Poisson Regression with Error Corrupted High Dimensional Features

The University of California, San Francisco, Pennsylvania State University

Supplementary Material

S.1 Regularity Conditions

We first restate and introduce some new notations to facilitate the theoretic derivations. For a matrix $\mathbf{M}$, let $\|\mathbf{M}\|_{\max}$ be the matrix maximum norm, $\|\mathbf{M}\|_\infty$ be the l_∞ norm and $\|\mathbf{M}\|_p$ be the l_p norm. Let $\mathcal{F}(\boldsymbol{\beta})$ be the σ-field generated by $\mathbf{X}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, i = 1, \ldots, n$. Further, let $\mathcal{F}_x$ be the sigma-field generated by $\mathbf{X}_i, i = 1, \ldots, n$. For a general vector $\mathbf{a}$, let $\|\mathbf{a}\|_\infty$ be the vector sup-norm, $\|\mathbf{a}\|_p$ be the vector l_p-norm. Let $\mathbf{e}_j$ be the unit vector with 1 on its jth entry. For a vector $\mathbf{v} = (v_1, \ldots, v_m)^{\mathrm{T}}$, let $\mathrm{supp}(\mathbf{v})$ be the set of indices with $v_i \neq 0$ and $\|\mathbf{v}\|_0 = |\mathrm{supp}(\mathbf{v})|$, where $|\mathcal{U}|$ stands for the cardinality of the set $\mathcal{U}$. Let $\mathbb{K}(s) \equiv \{\mathbf{v} \in \mathbb{R}^p : \|\mathbf{v}\|_2 \leqslant 1, \|\mathbf{v}\|_0 \leqslant s\}$. Let $\alpha_{\min}(\mathbf{M})$ and $\alpha_{\max}(\mathbf{M})$ be the minimal and maximal eigenvalues of the matrix $\mathbf{M}$, respectively. To simplify the notation, we define

$$\alpha_{\min}(\boldsymbol{\beta}) \equiv \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}\}],$$

and

$$\alpha_{\max}(\boldsymbol{\beta}) \equiv \alpha_{\max}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})\mathbf{X}\mathbf{X}^{\mathrm{T}}\}].$$

Further, we define $\|X\|_{\psi_1} \equiv \sup_{k\geqslant 1} k^{-1}E(|X|^k)^{1/k}$, and $\|X\|_{\psi_2} \equiv \sup_{k\geqslant 1} k^{-1/2}E(|X|^k)^{1/k}$. For notational convenience, let $A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i) \equiv \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)$ and $g(\mathbf{W}_i, \boldsymbol{\beta}, \mathbf{v}) \equiv \mathbf{v}^{\mathrm{T}}\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\}\mathbf{v}$.

(C1) For any $\boldsymbol{\beta}$ with $\|\boldsymbol{\beta}\|_2 \leqslant 2b_0$,

$$D_1 \leqslant \alpha_{\min}(\boldsymbol{\beta}) \leqslant \alpha_{\max}(\boldsymbol{\beta}) \leqslant D_2.$$

Here D_1, D_2 are positive constants.

(C2) For $j = 1, \ldots, p$, define $K_j \equiv \|U_{ij}\|_{\psi_2}$

$$K_j = (2\Omega_{jj})^{1/2} \sup_{k\geqslant 1} k^{-1/2}\pi^{-1/(2k)}\Gamma^{1/k}\{(k+1)/2\},$$

where Γ is the Gamma function, then there exist constants m_0, M_0 so that $m_0 < K_j^2 \sum_{i=1}^n Y_i^2/n < M_0$ uniformly for all j almost surely.

(C3) Define

$$K_Y(\mathbf{X}_i) \equiv \sup_{k\geqslant 1} k^{-1}E[|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)|^k|\mathbf{X}_i]^{1/k}.$$

There exist constants m_1, m_2, M_1, M_2 so that uniformly for all $j = 1, \ldots, p$,

$$m_1 < n^{-1}\sum_{i=1}^n X_{ij}^2 K_Y(\mathbf{X}_i)^2 < M_1$$

and

$$\max_i |X_{ij}| K_Y(\mathbf{X}_i)\{\log(n)\}^{-1} < M_2$$

almost surely.

(C4) The sample size n and the dimension of covariates p satisfy the relation $\log(n)\sqrt{\log(p)/n} \leqslant C$ for an absolute constant C.

(C5) For $\mathbf{e}_j$, $j = 1, \ldots, p$, define

$$\begin{aligned} K_{wij}(\boldsymbol{\beta}_0) \quad \equiv \quad & \sup_{k \geqslant 1} k^{-1/2} E[|(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j \\ & -E\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j \\ & |\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\}|^k |\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i]^{1/k}, \end{aligned}$$

which is the conditional sub-Gaussian norm according to Definition 1 in Section S.4. Then $E\{K_{wij}(\boldsymbol{\beta}_0)^4\} < Q_0$. In addition, there exist constants m_3, M_3 and Q_1 so that (i)

$$m_3 < \sum_{i=1}^{n} K_{wij}(\boldsymbol{\beta}_0)^2 A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i)^2/n < M_3,$$

and (ii)

$$\left| \sum_{i=1}^{n} \{n\log(p)\}^{-1/2} E\{A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i)(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j \right. \\ \left. |\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j \right| < Q_1$$

uniformly for all $j = 1, \ldots, p$ in probability.

(C6) Let $\mathbf{v}$ be a unit vector, let $\boldsymbol{\beta}$ satisfy $\|\boldsymbol{\beta}\|_2 \leqslant 2b_0$, and let

$$\begin{aligned} K_{gvi}(\boldsymbol{\beta}) \quad \equiv \quad & \sup_{k>1} k^{-1} E(|[g(\mathbf{W}_i, \boldsymbol{\beta}, \mathbf{v}) \\ & -E\{g(\mathbf{W}_i, \boldsymbol{\beta}, \mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\}]|^k \end{aligned}$$

$$|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i)^{1/k},$$

which is the conditional sub-exponential norm according to Definition 2 in Section S.4. Then $E\{K_{gvi}(\boldsymbol{\beta})^4\} < Q_{01}$, and

$$E[\exp\{A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)K_{gvi}^2(\boldsymbol{\beta})\}] < Q_{02}.$$

In addition, for all $\mathbf{v}$,

$$m_4 < \sum_{i=1}^n |A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|^2 K_{gvi}(\boldsymbol{\beta})^2/n \quad < \quad M_4, \tag{S.1}$$

$$m_5 < \max_i |A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|K_{gvi}(\boldsymbol{\beta})/\mathrm{log}n \quad < \quad M_5, \tag{S.2}$$

and

$$\frac{1}{\sqrt{n}}\| \sum_{i=1}^n (A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E[\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} \\ |\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i] - E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\})\|_2 < Q_2, \tag{S.3}$$

in probability.

S.2 Examples to justify Regularity Conditions (C5) and (C6)

Example when Condition (C5) holds

When

$$M_{\boldsymbol{\Omega}} \equiv \|\boldsymbol{\Omega}\|_2 = O(1) \tag{S.4}$$

$$\text{and} \qquad \|\boldsymbol{\Sigma}_{\mathbf{X}}\|_2 = O(1), \tag{S.5}$$

where let $\boldsymbol{\Sigma}_{\mathbf{X}} = \mathrm{cov}(\mathbf{X})$, and note that $\|\boldsymbol{\beta}_0\|_2 \leqslant b_0$. Then, since $E(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{U}_i) = 0$ and $\mathrm{var}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{U}_i) = \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0 \leqslant \|\boldsymbol{\beta}_0\|_2^2\|\boldsymbol{\Omega}\|_2 \leqslant b_0^2 M_{\boldsymbol{\Omega}}$, we get $\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0 = O(1)$ and $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{U}_i = O_p(1)$. Therefore

$$\exp(2\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{U}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0) = O_p(1) \tag{S.6}$$

by the continuous mapping theorem. Similarly by (S.5), we have $\exp(2\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i) = O_p(1)$, and hence we have $E\{\exp(4\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - 2\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0)\} = O(1)$. Hence,

$$\begin{aligned} & \sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2 \exp(2\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0)/n \\ \leqslant \; & \frac{\sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^4/(2n) + \exp(4\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - 2\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0)}{2n} \\ = \; & Q_0/2 + E\{\exp(4\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - 2\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0)\}/2 + o_p(1). \end{aligned}$$

Hence the upper bound condition in the first statement is satisfied. Statement (ii) holds, $E[E\{A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i)(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j] = 0$. Further, $E\{A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i)(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j = O_p(1)$. This is because $A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i) = O_p(1)$, $|\mathbf{e}_j^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0| \leqslant \|\mathbf{e}_j\|_2\|\boldsymbol{\Omega}\boldsymbol{\beta}_0\|_2 = (\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\Omega}\boldsymbol{\beta}_0)^{1/2} \leqslant (\|\boldsymbol{\beta}_0\|_2^2 M_{\boldsymbol{\Omega}}^2)^{1/2} = b_0 M_{\boldsymbol{\Omega}}$ by (S.4) and $U_{ij} = O_p(1)$, $X_{ij} = O_p(1)$. Hence

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^n [E\{A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i)(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} \\ & - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j] = O_p(1), \end{aligned}$$

which suggests

$$\{n\log(p)\}^{-1/2}\sum_{i=1}^{n}[E\{A(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i)(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\}$$
$$-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j]=o_p(1),$$

in probability. Hence (ii) holds.

Example when Condition (C6) holds

Under (S.4) and (S.5), using the same arguments as those lead to (S.6), we have $A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)^4=O_p(1)$. Hence

$$\begin{aligned}&\sum_{i=1}^{n}A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)^2K_{gvi}(\boldsymbol{\beta})^2/n\\=\;&E\{A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)^4\}/2+E\{K_{gvi}(\boldsymbol{\beta})^4\}/2+o_p(1),\end{aligned}$$

which is bounded in probability. Hence the upper bound in (S.1) is satisfied. Further, it is easy to see that

$$\begin{aligned}&\Pr\{\max_i A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)K_{gvi}(\boldsymbol{\beta})/\sqrt{\log n}>\sqrt{2}\}\\\leqslant\;&\exp\{-2\log(n)+\log(n)\}E[\exp\{A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)K_{gvi}^2(\boldsymbol{\beta})\}]\\\leqslant\;&Q_{02}/n,\end{aligned}$$

in probability. Hence the upper bound in (S.2) is satisfied. Now recall that $g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})\equiv\mathbf{v}^{\mathrm{T}}\{(\mathbf{W}_i-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega})^{\otimes 2}-\boldsymbol{\Omega}\}\mathbf{v}$. (S.4) and (S.5) together also implies (S.3). To see this, for any unit vector $\mathbf{v}$, we have

$$\begin{aligned}
& \|A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} \\
& -E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\|_2 \\
= & \sup_{\mathbf{v}} A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)\mathbf{v}^{\mathrm{T}}E\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\}\mathbf{v} \\
& + \sup_{\mathbf{v}} \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}.
\end{aligned}$$

We can see that in the last line, the terms inside the expectations are functions of $A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)$, $\mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}$, $\mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{v}$, $\mathbf{v}^{\mathrm{T}}\mathbf{W}_i$, $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i$, $\mathbf{v}^{\mathrm{T}}\mathbf{X}_i$. We now show the boundedness of each term. $A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i) = O_p(1)$ as we have pointed out in (S.6). Further, $|\mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}| \leqslant \|\mathbf{v}\|_2\|\boldsymbol{\Omega}\boldsymbol{\beta}\|_2 \leqslant \|\mathbf{v}\|_2\sqrt{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\Omega}\boldsymbol{\beta}} = 2\|\mathbf{v}\|_2 b_0 M_{\boldsymbol{\Omega}}$. Further, because $\mathrm{var}(\mathbf{v}^{\mathrm{T}}\mathbf{U}_i) = \mathbf{v}^{\mathrm{T}}\boldsymbol{\Omega}\mathbf{v} = O(1)$, this leads to $|\mathbf{v}^{\mathrm{T}}\mathbf{U}_i| = O_p(1)$. Similarly, $\|\boldsymbol{\Sigma}_{\mathbf{X}}\|_2 = O(1)$. Moreover, because also $\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{v} \leqslant \|\boldsymbol{\Sigma}_{\mathbf{X}}\|_2 = O(1)$, $|\mathbf{v}^{\mathrm{T}}\mathbf{X}_i| = O_p(1)$. Therefore, $|\mathbf{v}^{\mathrm{T}}\mathbf{W}_i| \leqslant |\mathbf{v}^{\mathrm{T}}\mathbf{X}_i| + |\mathbf{v}^{\mathrm{T}}\mathbf{U}_i| = O_p(1)$. Further, $\mathrm{var}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{\beta} \leqslant 4b_0^2\|\boldsymbol{\Sigma}_{\mathbf{X}}\|_2 = O(1)$. Hence, $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i = O_p(1)$. By the continuous mapping theorem, we have

$$\begin{aligned}
& \mathrm{var}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)\mathbf{v}^{\mathrm{T}}E\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\}\mathbf{v} \\
& -\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\mathbf{v}] = O(1).
\end{aligned}$$

Further,

$$\begin{aligned}
& E[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} \\
& -\boldsymbol{\Omega}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}] = \mathbf{0}.
\end{aligned}$$

Therefore by the weak law of large numbers, (S.3) holds.

