Supplementary Material for "Regularization Methods for High-Dimensional Instrumental Variables Regression With an Application to Genetical Genomics"

Proof of Theorem 4

We first present two lemmas that are essential to the proof of Theorem 4, which concern the concentration of the empirical covariance matrix $\widehat{\mathbf{C}}$ around its population version $\mathbf{C}$ and the score vector

$$\frac{1}{n}\widehat{\mathbf{X}}^T(\mathbf{y}-\widehat{\mathbf{X}}\boldsymbol{\beta}_0)=\frac{1}{n}\widehat{\mathbf{X}}^T\boldsymbol{\eta}-\frac{1}{n}\widehat{\mathbf{X}}^T(\widehat{\mathbf{X}}-\mathbf{X})\boldsymbol{\beta}_0$$

around zero. These lemmas can be viewed as generalizations of Lemma A.3 and inequality (A.15), respectively. For ease of presentation, we condition on the event of probability $1-\pi_0$ that the two error bounds in Condition (C4) hold, and incorporate the probability π_0 into the result by the union bound.

Lemma S.1. *Under Conditions (C4)–(C6), if* $\mu_0 > 0$ *and the first-stage error bounds* e_1 *and* e_2 *satisfy*

$$s(2Le_1+e_2)\le \frac{\alpha}{(4-\alpha)\varphi}\wedge\frac{(\mu_0/2)^2}{s}, \tag{S.1}$$

then with probability at least $1-\pi_0$*, the following inequalities holds:*

$$\|(\widehat{\mathbf{C}}_{SS})^{-1}\|_\infty\le\frac{4-\alpha}{2(2-\alpha)}\varphi, \tag{S.2}$$

$$\|\widehat{\mathbf{C}}_{S^cS}(\widehat{\mathbf{C}}_{SS})^{-1}\|_\infty\le\left\{\left(1-\frac{\alpha}{2}\right)\frac{\rho'(0+)}{\rho'_\mu(b_0/2)}\right\}\wedge(2cn^\nu), \tag{S.3}$$

and

$$\Lambda_{\min}(\widehat{\mathbf{C}}_{SS})>\mu\tau_0. \tag{S.4}$$

Proof. It follows from the arguments in the proof of Lemma A.1 and Condition (C4) that

$$\max_{1\le i,j\le p}\frac{1}{n}|\widehat{\mathbf{x}}_i^T\widehat{\mathbf{x}}_j-(\mathbf{Z}\boldsymbol{\gamma}_{0i})^T\mathbf{Z}\boldsymbol{\gamma}_{0j}|\le 2Le_1+e_2.$$

Consequently, by the assumption (S.1),

$$\varphi\|\widehat{\mathbf{C}}_{SS}-\mathbf{C}_{SS}\|_\infty \le \varphi s(2Le_1+e_2) \le \frac{\alpha}{4-\alpha} \tag{S.5}$$

and

$$\varphi\|\widehat{\mathbf{C}}_{S^cS}-\mathbf{C}_{S^cS}\|_\infty \le \frac{\alpha}{4-\alpha}. \tag{S.6}$$

Then inequality (S.2) follows as in the proof of Lemma A.3.

To show inequality (S.3), by (S.2), (S.5), (S.6), and Condition (C6), we have

$$\begin{aligned}
&\|\widehat{\mathbf{C}}_{S^cS}(\widehat{\mathbf{C}}_{SS})^{-1}-\mathbf{C}_{S^cS}(\mathbf{C}_{SS})^{-1}\|_\infty\\
&\quad\le \|\widehat{\mathbf{C}}_{S^cS}-\mathbf{C}_{S^cS}\|_\infty\|(\widehat{\mathbf{C}}_{SS})^{-1}\|_\infty+\|\mathbf{C}_{S^cS}(\mathbf{C}_{SS})^{-1}\|_\infty\|\widehat{\mathbf{C}}_{SS}-\mathbf{C}_{SS}\|_\infty\|(\widehat{\mathbf{C}}_{SS})^{-1}\|_\infty\\
&\quad\le \frac{\alpha}{(4-\alpha)\varphi}\frac{4-\alpha}{2(2-\alpha)}\varphi+\left[\left\{(1-\alpha)\frac{\rho'(0+)}{\rho'_\mu(b_0/2)}\right\}\wedge(cn^\nu)\right]\frac{\alpha}{(4-\alpha)\varphi}\frac{4-\alpha}{2(2-\alpha)}\varphi\\
&\quad\le \frac{\alpha}{2(2-\alpha)}+\left\{\frac{\alpha(1-\alpha)}{2(2-\alpha)}\frac{\rho'(0+)}{\rho'_\mu(b_0/2)}\right\}\wedge\left(\frac{c}{2}n^\nu\right)\\
&\quad\le \left\{\frac{\alpha}{2}\frac{\rho'(0+)}{\rho'_\mu(b_0/2)}\right\}\wedge(cn^\nu),
\end{aligned}$$

where we have used the inequalities $\rho'(0+)/\rho'_\mu(b_0/2)\ge 1$ and $\alpha/\{2(2-\alpha)\}\le 1/2\le cn^\nu/2$. This, along with Condition (C6), implies (S.3).

