representation of $\mathbf{\Psi}$ in (2.1) is zero, $\mathbf{\Psi}^{(0,1)}\boldsymbol{R}^{(0,K)} + \mathbf{\Psi}^{(1)}\boldsymbol{R}^{(1,K)} = o(1)$. By condition (C5), no linear combinations (with nonvanishing coefficients) of the columns of $(\mathbf{\Psi}^{(0,1)}, \mathbf{\Psi}^{(1)})$ go to zero, so $\boldsymbol{R}^{(0,K)} = o(1)$ and $\boldsymbol{R}^{(1,K)} = o(1)$. Similar arguments show that

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}^{(0,0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{R}^{(0,1)} & \boldsymbol{R}^{(1,1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{R}^{(0,2)} & \mathbf{0} & \boldsymbol{R}^{(2,2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \ddots & \vdots \\ \boldsymbol{R}^{(0,K)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{R}^{(K,K)} \end{pmatrix} + o(1).$$

If $\boldsymbol{R} + \boldsymbol{R}^{\mathrm{T}} = o(1)$, then $\boldsymbol{R}^{(0,k)}$ can be chosen as $\mathbf{0}$ $(k = 1,\ldots,K)$. Also, noting that the upper triangular elements of $\mathbf{\Psi}^{(0,1)}$ and $\mathbf{\Psi}^{(k)}$ $(k=1,\ldots,K)$ are 0 and that the true values of $\psi_{jj}^{(0,1)}$ $(j = 1,\ldots,r_0)$ and $\psi_{jj}^{(k)}$ $(j = 1,\ldots,r_k; k = 1,\ldots,K)$ are bounded away from 0, we conclude that the upper triangular elements of $\boldsymbol{R}^{(k,k)}$ $(k = 0,\ldots,K)$ can be chosen as 0. Therefore, $\boldsymbol{R} + \boldsymbol{R}^{\mathrm{T}} = o(1)$ implies that $\boldsymbol{R} = o(1)$, which contradicts the definition of $\boldsymbol{R}$. We conclude that $d_\Psi^2 \mathrm{tr}\{\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}(\widetilde{\boldsymbol{R}}^{\mathrm{T}} + \widetilde{\boldsymbol{R}})\boldsymbol{A}(\boldsymbol{I}+\boldsymbol{A})^{-1}\widetilde{\boldsymbol{R}}\} > c_6 K^2 n^{-1} p_n$ for some $c_6 > 0$ and large enough n. Therefore, if $\tau_1 \to 0$ and τ_2 is bounded away from 0, then

$$n^{-1}\{p\ell_n(\boldsymbol{\theta}) - p\ell_n(\boldsymbol{\theta}_0)\}$$

$$\leq o_p(Kn^{-1}p_n) + O_p(n^{-1/2}p_n^{-1/2})\|\boldsymbol{H}\boldsymbol{\Sigma}_0^{-1/2}(\boldsymbol{\Psi} - \boldsymbol{\Psi}_0)\| - c_6K^2n^{-1}p_n + o_p(n^{-1}p_n)$$

$$\leq o_p(Kn^{-1}p_n) + O_p(Kn^{-1}p_n^{1/2}) - c_6K^2n^{-1}p_n + o_p(n^{-1}p_n),$$

where the right-hand side of the above inequality is negative for large enough K and n. The desired result follows. Finally, we can use similar arguments to show that the desired result holds when $\tau_1 \to 0$ and $\tau_2 \to 0$. □

Proof of Lemma S1. Let p be the dimension of $\boldsymbol{f}(\boldsymbol{Z})$ and $f_j(\boldsymbol{Z})$ be its jth component $(j = 1, \ldots, p)$. We have

$$\begin{aligned}
\mathrm{E}\Big\|n^{-1/2}\sum_{i=1}^n \boldsymbol{f}(\boldsymbol{Z}_i)\Big\|^2 &= \mathrm{E}\sum_{j=1}^p \Big\{n^{-1/2}\sum_{i=1}^n f_j(\boldsymbol{Z}_i)\Big\}^2 \\
&= \mathrm{E}\sum_{j=1}^p \Big\{n^{-1}\sum_{i=1}^n f_j(\boldsymbol{Z}_i)^2 + n^{-1}\sum_{i\neq k} f_j(\boldsymbol{Z}_i)f_j(\boldsymbol{Z}_k)\Big\} \\
&= \mathrm{E}\sum_{j=1}^p f_j(\boldsymbol{Z}_1)^2 = \mathrm{E}\|\boldsymbol{f}(\boldsymbol{Z}_1)\|^2.
\end{aligned}$$