S.3 Proofs of the Theorems

Proof of Theorem 1: Define

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}) &= \left[-n^{-1}\sum_{i=1}^{n}\{Y_i\mathbf{W}_i^{\mathrm{T}}\boldsymbol{\beta}\right.\\ &\quad \left.-\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\}+\lambda\|\boldsymbol{\beta}\|_1\right]\end{aligned}$$

be the objective function, hence $\mathcal{L}(\widehat{\boldsymbol{\beta}}) \leqslant \mathcal{L}(\boldsymbol{\beta}_0)$, where $\|\widehat{\boldsymbol{\beta}}\|_1 \leqslant b_0\sqrt{k}$. Define the error vector $\widehat{\mathbf{v}} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$, we expand $n^{-1}\sum_{i=1}^{n}\{Y_i\mathbf{W}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \exp(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{W}_i - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\Omega}\widehat{\boldsymbol{\beta}}/2)\}$ at $\boldsymbol{\beta}_0$ and obtain

$$\begin{aligned}0 &\geqslant \mathcal{L}(\widehat{\boldsymbol{\beta}})-\mathcal{L}(\boldsymbol{\beta}_0)\\ &= n^{-1}\sum_{i=1}^{n}\{Y_i\widehat{\mathbf{v}}^{\mathrm{T}}\mathbf{W}_i-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)\\ &\quad \times\widehat{\mathbf{v}}^{\mathrm{T}}(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta}_0)\}\\ &\quad +n^{-1}1/2\widehat{\mathbf{v}}^{\mathrm{T}}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}^{*}/2)\\ &\quad \times\{(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta}^{*})^{\otimes 2}-\boldsymbol{\Omega}\}\widehat{\mathbf{v}}+\lambda\|\boldsymbol{\beta}_0+\widehat{\mathbf{v}}\|_1-\lambda\|\boldsymbol{\beta}_0\|_1,\end{aligned}$$

where $\boldsymbol{\beta}^{*}$ is on the line connecting $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. Hence we have the inequality that

$$\begin{aligned}&n^{-1}1/2\widehat{\mathbf{v}}^{\mathrm{T}}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}^{*}/2)\\ &\times\{(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta}^{*})^{\otimes 2}-\boldsymbol{\Omega}\}\widehat{\mathbf{v}}\end{aligned}$$

$$\begin{aligned}&\leqslant -n^{-1}\sum_{i=1}^{n}\{Y_i\widehat{\mathbf{v}}^{\mathrm{T}}\mathbf{W}_i-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)\\&\quad\times\widehat{\mathbf{v}}^{\mathrm{T}}(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta}_0)\}+\lambda\|\boldsymbol{\beta}_0\|_1-\lambda\|\boldsymbol{\beta}_0+\widehat{\mathbf{v}}\|_1. \qquad (S.1)\end{aligned}$$

We first derive the upper bound of (S.1).

First note that let

$$\begin{aligned}\phi &\geqslant 3\max\{4e\sqrt{M_1}, 8eM_2C, 2c_{10}M_3Q_1(1+r)/m_3,\\&\quad \sqrt{2}\sqrt{36e^2M_0}, 1\},\end{aligned}$$

so

$$\begin{aligned}&\|n^{-1}\sum_{i=1}^{n}\{Y_i\mathbf{W}_i-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)\\&\times(\mathbf{W}_i-\Omega\boldsymbol{\beta}_0)\}\|_\infty\\&\leqslant \phi\sqrt{\log(p)/n}\\&\leqslant \phi\{\log(p)/n\}^{1/4}, \qquad (S.2)\end{aligned}$$

with probability at least $1-6/p$ by Lemma 3. Hence

$$\begin{aligned}&|n^{-1}\sum_{i=1}^{n}\{\widehat{\mathbf{v}}^{\mathrm{T}}Y_i\mathbf{W}_i-\widehat{\mathbf{v}}^{\mathrm{T}}\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)\\&\times(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta}_0)\}|\\&\leqslant \|\widehat{\mathbf{v}}\|_1\phi\{\log(p)/n\}^{1/4}\\&= (\|\widehat{\mathbf{v}}_S\|_1+\|\widehat{\mathbf{v}}_{S^c}\|_1)\phi\{\log(p)/n\}^{1/4} \qquad (S.3)\end{aligned}$$

with probability at least $1-6/p$, where S is the index set of the nonzero elements in $\boldsymbol{\beta}_0$. Here for a vector $\mathbf{a}=(a_1,\ldots,a_m)^{\mathrm{T}}$, and an index set S,

$\mathbf{a}_S = \{a_1 I(1 \in S), \ldots, a_m I(m \in S)\}^{\mathrm{T}}$. On the other hand, we have

$$\begin{aligned}
& \|\boldsymbol{\beta}_0 + \widehat{\mathbf{v}}\|_1 + \|\widehat{\mathbf{v}}_S\|_1 \\
\geqslant\ & \|\boldsymbol{\beta}_0 + \widehat{\mathbf{v}} - \widehat{\mathbf{v}}_S\|_1 \\
=\ & \|\boldsymbol{\beta}_0 + \widehat{\mathbf{v}}_{S^c}\|_1 \\
=\ & \|\boldsymbol{\beta}_{0S}\|_1 + \|\widehat{\mathbf{v}}_{S^c}\|_1.
\end{aligned}$$

Hence

$$\begin{aligned}
& \|\boldsymbol{\beta}_0 + \widehat{\mathbf{v}}\|_1 - \|\boldsymbol{\beta}_0\|_1 && \text{(S.4)} \\
\geqslant\ & \{\|\boldsymbol{\beta}_0\|_1 - \|\widehat{\mathbf{v}}_S\|_1\} + \|\widehat{\mathbf{v}}_{S^c}\|_1 - \|\boldsymbol{\beta}_0\|_1 && \text{(S.5)} \\
=\ & \|\widehat{\mathbf{v}}_{S^c}\|_1 - \|\widehat{\mathbf{v}}_S\|_1.
\end{aligned}$$

Combine (S.3) and (S.4), and recall that $\lambda > 8/3\phi\{\log(p)/n\}^{1/4}$, we have that the right hand side of (S.1) is upper bounded by

$$\begin{aligned}
& (\|\widehat{\mathbf{v}}_S\|_1 + \|\widehat{\mathbf{v}}_{S^c}\|_1)\phi\{\log(p)/n\}^{1/4} + \lambda\|\widehat{\mathbf{v}}_S\|_1 - \lambda\|\widehat{\mathbf{v}}_{S^c}\|_1 \\
\leqslant\ & 11/8\lambda\|\widehat{\mathbf{v}}_S\|_1 - 5/8\lambda\|\widehat{\mathbf{v}}_{S^c}\|_1,
\end{aligned}$$

i.e.

$$\begin{aligned}
& n^{-1}1/2\widehat{\mathbf{v}}^{\mathrm{T}}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}^*/2) \\
& \times\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}^*)^{\otimes 2} - \boldsymbol{\Omega}\}\widehat{\mathbf{v}} \\
\leqslant\ & (\|\widehat{\mathbf{v}}_S\|_1 + \|\widehat{\mathbf{v}}_{S^c}\|_1)\phi\{\log(p)/n\}^{1/4} + \lambda\|\widehat{\mathbf{v}}_S\|_1 - \lambda\|\widehat{\mathbf{v}}_{S^c}\|_1
\end{aligned}$$

$$\leqslant \quad 11/8\lambda\|\widehat{\mathbf{v}}_S\|_1 - 5/8\lambda\|\widehat{\mathbf{v}}_{S^c}\|_1. \tag{S.6}$$

Further, because $\|\boldsymbol{\beta}^*\| \leqslant \|\widehat{\boldsymbol{\beta}}\|_2 + \|\boldsymbol{\beta}_0\|_2 \leqslant 2b_0$, Lemma 4 implies that $n^{-1}\sum_{i=1}^n \exp({\boldsymbol{\beta}^*}^{\mathrm{T}}\mathbf{W}_i - {\boldsymbol{\beta}^*}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}^*/2)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}^*)^{\otimes 2} - \boldsymbol{\Omega}\}$ satisfies the lower and upper-RE conditions.

Hence,

$$\begin{aligned}
& n^{-1}\widehat{\mathbf{v}}^{\mathrm{T}}\sum_{i=1}^n \exp({\boldsymbol{\beta}^*}^{\mathrm{T}}\mathbf{W}_i - {\boldsymbol{\beta}^*}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}^*/2) \\
& \times\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}^*)^{\otimes 2} - \boldsymbol{\Omega}\}\widehat{\mathbf{v}} \\
\geqslant\; & \alpha_{\min}[E\{\exp({\boldsymbol{\beta}^*}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 - 1/(2c)\}\|\widehat{\mathbf{v}}\|_2^2 \\
& -\tau_1(n,p)\|\widehat{\mathbf{v}}\|_1^2 \\
=\; & \alpha_{\min}[E\{\exp({\boldsymbol{\beta}^*}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 - 1/(2c)\}\|\widehat{\mathbf{v}}\|_2^2 \\
& -\tau(n,p)\|\widehat{\mathbf{v}}\|_1^2.
\end{aligned} \tag{S.7}$$

Here $\tau_1(n,p)$ is the $\tau(n,p)$ given in Lemma 4, and $\tau(n,p)$ is defined in the statement of Theorem 1. The above equality holds because first

$$\begin{aligned}
& \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}} \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]/\{(2c\sqrt{s})\} \\
=\; & \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}} \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]/\{(2c\sqrt{c_1})\} \\
& \times\{\log(p)/n\}^{1/4} \\
\leqslant\; & \phi/b_0\{\log(p)/n\}^{1/4},
\end{aligned}$$

by the definition of ϕ in the statement. Hence

$$\sqrt{s}\tau(n,p)$$

$$
\begin{aligned}
&= \min\left[\sup_{\{\boldsymbol\beta:\|\boldsymbol\beta\|_2\leqslant 2b_0\}} \alpha_{\min}[E\{\exp(\boldsymbol\beta^{\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]\right.\\
&\quad \left./\{(2c\sqrt{s})\}, \phi/b_0\{\log(p)/n\}^{1/4}\right]\\
&= \sup_{\{\boldsymbol\beta:\|\boldsymbol\beta\|_2\leqslant 2b_0\}} \alpha_{\min}[E\{\exp(\boldsymbol\beta^{\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]/\{(2c\sqrt{s})\}\\
&= \sqrt{s}\tau_1(n,p).
\end{aligned}
$$

Now combine (S.6) and (S.7), we have

$$
\begin{aligned}
&\quad -1/2\tau(n,p)\|\widehat{\mathbf v}\|_1^2\\
&\leqslant 1/2\alpha_{\min}[E\{\exp(\boldsymbol\beta^{*\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]\{1-1/(2c)\}\|\widehat{\mathbf v}\|_2^2\\
&\quad -1/2\tau(n,p)\|\widehat{\mathbf v}\|_1^2\\
&\leqslant n^{-1}1/2\widehat{\mathbf v}^{\mathrm T}\sum_{i=1}^n \exp(\boldsymbol\beta^{*\mathrm T}\mathbf W_i-\boldsymbol\beta^{*\mathrm T}\boldsymbol\Omega\boldsymbol\beta^*/2)\\
&\quad \times\{(\mathbf W_i-\boldsymbol\Omega\boldsymbol\beta^*)^{\otimes 2}-\boldsymbol\Omega\}\widehat{\mathbf v}\\
&\leqslant 11/8\lambda\|\widehat{\mathbf v}_S\|_1-5/8\lambda\|\widehat{\mathbf v}_{S^c}\|_1
\end{aligned}
$$

as long as $2c>1$. Further $\|\widehat{\mathbf v}\|_1\leqslant\|\widehat{\boldsymbol\beta}\|_1+\|\boldsymbol\beta_0\|_1\leqslant 2b_0\sqrt{k}$, and $\sqrt{s}\tau(n,p)\leqslant \phi\{\log(p)/n\}^{1/4}/b_0$. Therefore,

$$
1/2\tau(n,p)\|\widehat{\mathbf v}\|_1^2\leqslant\phi\{\log(p)/n\}^{1/4}\|\widehat{\mathbf v}\|_1\leqslant 3/8\lambda\|\widehat{\mathbf v}\|_1.
$$

