Submitted to the Annals of Applied Statistics

SUPPLEMENT TO "GENERALIZED MATRIX DECOMPOSITION REGRESSION: ESTIMATION AND INFERENCE FOR TWO-WAY STRUCTURED DATA"

BY YUE WANG[1], ALI SHOJAIE[2], TIMOTHY RANDOLPH[3], PARKER KNIGHT[4] AND JING MA[5]

[1]*Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus,* yue.2.wang@cuanschutz.edu

[2]*Department of Biostatistics, University of Washington,* ashojaie@uw.edu

[3]*Clinical Research Division, Fred Hutchinson Cancer Research Center,* trandolp@fredhutch.org

[4]*Department of Biostatistics, Harvard University,* pknight@g.harvard.edu

[5]*Public Health Sciences Division, Fred Hutchinson Cancer Research Center,* jingma@fredhutch.org

In this supplement, we prove the theoretical results in the main paper.

1. Derivations of Eq. (7) in the main text. We only focus on the high-dimensional case where $K = n < p$, noting that similar derivations can be straightforwardly applied to the case where $K > p$. It follows from the definition of $\widehat{\boldsymbol{\beta}}_{\text{KPR}}(\eta)$ that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{KPR}}(\eta) &= \left(\mathbf{X}^\intercal \mathbf{H}\mathbf{X} + \eta \mathbf{Q}^{-1}\right)^{-1} \mathbf{X}^\intercal \mathbf{H}\mathbf{y} \\ &= \mathbf{Q}(\mathbf{X}^\intercal \mathbf{H}\mathbf{X}\mathbf{Q} + \eta \mathbf{I}_p)^{-1}\mathbf{X}^\intercal \mathbf{H}\mathbf{y} \\ &= \mathbf{Q}\mathbf{X}^\intercal \mathbf{H}\left(\mathbf{X}\mathbf{Q}\mathbf{X}^\intercal \mathbf{H} + \eta \mathbf{I}_n\right)^{-1} \mathbf{y}.\end{aligned}$$

Since the GMD of $\mathbf{X}$ yields $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\intercal$, we get

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{KPR}}(\eta) &= \mathbf{Q}\mathbf{V}\mathbf{S}\mathbf{U}^\intercal \mathbf{H}\left(\mathbf{U}\left(\mathbf{S}^2 + \eta \mathbf{I}_n\right)\mathbf{U}^\intercal \mathbf{H}\right)^{-1} \mathbf{y} \\ &= \mathbf{Q}\mathbf{V}\mathbf{S}^{-1}\mathbf{S}^2(\mathbf{S}^2 + \eta \mathbf{I}_n)^{-1}\mathbf{U}^\intercal \mathbf{H}\mathbf{y}.\end{aligned} \tag{1}$$

The last equality in (1) comes from the fact that $\mathbf{U}$ is a $n \times n$ invertible matrix and $\mathbf{U}^\intercal \mathbf{H}\mathbf{U} = \mathbf{I}_n$. Denoting $\mathbf{W}_\eta = \mathbf{S}^2(\mathbf{S}^2 + \eta \mathbf{I}_n)^{-1}$, we can write

$$\widehat{\boldsymbol{\beta}}_{\text{KPR}}(\eta) = \mathbf{Q}\mathbf{V}\mathbf{S}^{-1}\mathbf{W}_\eta \mathbf{U}^\intercal \mathbf{H}\mathbf{y}. \tag{2}$$