The desired result follows from Markov's inequality. □

Proof of Lemma S2. Since M is bounded, it suffices to prove that

$$\begin{aligned}
&\Big|(\mathbb{P}_n - \mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta\Big| + \Big|(\mathbb{P}_n - \mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\mathrm{E}\Big\{\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}}) \mid Y, \boldsymbol{S}^{(O)}; \widetilde{\boldsymbol{\theta}}\Big\}\boldsymbol{d}_\theta\Big| \\
&\quad + \Big|(\mathbb{P}_n - \mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\mathrm{E}\Big\{\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \widetilde{\boldsymbol{\theta}}\Big\}\boldsymbol{d}_\theta\Big| \qquad \text{(S14)}
\end{aligned}$$

is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$. Consider the first term of (S14). Let $v(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})$ be an arbitrary element of $\boldsymbol{V}(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})$. By the Taylor series expansion,

$$v(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}}) = v(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\xi}_0) + \dot{\boldsymbol{v}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\xi}_0)^{\mathrm{T}}\widetilde{\boldsymbol{d}}_{(\alpha\beta\xi)} + \frac{1}{2}\widetilde{\boldsymbol{d}}_{(\alpha\beta\xi)}^{\mathrm{T}}\ddot{\boldsymbol{v}}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\xi}^*)\widetilde{\boldsymbol{d}}_{(\alpha\beta\xi)},$$

where $\dot{\boldsymbol{v}}$ and $\ddot{\boldsymbol{v}}$ are, respectively, the first and second derivatives of v, $\widetilde{\boldsymbol{d}}_{(\alpha\beta\xi)} = (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}} - \boldsymbol{\alpha}_0^{\mathrm{T}}, \widetilde{\boldsymbol{\beta}}^{\mathrm{T}} - \boldsymbol{\beta}_0^{\mathrm{T}}, \widetilde{\boldsymbol{\xi}}^{\mathrm{T}} - \boldsymbol{\xi}_0^{\mathrm{T}})^{\mathrm{T}}$, and $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \boldsymbol{\xi}^*)$ lies between $(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})$ and $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\xi}_0)$. We apply $(\mathbb{P}_n - \mathbb{P})$ to both sides of the above equation. By condition (C1), the right-hand side of the resulting equation is bounded above by

$$O_p(n^{-1/2}) + O_p(n^{-1/2}p_n^{1/2})\|\widetilde{\boldsymbol{d}}_{(\alpha\beta\xi)}\| + O_p(p_n)\|\widetilde{\boldsymbol{d}}_{(\alpha\beta\xi)}\|^2 = O_p(n^{-1/2}).$$

Therefore,

$$(\mathbb{P}_n - \mathbb{P})\boldsymbol{d}_{(\alpha\beta\xi)}^{\mathrm{T}}\boldsymbol{V}_{\mathcal{S}}(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}})\boldsymbol{d}_{(\alpha\beta\xi)} = O_p(n^{-1/2}p_n)\|\boldsymbol{d}_{(\alpha\beta\xi)}\|^2 = o_p(1)\|\boldsymbol{d}_{(\alpha\beta\xi)}\|^2,$$

which, by condition (C1) and (S11), is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

To derive the bound for the remaining terms of $(\mathbb{P}_n - \mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta$, we note that for any $(\boldsymbol{\Psi}, \boldsymbol{\Omega})$ that satisfy condition (C5), $\|\boldsymbol{I} - \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}\| = \|(\boldsymbol{I} + \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi})^{-1}\| = O(p_n^{-1})$. Furthermore, for any matrix $\boldsymbol{C}$ of appropriate dimensions,

$$\begin{aligned}
\|\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{C}\| &= \|\boldsymbol{\Psi}^{\mathrm{T}}\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}(\boldsymbol{I} + \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\}\boldsymbol{C}\| \\
&\le \|\boldsymbol{L}\|\|\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}^{\mathrm{T}}\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}(\boldsymbol{I} + \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\}\boldsymbol{C}\| \\
&= \|\boldsymbol{L}\|\|(\boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}\| = O(p_n^{-1/2})\|\boldsymbol{C}\|,
\end{aligned}$$

and $\|\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{C}\| = O(\|\boldsymbol{C}\|^2) + O(\|\boldsymbol{C}\|^2)O(\|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\|^2)O\{\|(\boldsymbol{I} + \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi})^{-1}\|\} = O(\|\boldsymbol{C}\|^2)$. Let $\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{\Omega}}^{-1}\{(\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})(\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})^{\mathrm{T}} - \widetilde{\boldsymbol{\Omega}}\}\widetilde{\boldsymbol{\Omega}}^{-1}$. We can show that

$$\begin{aligned}
\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{B}}\boldsymbol{D}_\Psi\| &\le \|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})\|^2 + \|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\| \\
&\le \|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\|(1 + \|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2),
\end{aligned}$$

so $(\mathbb{P}_n - \mathbb{P})\mathrm{tr}\big\{(\boldsymbol{I} - \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Psi}})\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{B}}\boldsymbol{D}_\Psi\big\} = O_p(n^{-1/2})\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\|$. In addition,

$$\|\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{B}}\boldsymbol{D}_\Psi\| \le \{1 + O(p_n^{-1/2})\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2\}\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\|^{1/2},$$

so $(\mathbb{P}_n - \mathbb{P})\mathrm{tr}\big(\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{B}}\boldsymbol{D}_\Psi\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\big) = O_p(n^{-1/2}p_n^{1/2})\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\|$. Likewise,

$$(\mathbb{P}_n - \mathbb{P})\mathrm{tr}\big(\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{\Psi}}\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\big) = O_p(n^{-1/2})\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\|.$$

Let $H_1(\boldsymbol{\theta}), \ldots, H_6(\boldsymbol{\theta})$ denote the terms on the left-hand side of (S5)–(S10), respectively. We conclude that

$$(\mathbb{P}_n - \mathbb{P})H_1(\widetilde{\boldsymbol{\theta}}) = (\mathbb{P}_n - \mathbb{P})O(1 + p_n^{-1/2}\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2)\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\| = o_p(\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\|). \tag{S15}$$