Combining the above two displays, we have

$$
\begin{aligned}
0 &\leqslant 11/8\lambda\|\widehat{\mathbf v}_S\|_1-5/8\lambda\|\widehat{\mathbf v}_{S^c}\|_1+3/8\lambda\|\widehat{\mathbf v}\|_1\\
&= 11/8\lambda\|\widehat{\mathbf v}_S\|_1-5/8\lambda\|\widehat{\mathbf v}_{S^c}\|_1+3/8\lambda\|\widehat{\mathbf v}_S\|_1
\end{aligned}
$$

$$
\begin{aligned}
& +3/8\lambda\|\widehat{\mathbf{v}}_{S^c}\|_1 \\
= \quad & 7/4\lambda\|\widehat{\mathbf{v}}_S\|_1 - 1/4\lambda\|\widehat{\mathbf{v}}_{S^c}\|_1.
\end{aligned}
$$

Hence $\|\widehat{\mathbf{v}}_{S^c}\|_1 \leqslant 7\|\widehat{\mathbf{v}}_S\|_1$ and

$$
\begin{aligned}
\|\widehat{\mathbf{v}}\|_1 &= \|\widehat{\mathbf{v}}_S\|_1 + \|\widehat{\mathbf{v}}_{S^c}\|_1 \leqslant 8\|\widehat{\mathbf{v}}_S\|_1 \\
&\leqslant 8\sqrt{k}\|\widehat{\mathbf{v}}_S\|_2 \leqslant 8\sqrt{k}\|\widehat{\mathbf{v}}\|_2. \qquad \text{(S.8)}
\end{aligned}
$$

By Lemma 4, and recall that $c = 128$, we have

$$
\begin{aligned}
& n^{-1}\widehat{\mathbf{v}}^{\mathrm{T}}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}^*/2) \qquad \text{(S.9)} \\
& \times\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta}^*)^{\otimes 2} - \boldsymbol{\Omega}\}\widehat{\mathbf{v}} \\
\geqslant \quad & \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 - 1/(2c)\}\|\widehat{\mathbf{v}}\|_2^2 \\
& -\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]/(2cs)\|\widehat{\mathbf{v}}\|_1^2 \\
\geqslant \quad & (\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 - 1/(2c)\} \\
& -\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]/4)\|\widehat{\mathbf{v}}\|_2^2 \\
= \quad & 191/256\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\|\widehat{\mathbf{v}}\|_2^2. \qquad \text{(S.10)}
\end{aligned}
$$

Combining with the upper bound (S.6) we have

$$
\begin{aligned}
& 191/256\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\|\widehat{\mathbf{v}}\|_2^2 \\
\leqslant \quad & 2(\|\widehat{\mathbf{v}}_S\|_1 + \|\widehat{\mathbf{v}}_{S^c}\|_1)\phi\{\log(p)/n\}^{1/4} + 2\lambda\|\widehat{\mathbf{v}}_S\|_1 - 2\lambda\|\widehat{\mathbf{v}}_{S^c}\|_1 \\
\leqslant \quad & 2\|\widehat{\mathbf{v}}\|_1\phi\{\log(p)/n\}^{1/4} + 2\lambda\|\widehat{\mathbf{v}}\|_1 \\
\leqslant \quad & 4\max\{\phi\{\log(p)/n\}^{1/4}, \lambda\}\|\widehat{\mathbf{v}}\|_1
\end{aligned}
$$

$$\leqslant \quad 32\sqrt{k}\max\{\phi\{\log(p)/n\}^{1/4}, \lambda\}\|\widehat{\mathbf{v}}\|_2$$

$$= \quad 32\sqrt{k}\lambda\|\widehat{\mathbf{v}}\|_2.$$

Hence

$$\|\widehat{\mathbf{v}}\|_2 \leqslant \frac{2^{13}}{191}\frac{\sqrt{k}\lambda}{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}$$

and combine with (S.8)

$$\|\widehat{\mathbf{v}}\|_1 \leqslant \frac{2^{16}}{191}\frac{k\lambda}{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}.$$

This proves the results. □

Proof of Theorem 2: The conclusion is the same as those in Theorem 2 and (31) in Agarwal et al. (2012), where their optimization problem is

$$\begin{aligned}\widehat{\boldsymbol{\beta}} \quad &= \quad \mathrm{argmin}_{\|\boldsymbol{\beta}\|_1\leqslant b_0\sqrt{k}}\left[-n^{-1}\sum_{i=1}^{n}\{Y_i\mathbf{W}_i^{\mathrm{T}}\boldsymbol{\beta}\right.\\ &\qquad \left. -\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\} + \lambda\|\boldsymbol{\beta}\|_1\right].\end{aligned}$$

And their τ_l, τ_u are $\tau(n,p)$, γ_l, γ_u are $2a_1, 2a_2$, $\bar{\rho}$ is $b_0\sqrt{k}$, and $\mathcal{R}(\Pi_{\mathcal{M}^{\perp}}(\boldsymbol{\theta}^*))$ is $\|\boldsymbol{\beta}_{0S^c}\|_1 = 0$ in this theorem.

Carefully examining the proof of Theorem 2 in Agarwal et al. (2012) reveals that the proof holds when the lower-RE and upper-RE hold for the second derivative of $\mathcal{L}_1(\boldsymbol{\beta})$ at $\boldsymbol{\beta}$ in the feasible set, $\lambda \geqslant 2\|\partial\mathcal{L}_1(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0\|_\infty$ and $\mathcal{L}_1(\boldsymbol{\beta})$ is convex in the feasible set of $\boldsymbol{\beta}$.

In Lemma 4, we have already shown that the second derivative of $\mathcal{L}_1(\boldsymbol{\beta})$ at $\boldsymbol{\beta}$ in the feasible set satisfies the lower- and upper-RE conditions. In addition, we have shown in (S.9) that the second derivative of $\mathcal{L}_1(\boldsymbol{\beta})$ at $\boldsymbol{\beta}$ in the feasible set is positive definite under the conditions in the theorem statement. Further because

$$\begin{aligned}\lambda &\geqslant 8/3\phi\{\log(p)/n\}^{1/4} \\ &\geqslant 2\phi\{\log(p)/n\}^{1/4} \geqslant 2\|\partial\mathcal{L}_1(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0\|_\infty,\end{aligned}$$

where the last inequality holds by (S.2), so λ satisfies $\lambda \geqslant 2\|\partial\mathcal{L}_1(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0\|_\infty$ in Theorem 2 in Agarwal et al. (2012). Hence, the $\mathcal{L}_1(\boldsymbol{\beta})$ is convex on the feasible set and $\lambda \geqslant 2\|\partial\mathcal{L}_1(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0\|_\infty$ are satisfied simultaneously. Therefore, the result follows by using the same argument as those lead to Theorem 2 in Agarwal et al. (2012). □

Proof of Theorem 3: We will show that the theorem holds when the assumptions in Theorem 2 are satisfied, hence we start with verifying the assumptions in Theorem 2. The same argument as in Theorem 2 leads to that $\lambda \geqslant 8/3\|\partial\mathcal{L}_1(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}_0\|_\infty$, and that for any $\boldsymbol{\beta}$ in the feasible set, $\partial^2\mathcal{L}_1(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}$ satisfies the lower-RE and upper-RE conditions with parameters $\{a_1, \tau(n,p)\}$ and $\{a_2, \tau(n,p)\}$ as specified. We now verify the remaining assumptions in Theorem 2.

First, by the assumption that $k = o[\{n/\log(p)\}^{1/2}]$ and the fact that

$\tau(n,p) = O\{\sqrt{\log(p)/n}\}$, we have

$$\bar{\gamma}_l = O(1),$$

and

$$\frac{64\psi^2(\mathcal{M})\tau(n,p)}{\bar{\gamma}_l} = O\{k\sqrt{\log(p)/n}\} = o(1).$$

When $n, p \to \infty$, this leads to $\xi(\mathcal{M}) \to 1$. Further because $\tau(n,p)\psi^2(\mathcal{M}) = O[\{\log(p)/n\}^{1/2}k] = o(1)$,

$$\bar{\gamma}_l = 2\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 - 1/(2c)\} + o(1).$$

Taking into account that

$$\begin{aligned} & \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 - 1/(2c)\} \\ < \ & \alpha_{\max}[E\{\exp(\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1 + 1/(2c)\}, \end{aligned}$$

we have $d_1 < \kappa(\mathcal{M}) < 1 - d_1$ for some small positive constant d_1. Thus the assumption $\kappa(\mathcal{M}) \in [0, 1)$ in Theorem 2 holds. Further, we can easily check that

$$\beta(\mathcal{M}) = O\{\sqrt{\log(p)/n}\},$$

hence

$$\frac{32b_0\sqrt{k}}{1 - \kappa(\mathcal{M})}\xi(\mathcal{M})\beta(\mathcal{M}) = o[\{\log(p)/n\}^{1/4}].$$

Now $\tau(n,p)/\lambda^2 = O(1)$, by Theorem 2, we have

$$\begin{aligned}\|\boldsymbol{\beta}^t - \widehat{\boldsymbol{\beta}}\|_2^2 &\leqslant \frac{2\delta^2}{\bar{\gamma}_l} + \frac{16\delta^2\tau(n,p)}{\bar{\gamma}_l\lambda^2} + \frac{4\tau(n,p)\{6\psi(\mathcal{M})\}^2}{\bar{\gamma}_l} \\ &= O_p(\delta^2) + o(1) \\ &= o(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2^2) + o(1).\end{aligned}$$

The second last equality holds because $4\tau(n,p)\{6\psi(\mathcal{M})\}^2/\bar{\gamma}_l = 4\tau(n,p)36k/\bar{\gamma}_l = o(1)$, and the last equality holds because we selected $\delta^2 = \epsilon^2(\mathcal{M})/\{1 - \kappa(\mathcal{M})\} = o(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2^2)$. □

S.4 Definition of sub-Gaussian and sub-Exponential random variables

Proof of Lemma 1: 1. $\Longrightarrow$ 2. Assume property 1 holds. Recall that for every non-negative random variable Z, we have

$$E(Z|\mathcal{F}) = \int_0^\infty \Pr(Z \geqslant u|\mathcal{F})du$$

Let $Z = |X|^k$ and change of variable $u = t^k$, we obtain

$$\begin{aligned}E(|X|^k|\mathcal{F}) &= \int_0^\infty \Pr(|X| > t|\mathcal{F})kt^{k-1}dt \\ &\leqslant \int_0^\infty e^{1-\{t/K_1(\mathcal{F})\}^2}kt^{k-1}dt \\ &= ek/2K_1(\mathcal{F})^k\Gamma(k/2) \\ &\leqslant 2e(k/2)^{k/2}K_1(\mathcal{F})^k\end{aligned}$$

$$\leqslant \quad (2e)^k (k/2)^{k/2} K_1(\mathcal{F})^k$$

Taking the kth root yields property 2 with $K_2(\mathcal{F}) = \sqrt{2} e K_1(\mathcal{F})$.