Finally, it follows from the Hoffman–Wielandt inequality (Horn and Johnson 1985) and the assumption (S.1) that

$$|\Lambda_{\min}(\widehat{\mathbf{C}}_{SS})-\Lambda_{\min}(\mathbf{C}_{SS})|^2\le\|\widehat{\mathbf{C}}_{SS}-\mathbf{C}_{SS}\|_F^2\le s^2(2Le_1+e_2)\le\left(\frac{\mu_0}{2}\right)^2.$$

In view of the definition of μ_0, inequality (S.4) follows. This completes the proof of the lemma.

Lemma S.2. *Under Conditions (C4)–(C6), if the first-stage error bounds satisfy* $e_1=O(1)$ *and* $e_2=O(1)$*, then there exist constants* $c_0,c_1,c_2>0$ *such that, if we choose*

$$\mu\ge C_0n^\nu\sqrt{\frac{\log p+\log q}{n}}\vee e_2,$$

where $C_0=c_0L\max(\sigma_{p+1},M\sigma_{\max},M)$*, then with probability at least* $1-\pi_0-c_1(pq)^{-c_2}$*, it holds that*

$$\left\|\frac{1}{n}\widehat{\mathbf{X}}^T\boldsymbol{\eta}-\frac{1}{n}\widehat{\mathbf{X}}^T(\widehat{\mathbf{X}}-\mathbf{X})\boldsymbol{\beta}_0\right\|_\infty<\frac{\alpha}{6cn^\nu}\mu\rho'(0+). \tag{S.7}$$

Proof. As in the proof of Lemma A.2, we write $n^{-1}\widehat{\mathbf{X}}^T\boldsymbol{\eta}-n^{-1}\widehat{\mathbf{X}}^T(\widehat{\mathbf{X}}-\mathbf{X})\boldsymbol{\beta}_0=T_1+\cdots+T_6$. Letting $t_0=\alpha\mu\rho'(0+)/(6cn^\nu)$, we bound the six terms similarly as follows:

$$P\left(\|T_1\|_\infty\ge\frac{t_0}{6}\right)\le P\left(\left\|\frac{1}{n}\mathbf{Z}^T\boldsymbol{\eta}\right\|_\infty\ge\frac{t_0}{6e_1}\right)\le q\exp\left\{-\frac{n}{2\sigma_{p+1}^2}\left(\frac{t_0}{6e_1}\right)^2\right\},$$

$$P\left(\|T_2\|_\infty \geq \frac{t_0}{6}\right) \leq P\left(\left\|\frac{1}{n}\mathbf{Z}^T\boldsymbol{\eta}\right\|_\infty \geq \frac{t_0}{6L}\right) \leq q\exp\left\{-\frac{n}{2\sigma_{p+1}^2}\left(\frac{t_0}{6L}\right)^2\right\},$$

$$P\left(\|T_3\|_\infty \geq \frac{t_0}{6}\right) \leq P\left(\max_{1\leq i\leq q,\,1\leq j\leq p}\left|\frac{1}{n}\mathbf{z}_i^T\boldsymbol{\varepsilon}_j\right|_\infty \geq \frac{t_0}{6Me_1}\right) \leq pq\exp\left\{-\frac{n}{2\sigma_{\max}^2}\left(\frac{t_0}{6Me_1}\right)^2\right\},$$

$$P\left(\|T_4\|_\infty \geq \frac{t_0}{6}\right) \leq P\left(\max_{1\leq i\leq q,\,1\leq j\leq p}\left|\frac{1}{n}\mathbf{z}_i^T\boldsymbol{\varepsilon}_j\right|_\infty \geq \frac{t_0}{6LM}\right) \leq pq\exp\left\{-\frac{n}{2\sigma_{\max}^2}\left(\frac{t_0}{6LM}\right)^2\right\},$$

$$\|T_5\|_\infty \leq M\max_{1\leq i,j\leq p}\frac{1}{n}\|\mathbf{Z}(\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{0i})\|_2\|\mathbf{Z}(\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{0j})\|_2 \leq Me_2,$$

and

$$\|T_6\|_\infty \leq LM\max_{1\leq j\leq p}\frac{1}{\sqrt{n}}\|\mathbf{Z}(\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{0j})\|_2 \leq LM\sqrt{e_2}.$$

Combining these bounds and in view of the assumptions $e_1 = O(1)$ and $e_2 = O(1)$, there exist constants $c_0, c_1, c_2 > 0$ such that, if we choose

$$\mu \geq C_0 n^\nu \sqrt{\frac{\log p + \log q}{n}} \vee e_2,$$

where $C_0 = c_0 L\max(\sigma_{p+1}, M\sigma_{\max}, M)$, then with probability at least $1 - \pi_0 - c_1(pq)^{-c_2}$, the desired inequality holds. The completes the proof of the lemma.