For any matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ of appropriate dimensions,

$$\begin{aligned}
&\boldsymbol{C}_1^{\mathrm{T}}(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}_0^{-1})\boldsymbol{C}_2\\
&= \boldsymbol{C}_1^{\mathrm{T}}\{\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}_0^{-1} - \widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}}(\boldsymbol{I} + \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}})^{-1}\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\}\boldsymbol{C}_2\\
&= \boldsymbol{C}_1^{\mathrm{T}}(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}_0^{-1})\boldsymbol{C}_2 - \boldsymbol{C}_1^{\mathrm{T}}\{\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}}(\boldsymbol{I} + \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}})^{-1}\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\\
&\quad - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\}\boldsymbol{C}_2\\
&= \boldsymbol{C}_1^{\mathrm{T}}(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}_0^{-1})\boldsymbol{C}_2 - \boldsymbol{C}_1^{\mathrm{T}}(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\Psi}_0(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{C}_2\\
&\quad - \boldsymbol{C}_1^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)(\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{C}_2\\
&\quad - \boldsymbol{C}_1^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}}\{(\boldsymbol{I} + \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}})^{-1} - (\boldsymbol{I} + \boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{C}_2\\
&\quad - \boldsymbol{C}_1^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}}(\boldsymbol{I} + \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}})^{-1}(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{C}_2\\
&\quad - \boldsymbol{C}_1^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}}(\boldsymbol{I} + \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}\widetilde{\boldsymbol{\Psi}})^{-1}\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}(\widetilde{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}_0^{-1})\boldsymbol{C}_2.
\end{aligned} \tag{S16}$$

The Frobenius norm of each term on the right-hand side of (S16) and thus $\|\boldsymbol{C}_1^{\mathrm{T}}(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}_0^{-1})\boldsymbol{C}_2\|$ are bounded by $O(n^{-1/2}p_n^{1/2})\|\boldsymbol{C}_1\|\|\boldsymbol{C}_2\|$. Thus,

$$\|\boldsymbol{D}_\Psi^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_\Psi\| \leq \|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\| + \|\boldsymbol{D}_\Psi^{\mathrm{T}}(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_\Psi\| = \|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\| + O(n^{-1/2}p_n^{1/2})\|\boldsymbol{D}_\Psi\|^2.$$

By the proof of Lemma 1, $\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\| > cp_n^{-1}\|\boldsymbol{D}_\Psi\|^2$ for large enough n and some constant $c > 0$, such that the second term on the right-hand side of the equality above is dominated by the first term in the expression. We conclude that the left-hand side of (S15) is dominated by $\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|$ and thus $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

Clearly,

$$
\begin{aligned}
&\mathrm{tr}\big\{\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Psi}\big\}\\
\le&\,\mathrm{tr}\big\{\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Psi}}\big\}^{1/2}\\
&\times\mathrm{tr}\big(\boldsymbol{D}_{\Psi}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Psi}\big)^{1/2}\\
\le&\,\|\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{\Psi}}\|^{1/2}\\
&\times\|\boldsymbol{D}_{\Psi}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Psi}\|^{1/2}\mathrm{tr}\big(\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\big)^{1/2}\\
=&\,O(p_n^{-1/2})\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2\|\boldsymbol{D}_{\Psi}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Psi}\|^{1/2}\mathrm{tr}\big(\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\big)^{1/2}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
(\mathbb{P}_n-\mathbb{P})H_2(\widetilde{\boldsymbol{\theta}})&=(\mathbb{P}_n-\mathbb{P})O\{p_n^{-1/2}(1+\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2)\}\|\boldsymbol{D}_{\Psi}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Psi}\|^{1/2}\mathrm{tr}\big(\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\big)^{1/2}\\
&=o_p\big\{\|\boldsymbol{D}_{\Psi}^{\mathrm{T}}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Psi}\|^{1/2}\mathrm{tr}\big(\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\big)^{1/2}\big\}. \qquad \text{(S17)}
\end{aligned}
$$

We can show that

$$
\begin{aligned}
&\mathrm{tr}\{(\widetilde{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_{\Sigma}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_{\Sigma}\}\\
=&\,\mathrm{tr}[(\widetilde{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_{\Sigma}\{\boldsymbol{\Sigma}_0^{-1}-\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I}+\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\}\boldsymbol{D}_{\Sigma}]\\
=&\,\mathrm{tr}\{(\widetilde{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_{\Sigma}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma}\}-\mathrm{tr}\{(\widetilde{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_{\Sigma}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I}+\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma}\}\\
=&\sum_{j=1}^{p_n}(\boldsymbol{D}_{\Sigma}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma})_{jj}\mathrm{tr}\{\boldsymbol{e}_j^{\mathrm{T}}(\widetilde{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}_0^{-1})\boldsymbol{e}_j\}\\
&-\mathrm{tr}\{\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma}(\widetilde{\boldsymbol{\Omega}}^{-1}-\boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_{\Sigma}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0(\boldsymbol{I}+\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\}\\
\le&\,O(n^{-1/2}p_n^{1/2})\|\boldsymbol{D}_{\Sigma}\|^2+O(n^{-1/2}p_n^{1/2})\|\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma}\|^2\|(\boldsymbol{I}+\boldsymbol{\Psi}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}_0)^{-1}\|\\
=&\,o(\|\boldsymbol{D}_{\Sigma}\|^2), \qquad \text{(S18)}
\end{aligned}
$$