2. $\Longrightarrow$ 3. Assume property 2 holds. Let $K_3(\mathcal{F}) = \sqrt{2/(e-1)} e K_2(\mathcal{F})$. Writing the Taylor series of the exponential function, we obtain

$$E[\exp\{X^2/K_3^2(\mathcal{F})\}|\mathcal{F}] = 1 + \sum_{k=1}^{\infty} \frac{K_3(\mathcal{F})^{-2k} E(X^{2k}|\mathcal{F})}{k!}$$
$$\leqslant \quad 1 + \sum_{k=1}^{\infty} \frac{(e-1)^k/2^k e^{-2k} K_2^{-2k}(\mathcal{F}) E(X^{2k}|\mathcal{F})}{k!}$$
$$\leqslant \quad 1 + \sum_{k=1}^{\infty} \frac{(e-1)^k/2^k e^{-2k} K_2^{-2k}(\mathcal{F}) K_2^{2k}(\mathcal{F})(2k)^k}{k!}$$
$$\leqslant \quad 1 + \sum_{k=1}^{\infty} \frac{(e-1)^k e^{-2k} k^k}{(k/e)^k} = e.$$

The last inequality holds because $k! \geqslant (k/e)^k$.

3. $\Longrightarrow$ 1. Assume property 3 holds. Exponentiating and using Markov's inequality and then the property 3, we have

$$\Pr(|X| > t|\mathcal{F}) = \Pr[\exp\{X^2/K_3^2(\mathcal{F})\} \geqslant \exp\{t^2/K_3^2(\mathcal{F})\}|\mathcal{F}]$$
$$\leqslant \quad e^{-t^2/K_3^2(\mathcal{F})} E[\exp\{X^2/K_3^2(\mathcal{F})\}|\mathcal{F}] \leqslant e^{1-\{t/K_3(\mathcal{F})\}^2}.$$

Hence property 1 holds with $K_1(\mathcal{F}) = K_3(\mathcal{F})$.

2. $\Longrightarrow$ 4. Assume that $E(X|\mathcal{F}) = 0$ and property 2 holds. We will prove that property 4 holds with an appropriately large absolute constant C such that $K_4(\mathcal{F}) = CK_2(\mathcal{F})$. This will follow by estimating Taylor series for the

exponential function

$$
\begin{aligned}
& E\{\exp(tX)|\mathcal{F}\} \\
= \; & 1 + tE(X|\mathcal{F}) + \sum_{k=2}^{\infty} \frac{t^k E(X^k|\mathcal{F})}{k!} \\
\leqslant \; & 1 + \sum_{k=2}^{\infty} \frac{|t|^k k^{k/2} K_2^k(\mathcal{F})}{k!} \\
\leqslant \; & 1 + \sum_{k=2}^{\infty} \left\{ \frac{e|t|}{\sqrt{k}} K_2(\mathcal{F}) \right\}^k \\
= \; & 1 + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{2k}} K_2(\mathcal{F}) \right\}^{2k} \\
& + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{2k+1}} K_2(\mathcal{F}) \right\}^{2k+1} \\
\leqslant \; & 1 + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{2k}} K_2(\mathcal{F}) \right\}^{2k} \\
& + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{2k+1}} K_2(\mathcal{F}) \right\}^{2k+2} \\
& + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{2k+1}} K_2(\mathcal{F}) \right\}^{2k} \\
\leqslant \; & 1 + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{k}} K_2(\mathcal{F}) \right\}^{2k} \\
& + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{k+1}} K_2(\mathcal{F}) \right\}^{2k+2} + \sum_{k=1}^{\infty} \left\{ \frac{e|t|}{\sqrt{k}} K_2(\mathcal{F}) \right\}^{2k} \\
\leqslant \; & 1 + \sum_{k=1}^{\infty} 3 \left\{ \frac{e|t|}{\sqrt{k}} K_2(\mathcal{F}) \right\}^{2k} \\
\leqslant \; & 1 + \sum_{k=1}^{\infty} \left\{ \frac{3e|t|}{\sqrt{k}} K_2(\mathcal{F}) \right\}^{2k} \\
\leqslant \; & 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \left\{ 3e|t| K_2(\mathcal{F}) \right\}^{2k}
\end{aligned}
$$

$$= \exp\{t^2(3e)^2K_2(\mathcal{F})^2\}.$$

Thus, the property 4 holds with $K_4(\mathcal{F}) = 3eK_2(\mathcal{F})$. In the above derivation, the first inequality holds follows from $E(X|\mathcal{F}) = 0$ and property 2, the second one holds because $k! > (k/e)^k$.

4. $\Longrightarrow$ 1. Assume property 4 holds. Then for $\lambda > 0$, by the exponential Markov inequality, and using the bound on the moment generating function given in property 4, we obtain

$$\begin{aligned}
\Pr(X \geqslant t|\mathcal{F}) &= \Pr\{\exp(\lambda X) \\
&\geqslant \exp(\lambda t)|\mathcal{F}\} \\
&\leqslant \exp(-\lambda t)E\{\exp(\lambda X)|\mathcal{F}\} \\
&\leqslant \exp\{-\lambda t + \lambda^2 K_4^2(\mathcal{F})\}.
\end{aligned}$$

Choose $\lambda = t/\{2K_4^2(\mathcal{F})\}$, we conclude that $\Pr(X \geqslant t|\mathcal{F}) \leqslant \exp[-t^2/\{4K_4^2(\mathcal{F})\}]$. Repeating the argument for $-X$, we also obtain $\Pr(X < -t|\mathcal{F}) \leqslant \exp[-t^2/\{4K_4^2(\mathcal{F})\}]$. Combining these two bounds we have

$$\Pr(|X| \geqslant t|\mathcal{F}) \leqslant 2\exp[-t^2/\{4K_4^2(\mathcal{F})\}] \leqslant \exp[1 - t^2/\{4K_4^2(\mathcal{F})\}].$$

Hence property 1 holds with $K_1(\mathcal{F}) = 2K_4(\mathcal{F})$. Thus, the lemma is proved.

□

Lemma S.1. *Let X be a centered conditional sub-Gaussian random variable* *with respect to $\mathcal{F}$. Then*

$$E\{\exp(\lambda X)|\mathcal{F}\} \leqslant \exp\{(3e)^2\lambda^2\|X\|^2_{\psi_2(\mathcal{F})}\}.$$

Proof: We first note that property 2 in Lemma 1 holds with $K_2(\mathcal{F}) = \|X\|_{\psi_2(\mathcal{F})}$. Following the proof of Lemma 1, this implies that property 4 in Lemma 1 also holds with $K_4(\mathcal{F}) = 3e\|X\|_{\psi_2(\mathcal{F})}$, which proves the result in Lemma S.1. □

Proof of Lemma 2: The proof follows the similar argument as that of Lemma 1.

1. $\Longrightarrow$ 2. Assume property 1 holds. Recall that for every non-negative random variable Z, we have

$$E(Z|\mathcal{F}) = \int_0^\infty \Pr(Z \geqslant u|\mathcal{F})du$$

Let $Z = |X|^k$ and change of variable $u = t^k$, we obtain

$$\begin{aligned} E(|X|^k|\mathcal{F}) &= \int_0^\infty \Pr(|X| > t|\mathcal{F})kt^{k-1}dt \\ &\leqslant \int_0^\infty e^{1-t/K_1(\mathcal{F})}kt^{k-1}dt \\ &= \Gamma(k+1)eK_1(\mathcal{F})^k \leqslant k^k eK_1(\mathcal{F})^k, \end{aligned}$$

where the inequality hold because $k! \leqslant k^k$. Taking the kth root yields property 2 with $K_2(\mathcal{F}) = eK_1(\mathcal{F})$.

2. $\Longrightarrow$ 3. Assume property 2 holds. Let $K_3(\mathcal{F}) = e^2/(e-1)K_2(\mathcal{F})$. Writing

the Taylor series of the exponential function, we obtain

$$\begin{aligned}
& E[\exp\{X/K_3(\mathcal{F})\}|\mathcal{F}] \\
= & 1 + \sum_{k=1}^{\infty} \frac{K_3(\mathcal{F})^{-k} E(|X|^k|\mathcal{F})}{k!} \\
\leqslant & 1 + \sum_{k=1}^{\infty} \frac{K_3(\mathcal{F})^{-k} K_2(\mathcal{F})^k k^k}{k!} \\
= & 1 + \sum_{k=1}^{\infty} \frac{(e-1)^k k^k / e^{2k}}{k!} \\
\leqslant & 1 + \sum_{k=1}^{\infty} (e-1)^k / e^k = e.
\end{aligned}$$

The last inequality holds because $k! \geqslant (k/e)^k$.

3. $\implies$ 1. Assume property 3 holds. Exponentiating and using Markov's inequality and then the property 3, we have

$$\begin{aligned}
\Pr(|X| > t|\mathcal{F}) & = \Pr[\exp\{|X|/K_3(\mathcal{F})\} \\
& \geqslant \exp\{t/K_3(\mathcal{F})\}|\mathcal{F}] \\
& \leqslant e^{-t/K_3(\mathcal{F})} E[\exp\{|X|/K_3(\mathcal{F})\}|\mathcal{F}] \\
& \leqslant e^{1-t/K_3(\mathcal{F})}.
\end{aligned}$$

Hence property 1 holds with $K_1(\mathcal{F}) = K_3(\mathcal{F})$. □

S.5 Properties of Conditional sub-Gaussian and sub-Exponential Random Variables

Lemma S.2. *Let X be a centered conditional sub-exponential random variables with respect to $\mathcal{F}$. Then for λ such that $0 < \lambda \leqslant 1/(2e\|X\|_{\psi_1(\mathcal{F})})$, we have*

$$E\{\exp(\lambda X)|\mathcal{F}\} \leqslant \exp(2e^2\lambda^2\|X\|^2_{\psi_1(\mathcal{F})}).$$

Proof: From $E(X|\mathcal{F}) = 0$ and property 2 in Lemma 2, using Taylor expansion, we get

$$\begin{aligned} E\{\exp(\lambda X)|\mathcal{F}\} &= 1 + \lambda E(X|\mathcal{F}) + \sum_{k=2}^{\infty} \frac{\lambda^k E(X^k|\mathcal{F})}{k!} \\ &\leqslant 1 + \sum_{k=2}^{\infty} \frac{\lambda^k E(|X|^k|\mathcal{F})}{k!} \\ &\leqslant 1 + \sum_{k=2}^{\infty} \frac{\lambda^k k^k \|X\|^k_{\psi_1(\mathcal{F})}}{k!} \\ &\leqslant 1 + \sum_{k=2}^{\infty} \left\{e\lambda\|X\|_{\psi_1(\mathcal{F})}\right\}^k. \end{aligned}$$

The first inequality holds because $E(X|\mathcal{F}) = 0$ and $X^k \leqslant |X|^k$; The second inequality follows property 2. The third inequality holds because $k! > (k/e)^k$. If $0 < \lambda \leqslant 1/(2e\|X\|_{\psi_1(\mathcal{F})})$, the right hand side of the above equation is bounded by

$$1 + 2e^2\lambda^2\|X\|^2_{\psi_1(\mathcal{F})} \leqslant \exp(2e^2\lambda^2\|X\|^2_{\psi_1(\mathcal{F})}).$$

This completes the proof. □

Lemma S.3. *Let $X_1, \ldots, X_n$ be independent centered sub-Gaussian random variables with respect to the sub-sigma fileds $\mathcal{F}_1, \ldots, \mathcal{F}_n$ respectively. For sequence $a_1(\mathcal{F}), \ldots, a_n(\mathcal{F})$,*

$$\begin{aligned} & E\left\{\exp\left(\lambda \sum_{i=1}^n a_i(\mathcal{F}) X_i\right) |\mathcal{F}_i, i = 1, \ldots, n\right\} \\ \leqslant \; & \exp\left((3e)^2 \lambda^2 \sum_{i=1}^n \|a_i(\mathcal{F}) X_i\|^2_{\psi_2(\mathcal{F}_i)}\right) \end{aligned}$$

Proof: When X_i is centered sub-Gaussian, then $a_i X_i$ is also centered and sub-Gaussian. Hence, from Lemma S.1, we have

$$\begin{aligned} & E\left\{\exp\left(\lambda \sum_{i=1}^n a_i(\mathcal{F}) X_i\right) |\mathcal{F}_i, i = 1, \ldots, n\right\} \\ = \; & \prod_{i=1}^n E\{\exp(\lambda a_i(\mathcal{F}) X_i) | \mathcal{F}_i\} \\ \leqslant \; & \exp\left((3e)^2 \lambda^2 \sum_{i=1}^n \|a_i(\mathcal{F}) X_i\|^2_{\psi_2(\mathcal{F}_i)}\right). \end{aligned}$$