Proof of Theorem 4. One can easily show that $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^p$ is a strict local minimizer of problem (4) if the following conditions hold:

$$\frac{1}{n}\widehat{\mathbf{X}}_{\widehat{S}}^T(\mathbf{y} - \widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}) = \mu\rho_\mu'(|\widehat{\boldsymbol{\beta}}_{\widehat{S}}|) \circ \operatorname{sgn}(\widehat{\boldsymbol{\beta}}_{\widehat{S}}), \tag{S.8}$$

$$\left\|\frac{1}{n}\widehat{\mathbf{X}}_{\widehat{S}^c}^T(\mathbf{y} - \widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}})\right\|_\infty < \mu\rho'(0+), \tag{S.9}$$

and

$$\Lambda_{\min}(\widehat{\mathbf{C}}_{\widehat{S}\widehat{S}}) > \mu\tau(\rho_\mu; \widehat{\boldsymbol{\beta}}_{\widehat{S}}), \tag{S.10}$$

where $\circ$ denotes the Hadamard (entrywise) product, and $|\cdot|$, $\rho_\mu'(\cdot)$, and $\operatorname{sgn}(\cdot)$ are applied componentwise. It suffices to find a $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^p$ with the desired properties such that conditions (S.8)–(S.10) hold. Let $\widehat{\boldsymbol{\beta}}_{S^c} = \mathbf{0}$. The idea of the proof is to first determine $\widehat{\boldsymbol{\beta}}_S$ from (S.8), and then show that thus obtained $\widehat{\boldsymbol{\beta}}$ also satisfies (S.9) and (S.10).

From now on, we condition on the event of probability at least $1 - \pi_0 - c_1(pq)^{-c_2}$ that the inequalities in Lemmas S.1 and S.2 hold. Using similar arguments to those in the proof of Theorem 3, (S.8) with $\widehat{S}$ replaced by S can be written in the form

$$\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_{0S} = (\widehat{\mathbf{C}}_{SS})^{-1}\left\{\frac{1}{n}\widehat{\mathbf{X}}_S^T\boldsymbol{\eta} - \frac{1}{n}\widehat{\mathbf{X}}_S^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} - \mu\rho_\mu'(|\widehat{\boldsymbol{\beta}}_S|) \circ \operatorname{sgn}(\widehat{\boldsymbol{\beta}}_S)\right\}. \tag{S.11}$$

Define the function $f\colon \mathbb{R}^s \to \mathbb{R}^s$ by $f(\boldsymbol{\theta}) = \boldsymbol{\beta}_{0S} + (\widehat{\mathbf{C}}_{SS})^{-1}\{n^{-1}\widehat{\mathbf{X}}_S^T\boldsymbol{\eta} - n^{-1}\widehat{\mathbf{X}}_S^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} - \mu\rho_\mu'(|\boldsymbol{\theta}|) \circ \operatorname{sgn}(\boldsymbol{\theta})\}$, and let $\mathcal{K}$ denote the hypercube $\{\boldsymbol{\theta} \in \mathbb{R}^s\colon \|\boldsymbol{\theta} - \boldsymbol{\beta}_{0S}\|_\infty \leq 7\varphi\mu\rho'(0+)/4\}$.

It follows from (S.2), (S.7), and Condition (C4) that, for $\boldsymbol{\theta} \in \mathcal{K}$,

$$\begin{aligned}
\|f(\boldsymbol{\theta}) - \boldsymbol{\beta}_{0S}\|_\infty &\le \|(\widehat{\mathbf{C}}_{SS})^{-1}\|_\infty \left\{ \left\| \frac{1}{n}\widehat{\mathbf{X}}_S^T \boldsymbol{\eta} - \frac{1}{n}\widehat{\mathbf{X}}_S^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} \right\|_\infty + \mu\rho'(0+) \right\} \\
&\le \frac{4-\alpha}{2(2-\alpha)}\varphi \left\{ \frac{\alpha}{6cn^\nu}\mu\rho'(0+) + \mu\rho'(0+) \right\} \\
&\le \frac{3}{2}\varphi \left\{ \frac{1}{6}\mu\rho'(0+) + \mu\rho'(0+) \right\} = \frac{7}{4}\varphi\mu\rho'(0+),
\end{aligned}$$