where $(\boldsymbol{D}_{\Sigma}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma})_{jj}$ denotes the (j,j)th element of $\boldsymbol{D}_{\Sigma}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{D}_{\Sigma}$, and $\boldsymbol{e}_j$ is a p_n-vector with 1 at the jth component and 0 elsewhere. Therefore, $\mathrm{tr}(\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma}\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{D}_{\Sigma})-\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_{\Sigma}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_{\Sigma})$

is dominated by $\|\boldsymbol{D}_\Sigma\|^2$ and thus $\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)$. We conclude that the left-hand side of (S17) is dominated by $\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)^{1/2}$ and thus $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$. Similar arguments on (S7)–(S10) yield

$$H_3(\widetilde{\boldsymbol{\theta}}) = O(1+\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2)\mathrm{tr}\big(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\big), \tag{S19}$$

$$H_4(\widetilde{\boldsymbol{\theta}}) = O(1)\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|, \tag{S20}$$

$$H_5(\widetilde{\boldsymbol{\theta}}) = O(p_n^{-1/2}\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|)\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|^{1/2}, \tag{S21}$$

$$H_6(\widetilde{\boldsymbol{\theta}}) = O(\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|)\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|^{1/2}\mathrm{tr}\big(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\big)^{1/2}. \tag{S22}$$

We conclude that $(\mathbb{P}_n-\mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta$ is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

Next, we derive the bound for the second term of (S14). For any element $v(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}})$ of $\boldsymbol{V}(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}})$, the fourth-order Taylor series expansion of $\mathrm{E}\{v(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}}) \mid Y,\boldsymbol{S}^{(O)};\widetilde{\boldsymbol{\theta}}\}$ at $\boldsymbol{\theta}_0$, together with condition (C6), yields

$$(\mathbb{P}_n-\mathbb{P})\mathrm{E}\big\{v(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}}) \mid Y,\boldsymbol{S}^{(O)};\widetilde{\boldsymbol{\theta}}\big\} = \sum_{k=0}^{3} O_p(n^{-1/2}p_n^{k/2})\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|^k + O_p(p_n^2)\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|^4 = O_p(n^{-1/2}).$$

Therefore, $(\mathbb{P}_n-\mathbb{P})\boldsymbol{d}_{(\alpha\beta\xi)}^{\mathrm{T}}\mathrm{E}\big\{\boldsymbol{V}_{\mathcal{S}}(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\xi}}) \mid Y,\boldsymbol{S}^{(O)};\widetilde{\boldsymbol{\theta}}\big\}\boldsymbol{d}_{(\alpha\beta\xi)}$ is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

By condition (C6),

$$(\mathbb{P}_n-\mathbb{P})\mathrm{E}(\|\boldsymbol{S}\|^2 \mid Y,\boldsymbol{S}^{(O)};\widetilde{\boldsymbol{\theta}}) = O_p(n^{-1/2}p_n).$$

In light of (S15), (S17), and (S19)–(S22), we conclude that $(\mathbb{P}_n-\mathbb{P})\boldsymbol{d}_\theta^{\mathrm{T}}\mathrm{E}\big\{\ddot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}}) \mid Y,\boldsymbol{S}^{(O)};\widetilde{\boldsymbol{\theta}}\big\}\boldsymbol{d}_\theta$ is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

For the third term of (S14), we can show that

$$\begin{aligned}&\dot{\boldsymbol{\ell}}_{\theta_{\mathcal{S}}}^{(C)}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{d}_\theta\\ &= O\{\|\dot{\boldsymbol{\ell}}_{(\alpha\beta\xi)}^{(C)}(\widetilde{\boldsymbol{\theta}})\|\}\|\boldsymbol{d}_{(\alpha\beta\xi)}\| + O(\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|)\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|^{1/2} + O(1+p_n^{-1/2}\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2)\\ &\quad\times\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2} + O\{p_n^{1/2}+\|(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})\circ(\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X})\|\}\|\boldsymbol{D}_\Sigma\|,\end{aligned} \tag{S23}$$

where $\dot{\boldsymbol{\ell}}^{(C)}_{(\alpha\beta\xi)}$ denotes the subvector of $\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}$ that corresponds to $(\boldsymbol{\alpha}, \boldsymbol{\beta}_{\mathcal{S}}, \boldsymbol{\xi})$. Hence,

$$\boldsymbol{d}_\theta^{\mathrm{T}} \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} \boldsymbol{d}_\theta = O_p(p_n)(\|\boldsymbol{d}_{(\alpha\beta\xi)}\|^2 + \|\boldsymbol{D}_\Gamma^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Gamma\| + \|\boldsymbol{D}_\Psi^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi\| + \|\boldsymbol{D}_\Sigma\|^2),$$

and by condition (C6), $(\mathbb{P}_n - \mathbb{P}) \boldsymbol{d}_\theta^{\mathrm{T}} \mathrm{E}\big\{\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \widetilde{\boldsymbol{\theta}}\big\} \boldsymbol{d}_\theta$ is dominated by $\mathbb{P} \boldsymbol{d}_\theta^{\mathrm{T}} \boldsymbol{K}_3(\boldsymbol{\theta}_0) \boldsymbol{d}_\theta$. The desired result follows. □