2. Proof of Propositions 3.1 and 3.2. Recall $\widehat{\beta}_j^w(h_j) = \beta_j^w - \widehat{B}_j(h_j)$ and $\mathbf{A} = \mathbf{Q}\mathbf{V}\mathbf{W}\mathbf{S}^{-1}\mathbf{U}^\intercal \mathbf{H}\mathbf{L}_\psi^\intercal$. Plugging in the definition of β_j^w and $\widehat{B}_j(h_j)$, we get

$$\begin{aligned}\widehat{\beta}_j^w(h_j) &= \sum_{m \neq j} \xi_{jm}^w \beta_m^* + \xi_{jj}^w \beta_j^* - \sum_{m \neq j} \xi_{jm}^w \beta_m^{init} - h_j(\xi_{jj}^w - 1)\beta_j^{init} + z_j^w \\ &= \left((1 - h_j)\xi_{jj}^w + h_j\right)\beta_j^* + \sum_{m \neq j} \xi_{jm}^w(\beta_m^* - \beta_m^{init}) + \\ &\quad h_j(\xi_{jj}^w - 1)(\beta_j^* - \beta_j^{init}) + z_j^w,\end{aligned} \tag{3}$$

where $z_j^w = \sum_{i=1}^n a_{ji}\widetilde{\epsilon}_i$. To prove the asymptotic normality of z_j^w, we first check Lindeberg's condition; that is,

$$\lim_{n \to \infty} \frac{1}{s_{n,j}^2} \sum_{i=1}^n \mathbb{E}\left[a_{ji}^2 \widetilde{\epsilon}_i^2 \times I\left\{|a_{ji}\widetilde{\epsilon}_i| > t s_{n,j}\right\}\right] = 0, \text{ for all } t > 0,$$

where $s_{n,j}^2 = \sum_{i=1}^n a_{ji}^2$. Let ε be a random variable distributed like every $\widetilde{\epsilon}_i$. Then,

$$\begin{aligned}
\frac{1}{s_{n,j}^2}\sum_{i=1}^n \mathbb{E}\left[a_{ji}^2\widetilde{\epsilon}_i^2 \times I\{|a_{ji}\widetilde{\epsilon}_i| > ts_{n,j}\}\right] &= \frac{1}{s_{n,j}^2}\sum_{i=1}^n \left\{a_{ji}^2\mathbb{E}\left[\varepsilon^2 \times I\left\{|\varepsilon| > ts_{n,j}|a_{ji}|^{-1}\right\}\right]\right\} \\
&\le \frac{1}{s_{n,j}^2}\left(\sum_{i=1}^n a_{ji}^2\right)\mathbb{E}\left[\varepsilon^2 \times I\left\{|\varepsilon| > ts_{n,j}\left\{\max_{i=1,\ldots,n}|a_{ji}|\right\}^{-1}\right\}\right] \\
&= \mathbb{E}\left[\varepsilon^2 \times I\left\{|\varepsilon| > \frac{t\sqrt{\sum_{i=1}^n a_{ji}^2}}{\max_{i=1,\ldots,n}|a_{ji}|}\right\}\right].
\end{aligned}$$

Since

$$\lim_{n\to\infty}\frac{\max_{i=1,\ldots,n}|a_{ji}|}{\sqrt{\sum_{i=1}^n a_{ji}^2}} = 0,$$

by using the dominated convergence theorem, we get

$$\lim_{n\to\infty}\frac{1}{s_{n,j}^2}\sum_{i=1}^n \mathbb{E}\left[a_{ji}^2\widetilde{\epsilon}_i^2 \times I\left\{|a_{ji}\widetilde{\epsilon}_i| > ts_{n,j}\right\}\right] = 0 \text{ for all } t > 0.$$

Next, using the Lindeberg central limit theorem, we get $s_{n,j}^{-1}z_j^w \xrightarrow{d} N(0,1)$ as $n \to \infty$. Lastly, we find the explicit form of $s_{n,j}$ by noting that

$$s_{n,j}^2 = \sum_{i=1}^n a_{ji}^2 = (\mathbf{A}\mathbf{A}^\intercal)_{(j,j)} = \Omega_{jj}^w,$$

which completes the proof of Proposition 3.1.