that is, $f(\mathcal{K}) \subset \mathcal{K}$. Also, the last inequality and the assumption (14) imply that for $\boldsymbol{\theta} \in \mathcal{K}$, $\|\boldsymbol{\theta} - \boldsymbol{\beta}_{0S}\|_\infty \le b_0/2$, and hence $\operatorname{sgn}(\boldsymbol{\theta}) = \operatorname{sgn}(\boldsymbol{\beta}_{0S})$. Thus, in view of Condition (C4), f is a continuous function on the convex, compact hypercube $\mathcal{K}$. An application of Brouwer's fixed point theorem yields that equation (S.11) has a solution $\widehat{\boldsymbol{\beta}}_S$ in $\mathcal{K}$. Moreover, $\operatorname{sgn}(\widehat{\boldsymbol{\beta}}_S) = \operatorname{sgn}(\boldsymbol{\beta}_{0S})$, so that $\widehat{S} = S$. Therefore, we have found a $\widehat{\boldsymbol{\beta}}$ that satisfies the desired properties and (S.8).

To verify that $\widehat{\boldsymbol{\beta}}$ satisfies (S.9), by substituting (S.11), we write

$$\begin{aligned}
\frac{1}{n}\widehat{\mathbf{X}}_{S^c}^T(\mathbf{y} - \widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}) &= \frac{1}{n}\widehat{\mathbf{X}}_{S^c}^T\boldsymbol{\eta} - \frac{1}{n}\widehat{\mathbf{X}}_{S^c}^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} \\
&\quad - \widehat{\mathbf{C}}_{S^cS}(\widehat{\mathbf{C}}_{SS})^{-1} \left\{ \frac{1}{n}\widehat{\mathbf{X}}_S^T\boldsymbol{\eta} - \frac{1}{n}\widehat{\mathbf{X}}_S^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} - \mu\rho'_\mu(|\widehat{\boldsymbol{\beta}}_S|) \circ \operatorname{sgn}(\widehat{\boldsymbol{\beta}}_S) \right\}.
\end{aligned}$$

Also, we have $\|\widehat{\boldsymbol{\beta}}_S\|_\infty = \|\widehat{\boldsymbol{\beta}}_{0S} + (\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_{0S})\|_\infty \ge \|\widehat{\boldsymbol{\beta}}_{0S}\|_\infty - \|\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_{0S}\|_\infty \ge b_0 - b_0/2 = b_0/2$. This, together with (S.3), (S.7), and Condition (C4), leads to

$$\begin{aligned}
\left\| \frac{1}{n}\widehat{\mathbf{X}}_{S^c}^T(\mathbf{y} - \widehat{\mathbf{X}}\boldsymbol{\beta}) \right\|_\infty &\le \left\| \frac{1}{n}\widehat{\mathbf{X}}_{S^c}^T\boldsymbol{\eta} - \frac{1}{n}\widehat{\mathbf{X}}_{S^c}^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} \right\|_\infty + \|\widehat{\mathbf{C}}_{S^cS}(\widehat{\mathbf{C}}_{SS})^{-1}\|_\infty \\
&\quad \times \left\{ \left\| \frac{1}{n}\widehat{\mathbf{X}}_S^T\boldsymbol{\eta} - \frac{1}{n}\widehat{\mathbf{X}}_S^T(\widehat{\mathbf{X}}_S - \mathbf{X}_S)\boldsymbol{\beta}_{0S} \right\|_\infty + \mu\rho'_\mu(b_0/2) \right\} \\
&< \frac{\alpha}{6cn^\nu}\mu\rho'(0+) + 2cn^\nu \cdot \frac{\alpha}{6cn^\nu}\mu\rho'(0+) + \left(1 - \frac{\alpha}{2}\right)\frac{\rho'(0+)}{\rho'_\mu(b_0/2)} \cdot \mu\rho'_\mu(b_0/2) \\
&\le \frac{\alpha}{6}\mu\rho'(0+) + \frac{\alpha}{3}\mu\rho'(0+) + \left(1 - \frac{\alpha}{2}\right)\mu\rho'(0+) = \mu\rho'(0+).
\end{aligned}$$

Finally, it follows from (S.4) and the definition of τ_0 that $\Lambda_{\min}(\widehat{\mathbf{C}}_{SS}) > \mu\tau_0 \ge \mu\tau(\rho_\mu; \widehat{\boldsymbol{\beta}}_S)$, which verifies (S.10) and completes the proof.