Proof of Lemma S3. It suffices to prove that

$$\begin{aligned}
&\mathbb{P}\Big|\boldsymbol{d}_\theta^{\mathrm{T}} \Big\{\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) - \ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}_0)\Big\} \boldsymbol{d}_\theta\Big| + \mathbb{P}\Big|\boldsymbol{d}_\theta^{\mathrm{T}} \Big\{\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} - \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}_0) \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}_0)^{\mathrm{T}}\Big\} \boldsymbol{d}_\theta\Big| \\
&+ \mathbb{P}\Big|\boldsymbol{d}_\theta^{\mathrm{T}} \Big[\mathrm{E}\Big\{\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \mid Y, \boldsymbol{S}^{(O)}; \widetilde{\boldsymbol{\theta}}\Big\} - \mathrm{E}\Big\{\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \mid Y, \boldsymbol{S}^{(O)}; \boldsymbol{\theta}_0\Big\}\Big] \boldsymbol{d}_\theta\Big| \\
&+ \mathbb{P}\Big|\boldsymbol{d}_\theta^{\mathrm{T}} \Big[\mathrm{E}\Big\{\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \widetilde{\boldsymbol{\theta}}\Big\} - \mathrm{E}\Big\{\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}}) \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} \mid Y, \boldsymbol{S}^{(O)}; \boldsymbol{\theta}_0\Big\}\Big] \boldsymbol{d}_\theta\Big| \qquad \text{(S24)}
\end{aligned}$$

is dominated by $\mathbb{P} \boldsymbol{d}_\theta^{\mathrm{T}} \boldsymbol{K}_3(\boldsymbol{\theta}_0) \boldsymbol{d}_\theta$. Consider the first term of (S24). By condition (C1),

$$\boldsymbol{d}_{(\alpha\beta\xi)}^{\mathrm{T}} \mathbb{P}\{\boldsymbol{V}_{\mathcal{S}}(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\xi}}) - \boldsymbol{V}_{\mathcal{S}}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\xi}_0)\} \boldsymbol{d}_{(\alpha\beta\xi)} = o(\|\boldsymbol{d}_{(\alpha\beta\xi)}\|^2),$$

so the left-hand side of the above equality is dominated by $\mathbb{P} \boldsymbol{d}_\theta^{\mathrm{T}} \boldsymbol{K}_3(\boldsymbol{\theta}_0) \boldsymbol{d}_\theta$. The difference between the first term on the right-hand side of (S5) evaluated at $\widetilde{\boldsymbol{\theta}}$ versus at $\boldsymbol{\theta}_0$ is

$$\begin{aligned}
&\mathrm{tr}(\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}} \widetilde{\boldsymbol{\Omega}}^{-1} \widetilde{\boldsymbol{\Psi}} \boldsymbol{D}_\Psi^{\mathrm{T}} \widetilde{\boldsymbol{\Omega}}^{-1} \boldsymbol{D}_\Psi - \boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Psi}_0 \boldsymbol{D}_\Psi^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi) \\
&\le \|\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}} \widetilde{\boldsymbol{\Omega}}^{-1} \widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Psi}_0\| \|\boldsymbol{D}_\Psi^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi\| + \|\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}} \widetilde{\boldsymbol{\Omega}}^{-1} \widetilde{\boldsymbol{\Psi}}\| \|\boldsymbol{D}_\Psi^{\mathrm{T}} (\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}_0^{-1}) \boldsymbol{D}_\Psi\| \\
&= O(n^{-1/2} p_n^{3/2}) \|\boldsymbol{D}_\Psi^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi\|.
\end{aligned}$$

For the second term of the right-hand side of (S5),

$$\begin{aligned}
&\mathrm{tr}\big(\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}} \widetilde{\boldsymbol{\Omega}}^{-1} \boldsymbol{D}_\Psi \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}} \widetilde{\boldsymbol{\Omega}}^{-1} \boldsymbol{D}_\Psi - \boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi \boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi\big) \\
&= O\{\|(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi \boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi + \widetilde{\boldsymbol{\Psi}}^{\mathrm{T}} (\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}_0^{-1}) \boldsymbol{D}_\Psi \boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi\|\} \\
&\le \|(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} (\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)\|^{1/2} \|\boldsymbol{\Psi}_0^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Psi}_0\|^{1/2} \|\boldsymbol{D}_\Psi^{\mathrm{T}} \boldsymbol{\Omega}_0^{-1} \boldsymbol{D}_\Psi\| + O(n^{-1/2} p_n^{1/2}) \|\boldsymbol{D}_\Psi\|^2
\end{aligned}$$

$$= O(n^{-1/2}p_n^{3/2})\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|.$$

By similar arguments,

$$H_1(\widetilde{\boldsymbol{\theta}}) - H_1(\boldsymbol{\theta}_0) = \{O(n^{-1/2}p_n^{3/2})(1 + \|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2) + O(n^{-1/2}p_n^{1/2})\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|\}\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|.$$

For the right-hand side of (S6), note that

$$\begin{aligned}
&\mathrm{tr}\{(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\} \\
\le\ &\mathrm{tr}\{(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)\}^{1/2}\mathrm{tr}(\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi)^{1/2} \\
\le\ &\|(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}(\widetilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0)\|^{1/2}\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)^{1/2} \\
\le\ &O(n^{-1/2}p_n)\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)^{1/2}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathrm{tr}\{\widetilde{\boldsymbol{\Psi}}^{\mathrm{T}}(\widetilde{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}_0^{-1})\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\} &\le O(n^{-1/2}p_n^{1/2})\|\widetilde{\boldsymbol{\Psi}}\|\|\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\| \\
&\le O(n^{-1/2}p_n)\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)^{1/2}.
\end{aligned}$$