Next, we prove Proposition 3.2. Recall that $R_{jj}^w = \sigma^2\left\{\mathbf{Q}\mathbf{V}\mathbf{W}^2\mathbf{S}^{-2}\mathbf{V}^\intercal\mathbf{Q}\right\}_{(j,j)}$. Then, we can write

$$\Omega_{jj}^w = R_{jj}^w + \mathbf{e}_j^\intercal\mathbf{Q}\mathbf{V}\mathbf{W}\mathbf{S}^{-1}\mathbf{U}^\intercal\mathbf{H}^{1/2}(\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\mathbf{L}_\psi\mathbf{H}^{1/2} - \sigma^2\mathbf{I}_n)\mathbf{H}^{1/2}\mathbf{U}\mathbf{S}^{-1}\mathbf{W}\mathbf{V}^\intercal\mathbf{Q}\mathbf{e}_j,$$

where $\mathbf{e}_j \in \mathbb{R}^n$ is the vector with one in its j-th entry and zero elsewhere. By Assumption (A1), we know that $\|\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\mathbf{L}_\psi\mathbf{H}^{1/2} - \sigma^2\mathbf{I}_n\|_2 = \|\mathbf{L}_\psi\mathbf{H}\mathbf{L}_\psi^\intercal - \sigma^2\mathbf{I}_n\|_2 = o(1)$ as $n \to \infty$. Thus, we have

$$\mathbf{e}_j^\intercal\mathbf{Q}\mathbf{V}\mathbf{W}\mathbf{S}^{-1}\mathbf{U}^\intercal\mathbf{H}^{1/2}(\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\mathbf{L}_\psi\mathbf{H}^{1/2} - \sigma^2\mathbf{I}_n)\mathbf{H}^{1/2}\mathbf{U}\mathbf{S}^{-1}\mathbf{W}\mathbf{V}^\intercal\mathbf{Q}\mathbf{e}_j = o(R_{jj}^w).$$

This further leads to

$$\frac{\Omega_{jj}^w}{R_{jj}^w} = 1 + o(1), \text{ as } n \to \infty.$$

Since $z_j^w/\sqrt{\Omega_{jj}^w} \xrightarrow{d} N(0,1)$, we have

$$\frac{z_j^w}{\sqrt{R_{jj}^w}} = \frac{z_j^w}{\sqrt{\Omega_{jj}^w}} \times \sqrt{\frac{\Omega_{jj}^w}{R_{jj}^w}} \xrightarrow{d} N(0,1) \text{ as } n \to \infty.$$

This completes the proof of Proposition 3.2.

3. Proof of Theorem 3.3. Recall that $\check{\mathbf{X}} = \mathbf{H}^{1/2}\mathbf{X}\mathbf{D}\boldsymbol{\Delta}^{1/2}$ and $\widetilde{\boldsymbol{\beta}}^* = \mathbf{D}^\intercal \boldsymbol{\beta}^*$. The next lemma characterizes $\left\|\widetilde{\boldsymbol{\beta}}(\lambda) - \widetilde{\boldsymbol{\beta}}^*\right\|_1$, where $\widetilde{\boldsymbol{\beta}}(\lambda)$ is given in eq. (13) in the main text.

LEMMA 3.1. *Suppose that* $\|\mathbf{Q}\|_2 = 1$ *and each column of* $\mathbf{X}$ *has been scaled so that* $\|\mathbf{X}\mathbf{d}_j\|_{\mathbf{H}}^2 = n$ *for* $j = 1, \ldots, p$*, where* $\mathbf{d}_j$ *is the* j*-th eigenvector of* $\mathbf{Q}$*. For any* $c_0 > 0$*, if*

$$\lambda = 2\sqrt{2c^* n \log p (1+c_0) \|\mathbf{L}_\psi^\intercal \mathbf{H} \mathbf{L}_\psi\|_2},$$

then under Assumptions (A1)–(A3), we have

$$\left\|\widetilde{\boldsymbol{\beta}}(\lambda) - \widetilde{\boldsymbol{\beta}}^*\right\|_1 = O_p\left(s_0\sqrt{n^{-1}\log p}\right) \quad \text{as } n \to \infty.$$