□

Lemma S.4. *Let $X_1, \ldots, X_n$ be independent centered sub-exponential random variables with respect to the sub-sigma fileds $\mathcal{F}_1, \ldots, \mathcal{F}_n$ respectively. For any any sequence $a_1(\mathcal{F}), \ldots, a_n(\mathcal{F})$ and λ such that $0 < \lambda \leqslant \min_{i=1,\ldots,n}\{1/(2e\|a_i(\mathcal{F}_i) X_i\|_{\psi_1})\}$,*

$$\begin{aligned} & E\left\{\exp\left(\lambda \sum_{i=1}^n a_i(\mathcal{F}) X_i\right) |\mathcal{F}_i, i = 1, \ldots, n\right\} \\ \leqslant \; & \exp\left(2e^2 \lambda^2 \sum_{i=1}^n \|a_i(\mathcal{F}) X_i\|^2_{\psi_1(\mathcal{F}_i)}\right). \end{aligned}$$

Proof: When X_i is centered sub-exponential, then a_iX_i is also centered and sub-exponential. Hence, from Lemma S.2, we have

$$
\begin{aligned}
& E\left\{\exp\left(\lambda\sum_{i=1}^n a_i(\mathcal{F})X_i\right)|\mathcal{F}_i, i=1,\dots,n\right\} \\
= & \prod_{i=1}^n E\{\exp(\lambda a_i(\mathcal{F})X_i)|\mathcal{F}_i\} \\
\leqslant & \exp\left(2e^2\lambda^2\sum_{i=1}^n \|a_i(\mathcal{F})X_i\|^2_{\psi_1(\mathcal{F}_i)}\right).
\end{aligned}
$$

□

S.6 Properties under Regularity Conditions (C1) – (C6)

Lemma S.5. *For $r>0$, let $c_{10}\equiv \max[\sqrt{18e^2m_3^2/\{M_3Q_1^2(1+r)r\}},1]$. Assume Conditions (C1) – (C4) to hold. Then*

$$
\begin{aligned}
& \Pr\left[n^{-1}\sum_{i=1}^n \|Y_i-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i\|_\infty \right.\\
& \left. > \max(4e\sqrt{M_1}, 8eM_2C)\sqrt{log(p)}/\sqrt{n}\right] \leqslant 2p^{-1}, \\
& \Pr\left(n^{-1}\|\sum_{i=1}^n \left[\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})\right.\right. \\
& \left.-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\|_\infty\right] \\
> \quad & \left. 2c_{10}M_3Q_1(1+r)\sqrt{log(p)/n}/m_3\right) \leqslant 2p^{-1},
\end{aligned}
$$

and

$$\Pr\left[n^{-1}\|\sum_{i=1}^{n} Y_i(\mathbf{W}_i - \mathbf{X}_j)\|_\infty \right.$$
$$\left. > \sqrt{2}\sqrt{36e^2M_0}\sqrt{log(p)/n}\right] \leqslant 2p^{-1}.$$

Proof of Lemma S.5 Let $\mathbf{e}_i$ be the unit vector with the ith element 1 and $\mathcal{F}_x$ be the sigma field generated by $\mathbf{X}_i, i = 1, \ldots, n$. By Condition (C1), we can choose sufficiently large $K(\mathbf{X}_i)$, where $K(\mathbf{X}_i) > 1$, so that

$$\begin{aligned}
& E[\exp\{|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)|/K(\mathbf{X}_i)\}|\mathbf{X}_i] \\
\leqslant\ & E(\exp[\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}/K(\mathbf{X}_i)]|\mathbf{X}_i) \\
& +E(\exp[\{-Y_i + \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}/K(\mathbf{X}_i)]|\mathbf{X}_i) \\
\leqslant\ & E(\exp[\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}/K(\mathbf{X}_i)]|\mathbf{X}_i) \\
& + \exp[\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}/K(\mathbf{X}_i)] \\
=\ & \exp\left(\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)[\exp\{1/K(\mathbf{X}_i)\} - 1 - 1/K(\mathbf{X}_i)]\right) \\
& + \exp[\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}/K(\mathbf{X}_i)] \\
<\ & e/2 + e/2 = e.
\end{aligned}$$

Hence, $Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)$ and $-Y_i + \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)$ given $\mathbf{X}_i$ are conditional sub-exponential random variables following Definition 4. Let $0 < \lambda \leqslant \min_i 1/[2e\|\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j\|_{\psi_1(\mathcal{F}_x)}] = \min_i 1/\{2e|\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|K_Y(\mathbf{X}_i)\}$, we further have

$$\begin{aligned}
& \Pr\left[n^{-1}\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j > t|\mathcal{F}_x\right] \\
= \ & \Pr\left(\exp[\lambda\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j]\right. \\
& \left. > \exp(\lambda nt)|\mathcal{F}_x\right) \\
\leqslant \ & E\left(\exp[\lambda\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j]|\mathcal{F}_x\right) \\
& \times \exp(-\lambda nt) \\
\leqslant \ & \exp\left(2e^2\lambda^2\sum_{i=1}^{n}\|\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j\|^2_{\psi_1(\mathcal{F}_x)}\right. \\
& \left. -\lambda nt\right) \\
= \ & \exp\left(-\lambda nt + 2e^2\lambda^2\sum_{i=1}^{n}(\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j)^2\right. \\
& \left. [\sup_{k\geqslant 1} k^{-1}E\{|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)|^k|\mathcal{F}_x\}^{1/k}]^2\right) \\
= \ & \exp\left(-\lambda nt + 2e^2\lambda^2\sum_{i=1}^{n}|\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|^2 K_Y(\mathbf{X}_i)^2\right),
\end{aligned} \tag{S.11}$$

where the first inequality is due to the Markov inequality, the second inequality follows from Lemma S.4, and the last two equalities are due to the definitions of $\|\cdot\|_{\psi_1(\mathcal{F}_i)}$ and $K_Y(\mathbf{X}_i)$. Let

$$\lambda_1 = \frac{nt}{4e^2\sum_{i=1}^{n}|\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|^2 K_Y(\mathbf{X}_i)^2}, \quad \textit{and} \quad \lambda_2 = \frac{1}{2e\max_i|\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|K_Y(\mathbf{X}_i)}.$$

If $\lambda_1 < \lambda_2$, letting $\lambda = \lambda_1$ in (S.11), we get

$$\begin{aligned}
& \Pr\left\{n^{-1}\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j > t|\mathcal{F}_x\right\} \\
\leqslant\ & \exp\left[\left\{-\frac{n^2t^2}{8e^2\sum_{i=1}^{n}|\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|^2K_Y(\mathbf{X}_i)^2}\right\}\right] \\
\leqslant\ & \exp\left(-\frac{nt^2}{8e^2M_1}\right)
\end{aligned}$$

almost surely. If $\lambda_2 < \lambda_1$, letting $\lambda = \lambda_2$ in (S.11), we get

$$\begin{aligned}
& \Pr\left\{n^{-1}\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j > t|\mathcal{F}_x\right\} \\
\leqslant\ & \exp\{-\lambda_2 nt + 2e^2\lambda_2^2 nt/(4e^2\lambda_1)\} \\
\leqslant\ & \exp\{-\lambda_2 nt + 2e^2\lambda_2 nt/(4e^2)\} \\
=\ & \exp(-\lambda_2 nt/2) \\
=\ & \exp\left\{\frac{-nt}{4e\max_i|\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|K_Y(\mathbf{X}_i)}\right\} \\
\leqslant\ & \exp\left\{\frac{-nt}{4eM_2\log(n)}\right\}
\end{aligned}$$

almost surely. Thus, combining the above results, we get

$$\begin{aligned}
& \Pr\left\{n^{-1}\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j > t|\mathcal{F}_x\right\} \\
\leqslant\ & \exp\left\{-\min\left(\frac{nt^2}{8e^2M_1}, \frac{nt}{4eM_2\log(n)}\right)\right\} \qquad \text{(S.12)}
\end{aligned}$$

almost surely. Now taking expectations on both sides, we get

$$\Pr\left[n^{-1}\sum_{i=1}^{n}\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j > t\right] \leqslant \exp\left[-\min\left(\frac{nt^2}{8e^2M_1}, \frac{nt}{4eM_2\log(n)}\right)\right].$$

Note that the same derivation in (S.11) and below also applies to $-\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i$, hence we also have

$$\Pr\left[n^{-1}\sum_{i=1}^{n}\{-Y_i + \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j > t\right] \leqslant \exp\left[-\min\left(\frac{nt^2}{8e^2M_1}, \frac{nt}{4eM_2\log(n)}\right)\right].$$

Hence, we have

$$\Pr\left[n^{-1}\sum_{i=1}^{n}\|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j\|_\infty > t\right] \leqslant 2p\exp\left[-\min\left(\frac{nt^2}{8e^2M_1}, \frac{nt}{4eM_2\log(n)}\right)\right].$$

Inserting $t = c_{00}\sqrt{\log(p)}/\sqrt{n}$, we obtain

$$\begin{aligned} &\Pr\left[n^{-1}\sum_{i=1}^{n}\|\{Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i\|_\infty > c_{00}\sqrt{\log(p)}/\sqrt{n}\right] \\ \leqslant\ & 2p\exp\left[-\min\left(\frac{c_{00}^2\log(p)}{8e^2M_1}, \frac{nc_{00}\sqrt{\log(p)}/\sqrt{n}}{4eM_2\log(n)}\right)\right] \\ =\ & 2p\exp\left[-\min\left(\frac{c_{00}^2\log(p)}{8e^2M_1}, \frac{c_{00}\log(p)}{4eM_2\log(n)\{\sqrt{\log(p)}/\sqrt{n}\}}\right)\right] \\ \leqslant\ & 2p\exp\left[-\min\left(\frac{c_{00}^2\log(p)}{8e^2M_1}, \frac{c_{00}\log(p)}{4eM_2C}\right)\right]. \end{aligned}$$

The last equality holds by Condition (C4). Now let $c_{00} = \max(4e\sqrt{M_1}, 8eM_2C)$, we have

$$\Pr\left[n^{-1}\sum_{i=1}^{n}\|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i\|_\infty > \max(4e\sqrt{M_1}, 8eM_2C)\sqrt{\log(p)}/\sqrt{n}\right] \leqslant 2p^{-1}.$$

In addition, let $\mathcal{F}(\boldsymbol{\beta}_0)$ be the sigma field generated by $\mathbf{X}_i, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, i = 1, \ldots, n$, since $\mathbf{W}_i$ given $\mathbf{X}_i$ is normal, $\mathbf{W}_i$ given $\mathcal{F}(\boldsymbol{\beta}_0)$ is also normal hence is sub-gaussian. Recall that

$$K_{wij}(\boldsymbol{\beta}_0) = \sup_{k\geqslant 1} k^{-1/2}E[|(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j - E\{(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\}|^k|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i]^{1/k}.$$

Then letting

$$\lambda_j = \frac{nt}{18e^2 \sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2 |\exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)|^2}, \tag{S.13}$$

we have

$$\begin{aligned}
& \Pr\left[n^{-1} \sum_{i=1}^n \left\{ \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j - \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{X}_i) \mathbf{X}_i^{\mathrm{T}} \mathbf{e}_j \right\} > t | \mathcal{F}(\boldsymbol{\beta}_0) \right] \\
= \; & \Pr\left(\exp\left[\lambda_j \sum_{i=1}^n \left\{ \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j - \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{X}_i) \mathbf{X}_i^{\mathrm{T}} \mathbf{e}_j \right\} \right] \right. \\
& \left. > \exp(\lambda_j n t) | \mathcal{F}(\boldsymbol{\beta}_0) \right) \\
\leqslant \; & \exp(-\lambda_j n t) E \left[\exp\left\{ \lambda_j \sum_{i=1}^n \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j \right. \right. \\
& \left. \left. - \lambda_j \sum_{i=1}^n \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{X}_i) \mathbf{X}_i^{\mathrm{T}} \mathbf{e}_j \right\} | \mathcal{F}(\boldsymbol{\beta}_0) \right] \\
= \; & \exp(-\lambda_j n t) E \left(\exp\left[\lambda_j \sum_{i=1}^n \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j \right. \right. \\
& \left. \left. - \lambda_j E\{ \sum_{i=1}^n \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j | \boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i, \mathbf{X}_i \} \right] | \mathcal{F}(\boldsymbol{\beta}_0) \right) \\
& \times \exp\left(\lambda_j \sum_{i=1}^n [E\{ \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j | \boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i, \mathbf{X}_i \} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{X}_i) \mathbf{X}_i^{\mathrm{T}} \mathbf{e}_j] \right) \\
\leqslant \; & \exp(-\lambda_j n t) \exp\left[(3e)^2 \lambda_j^2 \sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2 \{ \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2) \}^2 \right] \\
& \times \exp\left(\lambda_j \sum_{i=1}^n [E\{ \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega})^{\mathrm{T}} \mathbf{e}_j | \boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i, \mathbf{X}_i \} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{X}_i) \mathbf{X}_i^{\mathrm{T}} \mathbf{e}_j] \right) \\
= \; & \exp\left\{ \frac{-n^2 t^2}{18e^2 \sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2 |\exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)|^2} \right\} \\
& \times \exp\left\{ \frac{n^2 t^2}{36e^2 \sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2 |\exp(\boldsymbol{\beta}_0^{\mathrm{T}} \mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{\Omega} \boldsymbol{\beta}_0/2)|^2} \right\}
\end{aligned}$$

$$\times \exp\left[\frac{nt}{18e^2\sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2|\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)|^2}\right.$$
$$\left.\times \sum_{i=1}^n [E\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j]\right]$$
$$\leqslant \exp\left\{\frac{-n^2t^2}{36e^2\sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2|\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)|^2}\right\}$$
$$\times \exp\left[\frac{nt}{18e^2\sum_{i=1}^n K_{wij}(\boldsymbol{\beta}_0)^2|\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)|^2}\right.$$
$$\left.\times |\sum_{i=1}^n E\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|\right].$$