By similar arguments,

$$\begin{aligned}
H_2(\widetilde{\boldsymbol{\theta}}) - H_2(\boldsymbol{\theta}_0) =&\{O(n^{-1/2}p_n) + O(n^{-1/2}p_n^{1/2})\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\| + O(n^{-1/2}p_n^{3/2})\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2\} \\
&\times \|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}\mathrm{tr}(\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Sigma)^{1/2}.
\end{aligned}$$

By (S18) and the arguments for bounding $H_1(\widetilde{\boldsymbol{\theta}}) - H_1(\boldsymbol{\theta}_0)$ and $H_2(\widetilde{\boldsymbol{\theta}}) - H_2(\boldsymbol{\theta}_0)$,

$$\begin{aligned}
H_3(\widetilde{\boldsymbol{\theta}}) - H_3(\boldsymbol{\theta}_0) &= \{O(n^{-1/2}p_n^{1/2})(1 + \|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2) + O(n^{-1/2}p_n)\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|\}\|\boldsymbol{D}_\Sigma\|^2, \\
H_4(\widetilde{\boldsymbol{\theta}}) - H_4(\boldsymbol{\theta}_0) &= O(n^{-1/2}p_n^{3/2})\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|, \\
H_5(\widetilde{\boldsymbol{\theta}}) - H_5(\boldsymbol{\theta}_0) &= O(n^{-1/2}p_n^{1/2})(1 + p_n^{1/2}\|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|)\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|^{1/2}\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2}, \\
H_6(\widetilde{\boldsymbol{\theta}}) - H_6(\boldsymbol{\theta}_0) &= O(n^{-1/2}p_n)(1 + \|\boldsymbol{S} - \widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|)\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|^{1/2}\|\boldsymbol{D}_\Sigma\|.
\end{aligned}$$

We conclude that the first term of (S24) is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

Next, we consider the second term of (S24). For some $\boldsymbol{\theta}^*$ between $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$,

$$\begin{aligned}
&\{\dot{\boldsymbol{\ell}}^{(C)}_{(\alpha\beta\xi)}(\widetilde{\boldsymbol{\theta}}) - \dot{\boldsymbol{\ell}}^{(C)}_{(\alpha\beta\xi)}(\boldsymbol{\theta}_0)\}^{\mathrm{T}}\boldsymbol{d}_{(\alpha\beta\xi)} = O(n^{-1/2}p_n^{1/2})\|\ddot{\boldsymbol{\ell}}^{(C)}_{(\alpha\beta\xi)}(\boldsymbol{\theta}^*)\|\|\boldsymbol{d}_{(\alpha\beta\xi)}\|,\\
&\quad \mathrm{tr}[\boldsymbol{D}_\Gamma^{\mathrm{T}}\{\dot{\boldsymbol{\ell}}^{(C)}_{\Gamma}(\widetilde{\boldsymbol{\theta}}) - \dot{\boldsymbol{\ell}}^{(C)}_{\Gamma}(\boldsymbol{\theta}_0)\}] = O(n^{-1/2}p_n)(1+\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|)\|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\|^{1/2},\\
&\quad \mathrm{tr}[\boldsymbol{D}_\Psi^{\mathrm{T}}\{\dot{\boldsymbol{\ell}}^{(C)}_{\Psi}(\widetilde{\boldsymbol{\theta}}) - \dot{\boldsymbol{\ell}}^{(C)}_{\Psi}(\boldsymbol{\theta}_0)\}] = O(n^{-1/2}p_n)(p_n^{1/2}+\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|+\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2)\|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\|^{1/2},\\
&\quad \mathrm{tr}[\boldsymbol{D}_\Sigma\{\dot{\boldsymbol{\ell}}^{(C)}_{\Sigma}(\widetilde{\boldsymbol{\theta}}) - \dot{\boldsymbol{\ell}}^{(C)}_{\Sigma}(\boldsymbol{\theta}_0)\}] = O(n^{-1/2}p_n^{1/2})(p_n^{1/2}+p_n^{1/2}\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|+\|\boldsymbol{S}-\widetilde{\boldsymbol{\Gamma}}\boldsymbol{X}\|^2)\|\boldsymbol{D}_\Sigma\|,
\end{aligned}$$

where $\ddot{\boldsymbol{\ell}}^{(C)}_{(\alpha\beta\xi)}$ denotes the submatrix of $\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}$ that corresponds to $(\boldsymbol{\alpha}, \boldsymbol{\beta}_{\mathcal{S}}, \boldsymbol{\xi})$. It follows from (S23) that

$$\begin{aligned}
&\mathbb{P}\left|\boldsymbol{d}_\theta^{\mathrm{T}}\left\{\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})^{\mathrm{T}} - \dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}_0)\dot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\boldsymbol{\theta}_0)^{\mathrm{T}}\right\}\boldsymbol{d}_\theta\right|\\
\le\,& O(n^{-1/2}p_n^{5/2})(\|\boldsymbol{d}_{(\alpha\beta\xi)}\|^2 + \|\boldsymbol{D}_\Gamma^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Gamma\| + \|\boldsymbol{D}_\Psi^{\mathrm{T}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{D}_\Psi\| + \|\boldsymbol{D}_\Sigma\|^2),
\end{aligned}$$

which is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$.