PROOF OF LEMMA 3.1. We first define some additional notations. For any index set $A \subset \{1, \ldots, p\}$, let $\check{\mathbf{X}}_A$ denote the submatrix of $\check{\mathbf{X}}$ with columns indexed by A, and let $\mathbf{P}_A$ denote the projection matrix onto the column space of $\check{\mathbf{X}}_A$. Define

$$\chi_m^* = \max_{|A|=m, \mathbf{s}\in\{\pm 1\}^m} \left| \boldsymbol{\varepsilon}^\intercal \frac{\check{\mathbf{X}}_A\left(\check{\mathbf{X}}_A'\check{\mathbf{X}}_A\right)^{-1}\mathbf{s}\lambda - (\mathbf{I} - \mathbf{P}_A)\check{\mathbf{X}}\boldsymbol{\Delta}^{-1/2}\widetilde{\boldsymbol{\beta}}^*}{\left\|\check{\mathbf{X}}_A\left(\check{\mathbf{X}}_A'\check{\mathbf{X}}_A\right)^{-1}\mathbf{s}\lambda - (\mathbf{I} - \mathbf{P}_A)\check{\mathbf{X}}\boldsymbol{\Delta}^{-1/2}\widetilde{\boldsymbol{\beta}}^*\right\|_2} \right|,$$

where $\boldsymbol{\varepsilon} = \mathbf{H}^{1/2}\boldsymbol{\epsilon}$, and $\lambda > 0$ is the tuning parameter. For all $m_0 \geq 0$, define the following Borel set

$$\Omega_{m_0} \equiv \{\chi_m^* \leq t_m \ \forall m \geq m_0\},$$

where

$$t_m = \sqrt{2\log p(m \vee 1)(1+c_0)\|\mathbf{L}_\psi^\intercal \mathbf{H} \mathbf{L}_\psi\|_2}$$

for some constant $c_0 > 0$. Let

$$r_1 \equiv \left(\frac{c^*\eta_1 n}{s_0 \lambda}\right)^{1/2}, \quad r_2 \equiv \left(\frac{c^*\eta_1^2 n}{s_0\lambda^2}\right)^{1/2} \text{ and } \quad C \equiv \frac{c^*}{c_*}.$$

It can be checked that with the η_1 specified in Assumption (A2) and the λ specified in Lemma 3.1, r_1 and r_2 are both $O(1)$ quantities. Let $M_1^* = 2 + 4r_1^2 + 4\sqrt{C}r_2 + 4C$, $\widetilde{S}_0 = \{j : \widetilde{\beta}_j(\lambda) \neq 0\}$ and $S_1 = \widetilde{S}_0 \cup S_0$. We follow the proof in Zhang et al. (2008), and summarize the following three major steps.

1. According to the proof of Theorem 1 in Zhang et al. (2008) (see Equation (5.26) in Zhang et al. (2008) and its derivations), we have $|S_1| \leq M_1^* s_0$ while conditioning on the event Ω_{s_0}.

2. In the second step, we show that the event Ω_{s_0} happens with high probability. In Zhang et al. (2008), this is proved for the setting where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I}_n)$. However, in our case, $\boldsymbol{\varepsilon}$ is not necessarily Gaussian, and its covariance matrix, $\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\mathbf{L}_\psi\mathbf{H}^{1/2}$, is not necessarily equal to $\mathbf{I}_n$ (up to a constant). To prove this, we first notice that for all s_0,

$$\mathbb{P}\left(\Omega_{s_0}^c\right) \leq \mathbb{P}\left(\Omega_0^c\right) \leq \sum_{m=0}^{\infty} 2^{m\vee 1}\binom{p}{m} p_m, \tag{4}$$