The first inequality holds by the Markov inequality, and the second inequality holds by Lemma S.3. Letting $t = 2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3$ for some constants $r > 0, c_{10} > 1$, where m_3, M_3, Q_1 are defined in Condition (C5), we get

$$\Pr\left[n^{-1}\sum_{i=1}^n \left\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j\right\}\right.$$
$$\left. > 2c_{10}Q_1M_3(1+r)\sqrt{\log(p)/n}/m_3|\mathcal{F}(\boldsymbol{\beta}_0)\right]$$
$$\leqslant \exp\left[-\left(\frac{nt^2}{36e^2M_3}\right)\right]\exp\left[\frac{t}{18e^2m_3}|\sum_{i=1}^n E\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\}\right.$$
$$\left. - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|\right]$$
$$= \exp\left[-\left(\frac{c_{10}^2Q_1^2(1+r)^2\log(p)M_3}{9e^2m_3^2}\right)\right]\exp\left[\frac{c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}}{9e^2m_3^2}\right.$$
$$\left.|\sum_{i=1}^n E\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})^{\mathrm{T}}\mathbf{e}_j|\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i, \mathbf{X}_i\} - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\mathbf{e}_j|\right]$$
$$\leqslant \exp\left[-\left\{c_{10}^2Q_1^2(1+r)^2\log(p)M_3/(9e^2m_3^2)\right\}\right]\exp\left\{c_{10}Q_1^2(1+r)\log(p)M_3/(9e^2m_3^2)\right\}$$

$$\begin{aligned}
&\leqslant \exp\left[-\left\{c_{10}^2Q_1^2(1+r)^2\log(p)M_3/(9e^2m_3^2)\right\}\right]\exp\left\{c_{10}^2Q_1^2(1+r)\log(p)M_3/(9e^2m_3^2)\right\}\\
&= \exp\left[-\left\{c_{10}^2Q_1^2(1+r)r\log(p)M_3/(9e^2m_3^2)\right\}\right]
\end{aligned}$$

almost surely, where the second inequality holds by Condition (C5).

Taking expectation on both sides of the above inequality, we have

$$\begin{aligned}
&\Pr\Bigg[n^{-1}\sum_{i=1}^n\left\{\exp(\boldsymbol\beta_0^{\rm T}\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega\boldsymbol\beta_0/2)(\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega)^{\rm T}\mathbf{e}_j-\exp(\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)\mathbf{X}_i^{\rm T}\mathbf{e}_j\right\}\\
&>2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3\Bigg]\\
\leqslant\ & \exp\left[-\left\{c_{10}^2Q_1^2(1+r)r\log(p)M_3/(9e^2m_3^2)\right\}\right].
\end{aligned}$$

We can easily check that the same derivation after (S.13) also applies to

$$-\exp(\boldsymbol\beta_0^{\rm T}\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega\boldsymbol\beta_0/2)(\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega)^{\rm T}\mathbf{e}_j+\exp(\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)\mathbf{X}_i^{\rm T}\mathbf{e}_j$$

and will lead to

$$\begin{aligned}
&\Pr\Bigg[n^{-1}\sum_{i=1}^n-\left\{\exp(\boldsymbol\beta_0^{\rm T}\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega\boldsymbol\beta_0/2)(\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega)^{\rm T}\mathbf{e}_j-\exp(\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)\mathbf{X}_i^{\rm T}\mathbf{e}_j\right\}\\
&>2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3\Bigg]\\
\leqslant\ & \exp\left[-\left\{c_{10}^2Q_1^2(1+r)r\log(p)M_3/(9e^2m_3^2)\right\}\right].
\end{aligned}$$

Thus, letting $c_{10}=\max[\sqrt{2}\sqrt{9e^2m_3^2/\{M_3Q_1^2(1+r)r}\},1]$, we have

$$\begin{aligned}
&\Pr\Bigg(n^{-1}\|\sum_{i=1}^n\left[\exp(\boldsymbol\beta_0^{\rm T}\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega\boldsymbol\beta_0/2)(\mathbf{W}_i-\boldsymbol\beta_0^{\rm T}\boldsymbol\Omega)-\exp(\boldsymbol\beta_0^{\rm T}\mathbf{X}_i)\mathbf{X}_i\|_\infty\right]\\
&>2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3\Bigg)\\
\leqslant\ & 2p\exp\left[-\left\{(c_{10}Q_1)^2(1+r)r\log(p)M_3/(9e^2m_3^2)\right\}\right]
\end{aligned}$$

$$\leqslant \quad 2p^{-1}.$$

Let $\mathcal{F}_Y$ be the sigma field generated by $Y_i, i = 1, \ldots, n$. Because $\mathbf{U}_i = \mathbf{W}_i - \mathbf{X}_i$ is normal and independent with Y_i, using the same argument, we have

$$\begin{aligned}
& \Pr\left[n^{-1}\sum_{i=1}^{n} Y_i(W_{ij} - X_{ij}) > t|\mathcal{F}_Y\right] \\
= \;& \Pr\left(\exp[\lambda\sum_{i=1}^{n} Y_i(W_{ij} - X_{ij})] > \exp(\lambda n t)\right) \\
\leqslant \;& E\left(\exp[\lambda\sum_{i=1}^{n} Y_i(W_{ij} - X_{ij})]\right)\exp(-\lambda n t) \\
\leqslant \;& \exp\left(-\lambda n t + (3e)^2\lambda^2\sum_{i=1}^{n} \|\{Y_i(W_{ij} - X_{ij})\}\|^2_{\psi_2(\mathcal{F}_{Y_i})}\right) \\
= \;& \exp\left(-\lambda n t + (3e)^2\lambda^2\sum_{i=1}^{n} Y_i^2[\sup_{k\geqslant 1} k^{-1/2}E\{|(W_{ij} - X_{ij})|^k\}^{1/k}]^2\right) \\
= \;& \exp\left(-\lambda n t + (3e)^2\lambda^2\sum_{i=1}^{n} Y_i^2 K_j^2\right).
\end{aligned}$$

The third inequality holds by Lemma S.3. Letting

$$\lambda = \frac{nt}{18e^2\sum_{i=1}^{n} Y_i^2 K_j^2}$$

we obtain

$$\begin{aligned}
& \Pr\left[n^{-1}\sum_{i=1}^{n} Y_i(W_{ij} - X_{ij}) > t|\mathcal{F}_Y\right] \\
\leqslant \;& \exp\left(-\frac{n^2t^2}{36e^2\sum_{i=1}^{n} Y_i^2 K_j^2}\right) \\
\leqslant \;& \exp\left(-\frac{nt^2}{36e^2 M_0}\right).
\end{aligned}$$

Take the expectation on both side, we have

$$\begin{aligned}
& \Pr\left[n^{-1}\sum_{i=1}^n Y_i(W_{ij}-X_{ij})>t\right] \\
\leqslant\ & \exp\left(-\frac{n^2t^2}{36e^2\sum_{i=1}^n Y_i^2K_j^2}\right) \\
\leqslant\ & \exp\left(-\frac{nt^2}{36e^2M_0}\right).
\end{aligned}$$

Using the same derivation on $-n^{-1}\sum_{i=1}^n Y_i(W_{ij}-X_{ij})$, we can also obtain

$$\Pr\left[n^{-1}\sum_{i=1}^n -Y_i(W_{ij}-X_{ij})>t\right]\leqslant \exp\left(-\frac{nt^2}{36e^2M_0}\right).$$

Hence, selecting $t=\sqrt{2}\sqrt{36e^2M_0}\sqrt{\log(p)/n}$ leads to

$$\begin{aligned}
& \Pr\left[n^{-1}\|\sum_{i=1}^n Y_i(\mathbf{W}_i-\mathbf{X}_j)\|_\infty>\sqrt{2}\sqrt{36e^2M_0}\sqrt{\log(p)/n}\right] \\
\leqslant\ & 2p\exp\left(-\frac{nt^2}{36e^2M_0}\right) \\
=\ & 2p^{-1}.
\end{aligned}$$

Proof of Lemma 3: By the triangle inequality we have

$$\begin{aligned}
& n^{-1}\|\sum_{i=1}^n Y_i\mathbf{W}_i-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})\|_\infty \\
\leqslant\ & n^{-1}\|\sum_{i=1}^n\{Y_i-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i\|_\infty \\
& +n^{-1}\|\sum_{i=1}^n\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i-\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})-\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\|_\infty \\
& +n^{-1}\|\sum_{i=1}^n Y_i(\mathbf{W}_i-\mathbf{X}_i)\|_\infty,
\end{aligned}$$

hence by Lemma S.5

$$
\begin{aligned}
& \Pr\left\{n^{-1}\|\sum_{i=1}^{n} Y_i\mathbf{W}_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega})\|_\infty\right.\\
& > 3\max\{\max(4e\sqrt{M_1}, 8eM_2C)\sqrt{\log(p)}/\sqrt{n},\\
& \left. 2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3, \sqrt{2}\sqrt{36e^2M_0}\sqrt{\log(p)/n}\}\right\}\\
\leqslant\ & \Pr\left[3\max\left\{n^{-1}\sum_{i=1}^{n}\|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i\|_\infty,\right.\right.\\
& n^{-1}\sum_{i=1}^{n}\|\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}) - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\|_\infty,\\
& \left. n^{-1}\|\sum_{i=1}^{n} Y_i(\mathbf{W}_i - \mathbf{X}_j)\|_\infty\right\} > 3\max\{\max(4e\sqrt{M_1}, 8eM_2C)\sqrt{\log(p)}/\sqrt{n},\\
& \left. 2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3, \sqrt{2}\sqrt{36e^2M_0}\sqrt{\log(p)/n}\right]\\
\leqslant\ & \Pr\left[n^{-1}\sum_{i=1}^{n}\|Y_i - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\}\mathbf{X}_i\|_\infty > \max(4e\sqrt{M_1}, 8eM_2C)\sqrt{\log(p)}/\sqrt{n}\right]\\
& + \Pr\left(n^{-1}\|\sum_{i=1}^{n}\left[\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}_0/2)(\mathbf{W}_i - \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\Omega}) - \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\|_\infty\right]\right.\\
& \left. > 2c_{10}M_3Q_1(1+r)\sqrt{\log(p)/n}/m_3\right)\\
& + \Pr\left[n^{-1}\|\sum_{i=1}^{n} Y_i(\mathbf{W}_i - \mathbf{X}_j)\|_\infty > \sqrt{2}\sqrt{36e^2M_0}\sqrt{\log(p)/n}\right]\\
\leqslant\ & 6p^{-1},
\end{aligned}
$$

where the last inequality is due to Lemma S.5. This proves the results. □

Lemma S.6. *Assume that Conditions (C1) and (C6) hold, and the variables $\mathbf{U}_i, \mathbf{X}_i$ have finite dimension p_1. Let $\mathbf{v}$ be a p_1-dimensional vector.*