For the third term of (S24), we note that

$$\begin{aligned}
&\mathbb{P}\int \boldsymbol{d}_\theta^{\mathrm{T}}\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta\{f(\boldsymbol{S}\mid Y,\boldsymbol{S}^{(O)};\widetilde{\boldsymbol{\theta}}) - f(\boldsymbol{S}\mid Y,\boldsymbol{S}^{(O)};\boldsymbol{\theta}_0)\}\,\mathrm{d}\boldsymbol{S}^{(M)}\\
=\,&\mathbb{P}\int \boldsymbol{d}_\theta^{\mathrm{T}}\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta\dot{\boldsymbol{f}}(\boldsymbol{S}\mid Y,\boldsymbol{S}^{(O)};\boldsymbol{\theta}^*)^{\mathrm{T}}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\,\mathrm{d}\boldsymbol{S}^{(M)}\\
\le\,&\left\|\mathbb{P}\int \boldsymbol{d}_\theta^{\mathrm{T}}\ddot{\boldsymbol{\ell}}^{(C)}_{\theta_{\mathcal{S}}}(\widetilde{\boldsymbol{\theta}})\boldsymbol{d}_\theta\dot{\boldsymbol{f}}(\boldsymbol{S}\mid Y,\boldsymbol{S}^{(O)};\boldsymbol{\theta}^*)\,\mathrm{d}\boldsymbol{S}^{(M)}\right\|\|\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|,
\end{aligned}$$

where $\boldsymbol{\theta}^*$ is some value between $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. By condition (C6), (S15), (S17), and (S19)–(S22), the first term on the right-hand side of the inequality above is $O(p_n^{3/2})\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$, so the third term of (S24) is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$. Likewise, by (S23) and condition (C6), the fourth term of (S24) is dominated by $\mathbb{P}\boldsymbol{d}_\theta^{\mathrm{T}}\boldsymbol{K}_3(\boldsymbol{\theta}_0)\boldsymbol{d}_\theta$. □

Table S1: Simulation results for the continuous outcome variable under a misspecified factor model

	Lasso				A-Lasso			
	Variables selected	FDR	TPR	Pred. error	Variables selected	FDR	TPR	Pred. error
MCAR; $p = 100$								
Complete	28.3	0.379	0.570	0.120	22.2	0.262	0.540	0.116
SMC	32.1	0.369	0.657	0.097	25.7	0.251	0.632	0.094
Imputed	33.6	0.379	0.679	0.090	26.4	0.246	0.655	0.088
Proposed	28.8	0.321	0.643	0.090	24.1	0.214	0.625	0.090
MAR; $p = 100$								
Complete	32.8	0.364	0.675	0.171	26.2	0.228	0.664	0.191
SMC	29.1	0.341	0.621	0.107	23.3	0.220	0.596	0.105
Imputed	29.0	0.330	0.624	0.116	22.7	0.196	0.595	0.117
Proposed	30.1	0.277	0.717	0.078	25.5	0.182	0.689	0.077
MCAR; $p = 300$								
Complete	34.2	0.532	0.500	0.149	27.4	0.432	0.502	0.133
SMC	35.7	0.485	0.587	0.118	29.1	0.380	0.588	0.107
Imputed	39.9	0.520	0.609	0.107	30.1	0.380	0.609	0.098
Proposed	30.0	0.446	0.544	0.116	26.0	0.334	0.567	0.101
MAR; $p = 300$								
Complete	39.7	0.507	0.612	0.163	29.7	0.362	0.615	0.190
SMC	31.3	0.441	0.559	0.129	25.5	0.335	0.554	0.119
Imputed	36.0	0.454	0.618	0.125	26.9	0.293	0.613	0.119
Proposed	32.3	0.358	0.679	0.094	28.2	0.269	0.679	0.079

NOTE: See NOTE to Table 1.

Table S2: Simulation results for the binary outcome variable under a misspecified factor model

	Lasso				A-Lasso			
	Variables selected	FDR	TPR	Pred. error	Variables selected	FDR	TPR	Pred. error
MCAR; $p = 100$								
Complete	24.3	0.382	0.484	1.558	18.4	0.271	0.442	1.417
SMC	27.5	0.373	0.563	1.263	20.5	0.240	0.512	1.205
Imputed	27.3	0.392	0.541	1.277	20.4	0.257	0.500	1.196
Proposed	27.5	0.391	0.546	1.258	20.5	0.257	0.500	1.190
MAR; $p = 100$								
Complete	28.6	0.373	0.587	1.210	21.2	0.230	0.540	1.089
SMC	26.3	0.351	0.556	1.287	20.1	0.224	0.515	1.172
Imputed	23.7	0.350	0.500	1.383	17.9	0.217	0.463	1.333
Proposed	28.3	0.357	0.595	1.198	21.5	0.226	0.549	1.072
MCAR; $p = 300$								
Complete	29.6	0.565	0.404	1.908	22.7	0.460	0.395	1.682
SMC	32.4	0.519	0.496	1.522	24.1	0.375	0.486	1.276
Imputed	34.9	0.572	0.472	1.568	24.9	0.425	0.465	1.334
Proposed	33.3	0.575	0.448	1.611	25.7	0.437	0.467	1.301
MAR; $p = 300$								
Complete	34.5	0.536	0.512	1.528	24.5	0.373	0.500	1.246
SMC	29.8	0.498	0.474	1.632	22.6	0.360	0.470	1.356
Imputed	28.8	0.498	0.458	1.586	21.9	0.359	0.456	1.464
Proposed	35.4	0.518	0.542	1.443	26.2	0.381	0.528	1.163

NOTE: See NOTE to Table 1.