where

$$p_m = \mathbb{P}\left(\left| \boldsymbol{\varepsilon}^\intercal \frac{\check{\mathbf{X}}_A\left(\check{\mathbf{X}}_A'\check{\mathbf{X}}_A\right)^{-1}\mathbf{s}\lambda - (\mathbf{I} - \mathbf{P}_A)\check{\mathbf{X}}\boldsymbol{\Delta}^{-1/2}\widetilde{\boldsymbol{\beta}}^*}{\left\|\check{\mathbf{X}}_A\left(\check{\mathbf{X}}_A'\check{\mathbf{X}}_A\right)^{-1}\mathbf{s}\lambda - (\mathbf{I} - \mathbf{P}_A)\check{\mathbf{X}}\boldsymbol{\Delta}^{-1/2}\widetilde{\boldsymbol{\beta}}^*\right\|_2} \right| \geq t_m.\right)$$

with $|A| = m$ and some $\mathbf{s} \in \{\pm 1\}^m$. Next, we find an upper bound for p_m. Since $\boldsymbol{\varepsilon} = \mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal \widetilde{\boldsymbol{\epsilon}}$, where $\widetilde{\boldsymbol{\epsilon}}$ has *i.i.d* sub-Gaussian entries with mean 0 and variance 1, using the Hoeffding's inequality (Wainwright, 2019), we get

$$p_m \le 2\exp\{-\frac{t_m^2}{\|\mathbf{L}_\psi^\intercal \mathbf{H}\mathbf{L}_\psi\|_2}\} = 2\exp(-(m\vee 1)(1+c_0)\log p). \tag{5}$$

Combining (4) and (5), we have

$$\begin{aligned}
\mathbb{P}\left(\Omega_{s_0}^c\right) &\le \sum_{m=0}^{\infty} 2^{(m\vee 1)+1}\binom{p}{m}(p^{-(1+c_0)})^{m\vee 1} \\
&\le \frac{4}{p^{1+c_0}} + 2\sum_{m=1}^{\infty} 2^m \binom{p}{m} p^{-m(1+c_0)} \\
&= \frac{4}{p^{1+c_0}} + 2\left((1+2p^{-(1+c_0)})^p - 1\right) \\
&\le \frac{4}{p^{1+c_0}} + 2\left(\exp(2p^{-c_0}) - 1\right); \qquad (6)
\end{aligned}$$

here, we use the binomial theorem and the fact that $\lim_{x\to 0}(1+x)^{1/x} = e$. Therefore, we have $\mathbb{P}\left(\Omega_{s_0}\right) = 1 - \mathbb{P}\left(\Omega_{s_0}^c\right) \to 1$ as $p \to \infty$.

3. We use the techniques in the proof of Theorem 3 in Zhang et al. (2008) to bound $\left\|\widetilde{\boldsymbol{\beta}}(\lambda) - \widetilde{\boldsymbol{\beta}}^*\right\|_1$ with the pre-specified η_1 and λ. Recall that on the event Ω_{s_0}, we have $|S_1| \le M_1^* s_0$. Let $\check{\mathbf{X}}_1$ and $\check{\mathbf{X}}_2$, respectively, denote the the submatrix of $\check{\mathbf{X}}$ with columns indexed by S_1 and S_1^c. Similarly, let $\widetilde{\boldsymbol{\beta}}_1^*$ and $\widetilde{\boldsymbol{\beta}}_2^*$, respectively, denote the subvector of $\widetilde{\boldsymbol{\beta}}^*$ with entries indexed by S_1 and S_1^c; $\widetilde{\boldsymbol{\beta}}_1(\lambda)$ and $\widetilde{\boldsymbol{\beta}}_2(\lambda)$, respectively, denote the subvector of $\widetilde{\boldsymbol{\beta}}(\lambda)$ with entries indexed by S_1 and S_1^c; $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$, respectively, denote the submatrix of $\boldsymbol{\Delta}$ with both rows and columns indexed by S_1 and S_1^c. Since $q^* \ge M_1^* s_0 + 1$, by Assumption (A3), the vector $\mathbf{v}_1 = \check{\mathbf{X}}_1\boldsymbol{\Delta}_1^{-1/2}\left(\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\right)$ satisfies

$$\|\mathbf{v}_1\|_2^2 \ge c_* n \left\|\boldsymbol{\Delta}_1^{-1/2}(\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*)\right\|_2^2. \tag{7}$$