For sufficiently large n, we have

$$\begin{aligned}
& \Pr\left(|\sum_{i=1}^{n} A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}| > nt\right) \\
\leqslant\ & 2\exp\left(-\min\left[\frac{nt^2}{16e^2M_4}, \frac{nt}{8eM_5 log(n)}\right]\right).
\end{aligned}$$

Proof: By Lemma 1 statement 3 and Lemma 2 statement 3, we can see that the square of a conditional sub-Gaussian variable is sub-exponential. Now because $\mathbf{v}^{\mathrm{T}}(\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega})$ given $\mathbf{X}_i$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i$ is normal, and recall that

$$g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) \equiv \mathbf{v}^{\mathrm{T}}\{(\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega})^{\otimes 2} - \boldsymbol{\Omega}\}\mathbf{v},$$

we have that

$$g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - E\{g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\}$$

is centered sub-exponential. Recall also that

$$A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i) \equiv \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2),$$

we have

$$\begin{aligned}
& \Pr\left(\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}] > t|\mathcal{F}(\boldsymbol{\beta})\right) \\
=\ & \Pr\left\{\exp\left(\lambda\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]\right) > \exp(\lambda t)|\mathcal{F}(\boldsymbol{\beta})\right\} \\
\leqslant\ & \exp(-\lambda t)E\left\{\exp\left(\lambda\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]\right)|\mathcal{F}(\boldsymbol{\beta})\right\} \\
=\ & \exp(-\lambda t)E\left\{\exp\left(\lambda\sum_{i=1}^{n}A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)[g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - E\{g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\}]\right)|\mathcal{F}(\boldsymbol{\beta})\right\}
\end{aligned}$$

$$
\begin{aligned}
& \times \exp\left(\lambda \sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E\{g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\} - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]\right) \\
\leqslant \; & \exp(-\lambda t)\exp\left\{2e^2\lambda^2\sum_{i=1}^{n}A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)K_{gvi}(\boldsymbol{\beta})^2\right\} \\
& \times \exp\left(\lambda \sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E\{g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\} - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]\right).
\end{aligned}
$$

The second inequality above holds by Lemma S.4. Further, let

$$
\lambda_1 = \frac{t}{4e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)}, \qquad \textit{and} \qquad \lambda_2 = \frac{1}{2e\max_i K_{gvi}(\boldsymbol{\beta})|A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|}.
$$

If $\lambda_1 < \lambda_2$, letting $\lambda = \lambda_1$, we get

$$
\exp(-\lambda t)\exp\left\{2e^2\lambda^2\sum_{i=1}^{n}A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)K_{gvi}(\boldsymbol{\beta})^2\right\} = \exp\left[\left\{-\frac{t^2}{8e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)}\right\}\right].
$$

If $\lambda_2 < \lambda_1$, letting $\lambda = \lambda_2$, we get

$$
\begin{aligned}
& \exp(-\lambda t)\exp\left\{2e^2\lambda^2\sum_{i=1}^{n}A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)K_{gvi}(\boldsymbol{\beta})^2\right\} \\
= \; & \exp\{-\lambda_2 t + 2e^2\lambda_2^2 t/(4e^2\lambda_1)\} \\
\leqslant \; & \exp\{-\lambda_2 t + 2e^2\lambda_2 t/(4e^2)\} \\
= \; & \exp(-\lambda_2 t/2) \\
= \; & \exp\left\{\frac{-t}{4e\max_i K_{gvi}(\boldsymbol{\beta})|A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|}\right\}.
\end{aligned}
$$

Combine the above result and let

$$
\lambda = \min\left(\frac{t}{4e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)}, \frac{1}{2e\max_i K_{gvi}(\boldsymbol{\beta})|A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|}\right)
$$

we have

$$
\begin{aligned}
&\Pr\left(\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})-\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]>t|\mathcal{F}(\boldsymbol{\beta})\right)\\
\leqslant\ & \exp\left(-\min\left[\frac{t^2}{8e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)},\frac{t}{4e\max_i K_{gvi}(\boldsymbol{\beta})|A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|}\right]\right)\\
&\times\exp\left[\sum_{i=1}^{n}\min\left\{\frac{t}{4e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)},\frac{1}{2e\max_i K_{gvi}(\boldsymbol{\beta})A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)}\right\}\right.\\
&\left.\times[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E\{g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\}-\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]\right].
\end{aligned}
$$

Replacing t with nt, for sufficiently large n and fixed t, we have

$$
\begin{aligned}
&\Pr\left(\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})-\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]>nt|\mathcal{F}(\boldsymbol{\beta})\right)\\
\leqslant\ & \exp\left(-\min\left[\frac{n^2t^2}{8e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)},\frac{nt}{4e\max_i K_{gvi}(\boldsymbol{\beta})|A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)|}\right]\right)\\
&\times\exp\left[\sum_{i=1}^{n}\min\left\{\frac{nt}{4e^2\sum_{i=1}^{n}K_{gvi}(\boldsymbol{\beta})^2A^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)},\frac{1}{2e\max_i K_{gvi}(\boldsymbol{\beta})A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)}\right\}\right.\\
&\left.\times[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)E\{g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i,\mathbf{X}_i\}-\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]\right]\\
\leqslant\ & \exp\left(-\min\left[\frac{nt^2}{8e^2M_4},\frac{nt}{4eM_5\log(n)}\right]\right)\\
&\times\exp\left[\min\left\{\frac{t}{4e^2m_4},\frac{1}{2em_5\log(n)}\right\}Q_2\sqrt{n}\right]\\
\leqslant\ & \exp\left(-\min\left[\frac{nt^2}{16e^2M_4},\frac{nt}{8eM_5\log(n)}\right]\right),
\end{aligned}
$$

in probability. The second inequality holds by (S.3).

Taking expectation on both sides of the above display, we have

$$
\Pr\left(\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v})-\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]>nt\right)
$$

$$\leqslant \exp\left(-\min\left[\frac{nt^2}{16e^2M_4}, \frac{nt}{8eM_5\log(n)}\right]\right).$$

Repeat the argument with

$$[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i, \boldsymbol{\beta}, \mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]$$

replaced by

$$-[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i, \boldsymbol{\beta}, \mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}],$$

we can obtain the left bound, hence prove the result. □

Lemma S.7. *Assume that Conditions (C1) and (C6) hold. If $\mathbf{X}_i, \mathbf{U}_i \in \mathbb{R}^p$, then for $s \geqslant 1$,*

$$pr\left(\sup_{\mathbf{v}\in\mathbb{K}(2s)} |\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i, \boldsymbol{\beta}, \mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]| > nt\right)$$
$$\leqslant 2\exp\left(-\min\left[\frac{nt^2}{324e^2M_4}, \frac{nt}{36eM_5 log(n)}\right] + 2slog(9p)\right).$$

Proof of Lemma S.7: For each subset $\mathcal{U} \subset (1, \ldots, p)$, we define the set $S_{\mathcal{U}}$ as $S_{\mathcal{U}} = \{\mathbf{v} \in \mathbb{R}^p, \|\mathbf{v}\|_2 \leqslant 1, \mathrm{supp}(\mathbf{v}) \subseteq \mathcal{U}\}$, and note that $\mathbb{K}(2s) = \cup_{|\mathcal{U}|\leqslant 2s} S_{\mathcal{U}}$. We define $\mathcal{A} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset S_{\mathcal{U}}$ to be a 1/3-cover of $S_{\mathcal{U}}$, if for every $\mathbf{v} \in S_{\mathcal{U}}$, there is some $\mathbf{u}_i \in \mathcal{A}$ such that $\|\mathbf{v} - \mathbf{u}_i\|_2 \leqslant 1/3$. Define $\Delta\mathbf{v} = \mathbf{v} - \mathbf{u}_j$ where $\mathbf{u}_j = \arg\min_{\mathbf{u}_i} \|\mathbf{v} - \mathbf{u}_i\|_2$. We have $\|\Delta\mathbf{v}\|_2 \leqslant 1/3$. The same as those shown in Lemma 15 in Loh & Wainwright (2012), by Ledoux & Talagrand (2013), we can construct $\mathcal{A}$ with $|\mathcal{A}| < 9^{2s}$. Define

$$\Phi(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{v}_1^{\mathrm{T}}\left[\sum_{i=1}^{n}\frac{A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} - E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}}{n}\right]\mathbf{v}_2.$$

We have

$$\begin{aligned}
& |\Phi(\mathbf{v},\mathbf{v})| \\
= \; & |\Phi(\Delta\mathbf{v}+\mathbf{u}_j,\Delta\mathbf{v}+\mathbf{u}_j)| \\
\leqslant \; & \max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)| + \max_i |\Phi(\Delta\mathbf{v},\mathbf{u}_i)| + \max_i |\Phi(\mathbf{u}_i,\Delta\mathbf{v})| + |\Phi(\Delta\mathbf{v},\Delta\mathbf{v})| \\
\leqslant \; & \max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)| + 2\max_i |\Phi(\Delta\mathbf{v},\mathbf{u}_i)| + |\Phi(\Delta\mathbf{v},\Delta\mathbf{v})|.
\end{aligned}$$

Hence,

$$\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(\mathbf{v},\mathbf{v})| \leqslant \max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)| + 2\sup_{\mathbf{v}\in S_{\mathcal{U}}}\max_i |\Phi(\Delta\mathbf{v},\mathbf{u}_i)| + \sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(\Delta\mathbf{v},\Delta\mathbf{v})|.$$

Since $\|3\Delta\mathbf{v}\|_2 \leqslant 1$ and $\mathrm{supp}(3\Delta\mathbf{v}) \subseteq \mathcal{U}$, $3\Delta\mathbf{v} \in S_{\mathcal{U}}$. It follows that

$$\begin{aligned}
& \sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(\mathbf{v},\mathbf{v})| \\
\leqslant \; & \max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)| + 2/3\sup_{\mathbf{v}\in S_{\mathcal{U}}}\max_i |\Phi(3\Delta\mathbf{v},\mathbf{u}_i)| + 1/9\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(3\Delta\mathbf{v},3\Delta\mathbf{v})| \\
\leqslant \; & \max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)| + 2/3\{\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(3\Delta\mathbf{v},3\Delta\mathbf{v})|\}^{1/2}\{\max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)|\}^{1/2} + 1/9\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(\mathbf{v},\mathbf{v})| \\
\leqslant \; & \max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)| + \sup_{\mathbf{v}\in S_{\mathcal{U}}}\{2/3|\Phi(\mathbf{v},\mathbf{v})| + 1/9|\Phi(\mathbf{v},\mathbf{v})|\}.
\end{aligned}$$

Hence, $\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\Phi(\mathbf{v},\mathbf{v})| \leqslant 9/2\max_i |\Phi(\mathbf{u}_i,\mathbf{u}_i)|$. By Lemma S.6 and a union bound, we have

$$\begin{aligned}
& \Pr\left(\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\sum_{i=1}^n [A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]| > 9/2nt\right) \\
\leqslant \; & 9^{2s}2\exp\left(-\min\left[\frac{nt^2}{16e^2M_4}, \frac{nt}{8eM_5\log(n)}\right]\right).
\end{aligned}$$

Now replacing t with $2/9t$, we have

$$\begin{aligned} &\Pr\left(\sup_{\mathbf{v}\in S_{\mathcal{U}}} |\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]| > nt\right) \\ \leqslant\ & 9^{2s}2\exp\left(-\min\left[\frac{nt^2}{324e^2M_4}, \frac{nt}{36eM_5\log(n)}\right]\right). \end{aligned}$$

Finally, taking a union bound over the $\binom{p}{2s}$ choices of $\mathcal{U}$, and noting that $\binom{p}{2s} \leqslant p^{2s}$, we have

$$\begin{aligned} &\Pr\left(\sup_{\mathbf{v}\in \mathbb{K}(2s)} |\sum_{i=1}^{n}[A(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i)g(\mathbf{W}_i,\boldsymbol{\beta},\mathbf{v}) - \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}]| > nt\right) \\ \leqslant\ & 2\exp\left(-\min\left[\frac{nt^2}{324e^2M_4}, \frac{nt}{36eM_5\log(n)}\right] + 2s\log(9p)\right). \end{aligned}$$

Lemma S.8. *Assume that Conditions (C1) and (C6) hold. For a fixed matrix* $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times p}$*, parameter* $s > 1$*, and tolerance* $\delta > 0$*, suppose we have the deviation condition*

$$|\mathbf{v}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{v}| \leqslant \delta, \forall \mathbf{v} \in \mathbb{K}(2s).$$

Then

$$|\mathbf{v}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{v}| \leqslant 27\delta(\|\mathbf{v}\|_2^2 + 1/s\|\mathbf{v}\|_1^2), \forall \mathbf{v} \in \mathbb{R}^p.$$

Proof: This is Lemma 12 in Loh & Wainwright (2012), we omit the proofs here.