Table S3: Selected features in the analysis of the TCGA data

Feature type	Feature name	Feature type	Feature name
Gene expression	WDR37 (22884)	Gene expression	SAMD8 (142891)
	FUT7 (2529)		USP38 (84640)
	DDIT4 (54541)		IMPG2 (50939)
	FCRLA (84824)		TRMT12 (55039)
	TRAFD1 (10906)		VAV3 (10451)
	C10orf2 (56652)		RMND1 (55005)
	STK17B (9262)		DNAJC5 (80331)
	STC1 (6781)		SSBP3 (23648)
	INHA (3623)		CCL3L3 (414062)
	ZNF683 (257101)		GOLGA6L5 (374650)
	C6orf62 (81688)		LASS5 (91012)
	CCBL1 (883)		ZNF124 (7678)
	ZC3HAV1 (56829)		YTHDF2 (51441)
	HOXA11AS (221883)		KLRF1 (51348)
	GRIP2 (80852)		ATG12 (9140)
	NCOA5 (57727)		XKR3 (150165)
	SCP2 (6342)		LOC389333 (389333)
	IGFBP1 (3484)		GDF7 (151449)
	GLUD1 (2746)	Protein expression	53BP1
	MTM1 (4534)		AMPK_alpha
	PTPRVP (148713)		ARID1A
	KCNF1 (3754)		Chromogranin-A-N-term
	TMEM22 (80723)		EGFR
	UPK2 (7379)		EGFR_pY1068
	RPL13AP20 (387841)		Ku80
	PCOLCE2 (26577)		p16_INK4a
	NUP62CL (54830)		

NOTE: For gene expressions, the Entrez gene ID is given in parentheses.

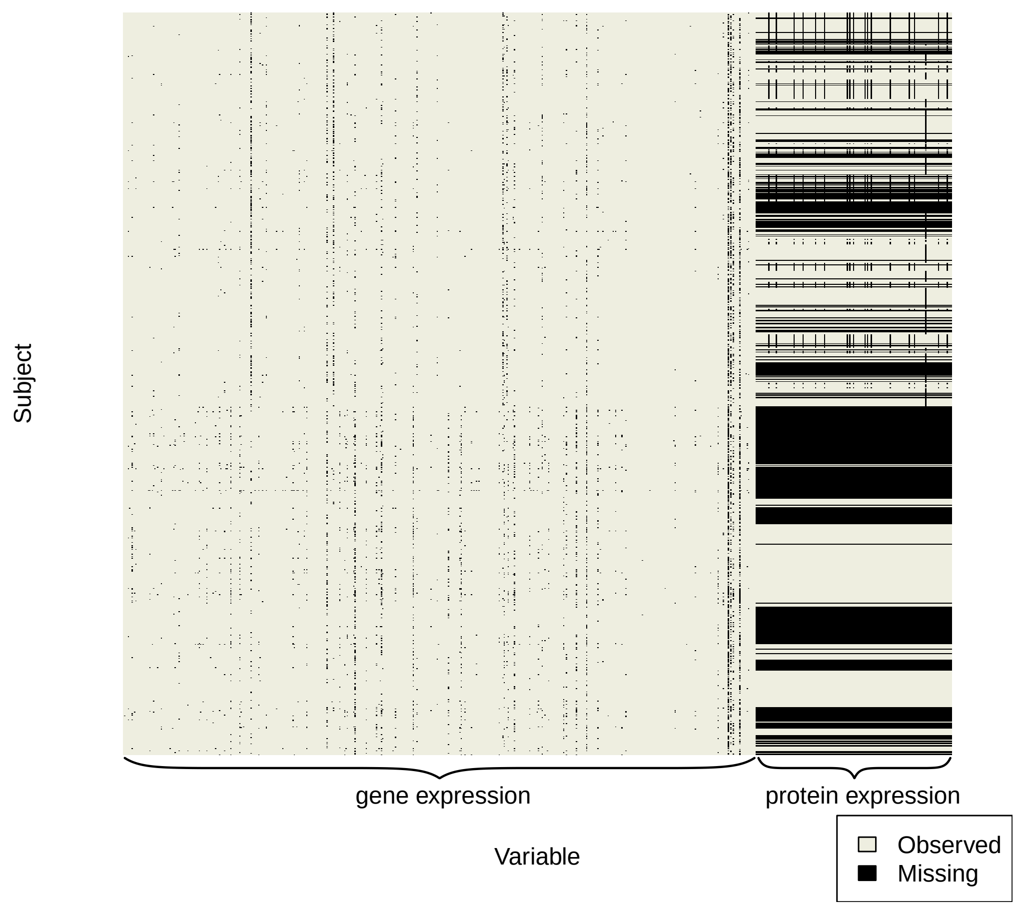

Figure S1: Missing-data pattern in the TCGA data.