Denoting by $\mathbf{P}_1$ the projection matrix onto the column space of $\check{\mathbf{X}}_1$, we have

$$\begin{aligned}
\|\mathbf{v}_1\|_2 &\le \|\mathbf{v}_1 + \mathbf{P}_1(\widetilde{\mathbf{y}} - \check{\mathbf{X}}_1\boldsymbol{\Delta}_1^{-1/2}\widetilde{\boldsymbol{\beta}}_1(\lambda))\|_2 + \|\mathbf{P}_1(\widetilde{\mathbf{y}} - \check{\mathbf{X}}_1\boldsymbol{\Delta}_1^{-1/2}\widetilde{\boldsymbol{\beta}}_1(\lambda))\|_2 \\
&\le \|\mathbf{P}_1\check{\mathbf{X}}_2\boldsymbol{\Delta}_2^{-1/2}\widetilde{\boldsymbol{\beta}}_2^* + \mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi\boldsymbol{\epsilon}\|_2 + \|\mathbf{P}_1(\widetilde{\mathbf{y}} - \check{\mathbf{X}}_1\boldsymbol{\Delta}_1^{-1/2}\widetilde{\boldsymbol{\beta}}_1(\lambda))\|_2 \\
&\le \|\check{\mathbf{X}}_2\boldsymbol{\Delta}_2^{-1/2}\widetilde{\boldsymbol{\beta}}_2^*\|_2 + \|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi\boldsymbol{\epsilon}\|_2 + n^{-1/2}\|\Sigma_{11}^{-1/2}\mathbf{g}_1\|_2, \qquad (8)
\end{aligned}$$

where $\Sigma_{11} \equiv n^{-1}\check{\mathbf{X}}_1^\intercal\check{\mathbf{X}}_1$, $\widetilde{\mathbf{y}} = \mathbf{H}^{1/2}\mathbf{y}$ and $\mathbf{g}_1 = \check{\mathbf{X}}_1^\intercal\left(\widetilde{\mathbf{y}} - \check{\mathbf{X}}\boldsymbol{\Delta}_1^{-1/2}\widetilde{\boldsymbol{\beta}}(\lambda)\right)$. Note that the KKT conditions of eq. (13) in the main text yield that $\|\mathbf{g}_1\|_\infty \le \lambda$. Then, by Assumption (A3), we have

$$\|\Sigma_{11}^{-1/2}\mathbf{g}_1\|_2 \le c_*^{-1/2}\|\mathbf{g}_1\|_\infty\sqrt{|S_1|} \le \lambda(M_1^* s_0/c_*)^{1/2}.$$

Also, since $S_1^c \subset S_0^c$, $\|\mathbf{X}\mathbf{d}_j\|_{\mathbf{H}}^2 = n$, and $\|\mathbf{Q}\|_2 = 1$, we have $\|\check{\mathbf{X}}_2\boldsymbol{\Delta}_2^{-1/2}\widetilde{\boldsymbol{\beta}}_2^*\|_2 \le \sqrt{n}\eta_1$. We next find an upper bound for $\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi\boldsymbol{\epsilon}\|_2$. Using the Hanson-Wright inequality (Theorem 2.1 in Rudelson et al., 2013), we have

$$\mathbb{P}\left(\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\widetilde{\boldsymbol{\epsilon}}\|_2 \ge \|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\|_F + t\right) \le \exp\left\{-\frac{ct^2}{\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\|_2^2}\right\}, \tag{9}$$

for some constant c. By properties of matrix norms, we have

$$\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\|_2^2 \le \|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2; \|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\|_F^2 \le \|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2|S_1|;$$

here, we use the facts that $\|\mathbf{P}_1\|_2 = 1$ and $\|\mathbf{P}_1\|_F = |S_1|$. Thus,

$$\mathbb{P}\left(\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\widetilde{\boldsymbol{\epsilon}}\|_2 \ge \sqrt{\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2|S_1|} + t\right) \le \exp\left\{-\frac{ct^2}{\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2}\right\}.$$