S.7 Verification of the Lower and Upper RE Conditions

Lemma S.9. *Assume that Conditions (C1) and (C6) hold and* $s \geqslant 1$,

$$n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2}-\boldsymbol{\Omega}\}$$

is an estimator for $E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}$, *satisfying the deviation condition*

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\mathbf{v}^{\mathrm{T}}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2}-\boldsymbol{\Omega}\}\mathbf{v}\\
& -\mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v}\\
\leqslant\ & \frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{54c}, \forall \mathbf{v}\in\mathbb{K}(2s)
\end{aligned}$$

for some constant c. *Then we have the lower-RE condition. That is, for any* $\mathbf{v}\in\mathbb{R}^p$,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\mathbf{v}^{\mathrm{T}}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2}-\boldsymbol{\Omega}\}\mathbf{v}\\
\geqslant\ & \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1-1/(2c)\}\|\mathbf{v}\|_2^2\\
& -\frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{2cs}\|\mathbf{v}\|_1^2.
\end{aligned}$$

We also have the upper-RE condition. That is, for any $\mathbf{v}\in\mathbb{R}^p$,

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\mathbf{v}^{\mathrm{T}}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i-\boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2}-\boldsymbol{\Omega}\}\mathbf{v}\\
\leqslant\ & \alpha_{\max}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\{1+1/(2c)\}\|\mathbf{v}\|_2^2\\
& +\frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{2cs}\|\mathbf{v}\|_1^2,
\end{aligned}$$

Proof: This result follows easily from Lemma S.8. Setting

$$\boldsymbol{\Gamma} = n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} - E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}.$$

and $\delta = \alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]/(54c)$, we have the bound

$$|\mathbf{v}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{v}| \leqslant \frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{2c}(\|\mathbf{v}\|_2^2 + 1/s\|\mathbf{v}\|_1^2).$$

Then

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\mathbf{v}^{\mathrm{T}}[\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}]\mathbf{v} \\ \geqslant\ & \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v} - \frac{\alpha_{\min}(E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\})}{2c}(\|\mathbf{v}\|_2^2 + 1/s\|\mathbf{v}\|_1^2) \end{aligned}$$

and

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\mathbf{v}^{\mathrm{T}}\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\}\mathbf{v} \\ \leqslant\ & \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v} + \frac{\alpha_{\min}(E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\})}{2c}(\|\mathbf{v}\|_2^2 + 1/s\|\mathbf{v}\|_1^2). \end{aligned}$$

Hence the lemma holds because

$$\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\|\mathbf{v}\|_2^2 \leqslant \mathbf{v}^{\mathrm{T}}E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}\mathbf{v} \leqslant \alpha_{\max}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\|\mathbf{v}\|_2^2.$$

□

Proof of Lemma 4:

Let

$$\boldsymbol{\Gamma} = n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} - E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\},$$

$$
\begin{aligned}
s &= \Bigg\{1/\{32C\max(M_4,M_5)\}\sqrt{\frac{n}{\log(p)}} \\
&\quad \min\left(\left[\frac{\sup_{\{\boldsymbol\beta:\|\boldsymbol\beta\|_2\leqslant 2b_0\}}\alpha_{\min}[E\{\exp(\boldsymbol\beta^{\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]}{54c}\right]^2/(81e^2),1\right)\Bigg\}
\end{aligned}
\quad (S.14)
$$

where C satisfies Condition (C4). Since $n/\log(p)\to\infty$ under Condition (C4), we always have $s>1$ for sufficiently large n.

Let

$$
t=\sup_{\{\boldsymbol\beta:\|\boldsymbol\beta\|_2\leqslant 2b_0\}}\frac{\alpha_{\min}[E\{\exp(\boldsymbol\beta^{\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]}{54c}.
$$

For $p\geqslant 9$, by Lemma S.7, we have

$$
\begin{aligned}
&\Pr\Bigg(\sup_{\mathbf v\in\mathbb K(2s)}\mathbf v^{\mathrm T}[n^{-1}\sum_{i=1}^n\exp(\boldsymbol\beta^{\mathrm T}\mathbf W_i-\boldsymbol\beta^{\mathrm T}\boldsymbol\Omega\boldsymbol\beta/2)\{(\mathbf W_i-\boldsymbol\Omega\boldsymbol\beta)^{\otimes 2}-\boldsymbol\Omega\} \\
&\quad -E\{\exp(\boldsymbol\beta^{\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]\mathbf v\geqslant\sup_{\{\boldsymbol\beta:\|\boldsymbol\beta\|_2\leqslant 2b_0\}}\frac{\alpha_{\min}[E\{\exp(\boldsymbol\beta^{\mathrm T}\mathbf X_i)\mathbf X_i\mathbf X_i^{\mathrm T}\}]}{54c}\Bigg) \\
\leqslant\;& 2\exp\left(-\min\left[\frac{nt^2}{324e^2M_4},\frac{nt}{36eM_5\log(n)}\right]+2s\log(9p)\right) \\
\leqslant\;& 2\exp\left(-\min\left[\frac{nt^2}{324\log(n)e^2\max(M_4,M_5)},\frac{nt}{36e\max(M_4,M_5)\log(n)}\right]+2s\log(9p)\right) \\
=\;& 2\exp\left[-n/\{4\log(n)\max(M_4,M_5)\}\min\left(\frac{t^2}{81e^2},\frac{t}{9e}\right)+2s\log(9p)\right] \\
\leqslant\;& 2\exp\left[-n/\{4\log(n)\max(M_4,M_5)\}\min\left(\frac{t^2}{81e^2},\frac{t}{9e}\right)+4s\log(p)\right] \\
\leqslant\;& 2\exp\left[-\sqrt{n\log(p)}1/\{4C\max(M_4,M_5)\}\min\left(\frac{t^2}{81e^2},\frac{t}{9e}\right)+4s\log(p)\right].
\end{aligned}
$$

where the last inequality is because of Condition (C4). If $t^2/(81e^2)>1$, then $t^2/(81e^2) > t/(9e) > 1$, hence

$$s = 1/\{32C\max(M_4, M_5)\}\sqrt{\frac{n}{\log(p)}},$$

and

$$\begin{aligned}
&\Pr\Bigg(\sup_{\mathbf{v}\in\mathbb{K}(2s)} \mathbf{v}^{\mathrm{T}}[n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} \\
&-E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\mathbf{v} \geqslant \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}} \frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{54c}\Bigg) \\
\leqslant\ & 2\exp\left[-\sqrt{n\log(p)}1/\{8C\max(M_4, M_5)\}\right].
\end{aligned}$$

On the other hand, if $t^2/(81e^2) \leqslant 1$, then $t^2/(81e^2) \leqslant t/(9e) \leqslant 1$, hence

$$\begin{aligned}
&\Pr\Bigg(\sup_{\mathbf{v}\in\mathbb{K}(2s)} \mathbf{v}^{\mathrm{T}}[n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} \\
&-E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\mathbf{v} \geqslant \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}} \frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{54c}\Bigg) \\
\leqslant\ & 2\exp\left[-\sqrt{n\log(p)}1/\{4C\max(M_4, M_5)\}\frac{t^2}{81e^2} + 1/\{8C\max(M_4, M_5)\}\sqrt{n\log(p)}\frac{t^2}{81e^2}\right] \\
=\ & 2\exp\left[-\sqrt{n\log(p)}1/\{8C\max(M_4, M_5)\}\frac{t^2}{81e^2}\right].
\end{aligned}$$

Combining the above results, we get

$$\begin{aligned}
&\Pr\Bigg(\sup_{\mathbf{v}\in\mathbb{K}(2s)} \mathbf{v}^{\mathrm{T}}[n^{-1}\sum_{i=1}^{n}\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}/2)\{(\mathbf{W}_i - \boldsymbol{\Omega}\boldsymbol{\beta})^{\otimes 2} - \boldsymbol{\Omega}\} \\
&-E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]\mathbf{v} \geqslant \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}} \frac{\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{54c}\Bigg) \\
\leqslant\ & 2\exp\left[-\sqrt{n\log(p)}1/\{8C\max(M_4, M_5)\}\min\left\{\frac{t^2}{81e^2}, 1\right\}\right].
\end{aligned}$$

Hence by Lemma S.9, the lemma holds by selecting

$$\tau(n,p) = \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}} \left(\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]/(2c)\right) \Bigg\{1/\{32C\max(M_4,M_5)\}\sqrt{\frac{n}{\log(p)}} \min\left(\left[\frac{\sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|_2\leqslant 2b_0\}}\alpha_{\min}[E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\}]}{54c}\right]^2/(81e^2),1\right)\Bigg\}^{-1}.$$

□

S.8 A Useful Topological Result

<u>**Lemma**</u> **S.10.** *For any constant $s > 1$, we have*

$$\mathbb{B}_1(\sqrt{s}) \cap \mathbb{B}_2(1) \subseteq \mathrm{cl}\{\mathrm{conv}\{\mathbb{B}_0(s) \cap \mathbb{B}_2(3)\}\}$$

where the l_k balls with radius r, $\mathbb{B}_k(r), k = 0, 1, 2$, are taken in p-dimensional space, and $\mathrm{cl}(\cdot)$ *and* $\mathrm{conv}(\cdot)$ *denote the topological closure and convex hull, respectively.*

Proof: From Lemma 11 in Loh & Wainwright (2012), we get

$$\mathbb{B}_1(\sqrt{s}) \cap \mathbb{B}_2(1) \subseteq 3\mathrm{cl}\{\mathrm{conv}\{\mathbb{B}_0(s) \cap \mathbb{B}_2(1)\}\}.$$

Here for a set A, $3A$ is defined as the set that satisfies $\sup_{\boldsymbol{\theta}\in 3A} < \boldsymbol{\theta}, \mathbf{z} >= 3\sup_{\boldsymbol{\theta}\in A} < \boldsymbol{\theta}, \mathbf{z} >$ for any z. Let U be a subset of $\{1, \ldots, p\}$ and $\mathbf{z}_U$ be the subvector of $\mathbf{z}$ with only the elements whose indices in U retained. Now when $A = \mathrm{cl}\{\mathrm{conv}\{\mathbb{B}_0(s) \cap \mathbb{B}_2(1)\}\}$, we get $\sup_{\boldsymbol{\theta}\in 3A} < \boldsymbol{\theta}, \mathbf{z} >= 3\max_{|U|=\lfloor s\rfloor}\sup_{\|\boldsymbol{\theta}_U\|_2\leqslant 1} <$

$\boldsymbol{\theta}_U, \mathbf{z}_U >= \max_{|U|=\lfloor s \rfloor} \sup_{\|\boldsymbol{\theta}_U\|_2 \leqslant 3} < \boldsymbol{\theta}_U, \mathbf{z}_U >= 3\|\mathbf{z}_S\|_2$, hence $3\mathrm{cl}\{\mathrm{conv}\{\mathbb{B}_0(s) \cap \mathbb{B}_2(1)\}\} = \mathrm{cl}\{\mathrm{conv}\{\mathbb{B}_0(s) \cap \mathbb{B}_2(3)\}\}$. Thus the results hold. □

Lemma S.10 implies that if a vector $\mathbf{v}$ satisfies $\|\mathbf{v}\|_1/\|\mathbf{v}\|_2 \leqslant \sqrt{s}$, then it automatically satisfies $\|\mathbf{v}\|_0 \leqslant s$.

Bibliography

Agarwal, A., Negahban, S., Wainwright, M. J. et al. (2012), 'Fast global convergence of gradient methods for high-dimensional statistical recovery', *The Annals of Statistics* **40**(5), 2452–2482.

Ledoux, M. & Talagrand, M. (2013), *Probability in Banach Spaces: isoperimetry and processes*, Springer Science & Business Media.

Loh, P.-L. & Wainwright, M. J. (2012), 'High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity', *The Annals of Statistics* pp. 1637–1664.