Letting $t^2 = \|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2|S_1|\log p$, there exists a Borel set $\widetilde{\Omega}$ with $\mathbb{P}\left(\widetilde{\Omega}\right) \to 1$ as $p \to \infty$, such that on this set

$$\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\widetilde{\boldsymbol{\epsilon}}\|_2 \le \sqrt{\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2|S_1|}\left(1 + \sqrt{\log p}\right). \tag{10}$$

Therefore, on the event $\Omega_{s_0} \cap \widetilde{\Omega}$, we have

$$\|\mathbf{P}_1\mathbf{H}^{1/2}\mathbf{L}_\psi^\intercal\widetilde{\boldsymbol{\epsilon}}\|_2 \le \sqrt{\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2 M_1^* s_0}\left(1 + \sqrt{\log p}\right).$$

Thus,

$$\begin{aligned}\left\|\boldsymbol{\Delta}_1^{-1/2}\left(\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\right)\right\|_2 &\le (c_* n)^{-1/2}\|\mathbf{v}_1\|_2 \\ &\le c_*^{-1/2}\eta_1 + n^{-1/2}\sqrt{\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2 M_1^* s_0}\left(1 + \sqrt{\log p}\right) \\ &\quad + n^{-1/2}\lambda\left(\frac{M_1^* s_0}{nc_*}\right)^{1/2}. \end{aligned} \tag{11}$$

Since $\lambda = 2\sqrt{2c^* n \log p(1 + c_0)\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2}$, $\|\mathbf{L}_\psi^\intercal\mathbf{H}\mathbf{L}_\psi\|_2 = \sigma^2 + o(1)$, $\eta_1 = O\left(\sqrt{n^{-1}s_0\log p}\right)$ and $M_1^* = O(1)$, we have

$$\left\|\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\right\|_2 \le \left\|\boldsymbol{\Delta}_1^{1/2}\right\|_2\left\|\boldsymbol{\Delta}_1^{-1/2}\left(\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\right)\right\|_2 = O_p\left(\sqrt{n^{-1}s_0\log p}\right).$$

Therefore,

$$\|\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\|_1 \le \sqrt{s_0}\|\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\|_2 = O_p\left(s_0\sqrt{n^{-1}\log p}\right).$$

Finally,

$$\|\widetilde{\boldsymbol{\beta}}(\lambda) - \widetilde{\boldsymbol{\beta}}^*\|_1 \le \|\widetilde{\boldsymbol{\beta}}_1(\lambda) - \widetilde{\boldsymbol{\beta}}_1^*\|_1 + \eta_1 = O_p\left(s_0\sqrt{n^{-1}\log p}\right).$$

This completes the proof. □

Now we use Lemma 3.1 to prove Theorem 3.3. First, note that since $s_0 = o\left((n/\log p)^r\right)$ for some $r \in (0, 1/2)$, we have

$$\left\|\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}}(\lambda)\right\|_1 = o_p\left\{\left(\frac{\log p}{n}\right)^{1/2-r}\right\}. \tag{12}$$

Thus,

$$\begin{aligned}|\zeta_j(h_j)| &= \left|\sum_{m=1}^p \xi_{jm}^w\left(\beta_m^* - \beta_m^{init}\right) - \left((1-h_j)\xi_{jj}^w + h_j\right)\left(\beta_j^* - \beta_j^{init}\right)\right| \\ &= \left|\left[\left(\mathbf{QVWV}^\intercal - (1-h_j)\boldsymbol{\Xi} - h_j\mathbf{I}_p\right)\left(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{init}\right)\right]_j\right| \\ &\le \left\|\left[\left(\mathbf{QVWV}^\intercal - (1-h_j)\boldsymbol{\Xi} - h_j\mathbf{I}_p\right)\mathbf{D}\right]_{(j,\cdot)}\right\|_\infty\left\|\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}}(\lambda)\right\|_1. \end{aligned} \tag{13}$$

Combining (12) and (13), it can be seen that

$$\lim_{n\to\infty} \Pr\left\{ |\zeta_j(h_j)| \le \left\| \left[\left(\mathbf{QVWV}^\intercal - (1-h_j)\mathbf{\Xi} - h_j \mathbf{I}_p \right) \mathbf{D} \right]_{(j,\cdot)} \right\|_\infty \left(\frac{\log p}{n} \right)^{1/2-r} \right\} = 1.$$

Finally, Theorem 3.3 directly follows from Propositions 3.1 and 3.2.

4. Proof of Theorem 3.4. We first note that $P_j^w(h_j) \le \alpha$ is equivalent to

$$\left| \widehat{\beta}_j^w(h_j) \right| \ge \Psi_j(h_j) + q_{(1-\alpha/2)} \sqrt{R_{jj}^w}.$$

Since $\left| \widehat{\beta}_j^w(h_j) \right| = \left| \left((1-h_j)\xi_{jj}^w + h_j \right) \beta_j^* + \zeta_j(h_j) + z_j^w \right|$, we have

$$\Pr\left\{ \left| \widehat{\beta}_j^w(h_j) \right| \ge \Psi_j(h_j) + q_{(1-\alpha/2)} \sqrt{R_{jj}^w} \right\} \ge$$
$$\Pr\left\{ \left| \left((1-h_j)\xi_{jj}^w + h_j \right) \beta_j^* \right| - |\zeta_j(h_j)| - |z_j^w| \ge \Psi_j(h_j) + q_{(1-\alpha/2)} \sqrt{R_{jj}^w} \right\}.$$

Hence, it suffices to show that as $n \to \infty$,

$$\Pr\left\{ \left| \left((1-h_j)\xi_{jj}^w + h_j \right) \beta_j^* \right| - |\zeta_j(h_j)| - |z_j^w| \ge \Psi_j(h_j) + q_{(1-\alpha/2)} \sqrt{R_{jj}^w} \right\} \ge \psi. \tag{14}$$

Since $(R_{jj}^w)^{-1/2} z_j^w \xrightarrow{d} N(0,1)$ as $n \to \infty$, if

$$\frac{\left| \left((1-h_j)\xi_{jj}^w + h_j \right) \beta_j^* \right| - |\zeta_j(h_j)| - \Psi_j(h_j) - q_{(1-\alpha/2)} \sqrt{R_{jj}^w}}{\sqrt{R_{jj}^w}} \ge q_{(1-\psi/2)}, \tag{15}$$

then (14) holds. Since $\lim_{n\to\infty} \Pr(|\zeta_j(h_j)| \le \Psi_j(h_j)) = 1$ (Theorem 3.3), we know that as $n \to \infty$, (15) holds if

$$\left| \beta_j^* \right| \ge |(1-h_j)\xi_{jj}^w + h_j|^{-1} \left(2\Psi_j(h_j) + \left(q_{(1-\alpha/2)} + q_{(1-\psi/2)} \right) \sqrt{R_{jj}^w} \right);$$

this completes the proof of Theorem 3.4.

REFERENCES

RUDELSON, M., VERSHYNIN, R. et al. (2013). Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability* **18**.

WAINWRIGHT, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint* **48**. Cambridge University Press.

ZHANG, C.-H., HUANG, J. et al. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics* **36** 1567–1594.