Submitted to Bernoulli
arXiv: arXiv:0000.0000

Supplement to "Inference of the Mean Outcome under an Optimal Treatment Regime in High Dimensions"

CHENGCHUN SHI[1,*] RUI SONG[1,**] WENBIN LU[1,†] and DAOWEN ZHANG[1,‡]

[1] *Department of statistics, North Carolina State University, Raleigh, NC 27695-8203*
E-mail: [*]cshi4@ncsu.edu; [**]rsong@ncsu.edu; [†]wenbin_lu@ncsu.edu; [‡]dzhang2@ncsu.edu

Proof of Theorem 5.1. Recall that $\rho(t,\lambda) = p_\lambda(t)/\lambda$. We will write $\rho_1(t) = \rho(t,\lambda_1)$. Let $\rho_1'(\cdot)$ denote the derivative of the function $\rho_1(\cdot)$ and $\bar{\rho}_1(t) = \rho_1'(|t|)$. For any vector $\boldsymbol{v} = (v_1,\dots,v_q)$, let $\bar{\rho}_1(\boldsymbol{v}) = \{\bar{\rho}_1(v_1),\dots,\bar{\rho}_1(v_q)\}^T$.

We first prove the oracle property of $\widehat{\boldsymbol{\alpha}}$. Similar to the proofs of Theorem 3 and Theorem 4 in [1], we break the proof into three steps. In the first step, we constrain the optimization function in (4.3) on the s_α-dimensional subspace $\{\boldsymbol{\alpha}\in\mathbb{R}^p : \boldsymbol{\alpha}_{\mathcal{M}_\alpha^c} = 0\}$ and show there exists some local maximizer $\widehat{\boldsymbol{\alpha}}$ with $O_p(\sqrt{s_\alpha/n})$ convergence rate. In the next step, we show with probability tending to 1, $\widehat{\boldsymbol{\alpha}}$ satisfies the following condition:

$$\max_{j\in\mathcal{M}_\alpha^c}\left|\sum_{i=1}^n X_{i,j}\{A_i - \pi(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})\}\right| \le n\lambda_1\rho_1'(0+). \tag{0.1}$$

This implies that $\widehat{\boldsymbol{\alpha}}$ is indeed a local minimizer of (4.3). Finally, we show $\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha}$ is asymptotically linear. The proof is then completed.

Step 1: Recall that $\widehat{\boldsymbol{\alpha}}$ is the minimizer of the penalized likelihood function,

$$\overline{Q}_\pi(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^n\left\{\log\{1+\exp(\boldsymbol{X}_i^T\boldsymbol{\alpha})\} - A_i\boldsymbol{\alpha}^T\boldsymbol{X}_i\right\} + \sum_{j=1}^p p_{\lambda_1}(|\alpha_j|).$$

Define the set

$$N_{\pi,\tau} = \left\{\boldsymbol{\alpha}\in\mathbb{R}^p : \|\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2 \le \sqrt{s_\alpha/n}\,\tau, \boldsymbol{\alpha}_{\mathcal{M}_\alpha^c} = 0\right\},$$

for $\tau = (0,+\infty)$. Consider the event

$$H_{\pi,n} = \left\{\overline{Q}_\pi(\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}) > \sup_{\substack{\boldsymbol{\alpha}\in\partial N_{\pi,\tau}\\ \boldsymbol{\alpha}_{\mathcal{M}_\alpha^c}=0}} \overline{Q}_\pi(\boldsymbol{\alpha})\right\},$$

where $\partial N_{\pi,\tau}$ denotes the boundary of $N_{\pi,\tau}$. Note that $\boldsymbol{\alpha}^*$ satisfies $\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha^c} = 0$. On the event $H_{\pi,n}$, there exists a local minimizer $\widehat{\boldsymbol{\alpha}}$ of $\overline{Q}(\boldsymbol{\alpha})$ on the s_α-dimensional subspace $\{\boldsymbol{\alpha} \in \mathbb{R}^p : \boldsymbol{\alpha}_{\mathcal{M}_\alpha^c} = 0\}$. Therefore, it suffices to show that for any sufficiently small $\varepsilon > 0$, there exists some $\tau > 0$ such that $\liminf \Pr(H_{\pi,n}) \geq 1 - \varepsilon$.

For any $\tau > 0$ and $\boldsymbol{\alpha} \in N_{\pi,\tau}$, a second order Taylor expansion around $\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}$ gives

$$\overline{Q}(\boldsymbol{\alpha}) - \overline{Q}(\boldsymbol{\alpha}^*) = (\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T \boldsymbol{v}_\pi - \frac{1}{2}(\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T \boldsymbol{D}_\pi (\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}), \quad (0.2)$$

where $\boldsymbol{v}_\pi = n^{-1}\sum_i \boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^*)\} - \lambda_1 \bar{\rho}_1(\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})$, and

$$\boldsymbol{D}_\pi = \frac{1}{n}\sum_i \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\{1 - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}^T_{i,\mathcal{M}_\alpha} + \mathrm{diag}\left\{p''_{\lambda_1}(|\boldsymbol{\alpha}^{**}_{\mathcal{M}_\alpha}|)\right\},$$

for some $\boldsymbol{\alpha}^{**}$ lying on the line segment joining $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$. When the second order derivative of p_λ doesn't exists, the second part of the matrix $\boldsymbol{D}_\pi$ can be replaced by a diagonal matrix with maximum absolute value bounded by $\lambda_1\kappa_\alpha$. Under Condition (C2), we have $\lambda_1\kappa_\alpha = o(1)$. This implies

$$\left\|\boldsymbol{D}_\pi - \frac{1}{n}\sum_i \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\{1 - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}^T_{i,\mathcal{M}_\alpha}\right\|_2 = o(1). \quad (0.3)$$

Similarly to (6.17), we have with probability tending to 1, $\max_{i\in\{1,\dots,n\}} |\boldsymbol{X}_i^T\boldsymbol{\alpha}^{**}| \leq 2\omega^*$. This further implies

$$\Pr\left(\pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\{1 - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\} \geq \frac{1}{\{1+\exp(2\omega^*)\}^2}\right) \to 1.$$

Hence, the following matrix is positive definite,

$$\frac{1}{n}\sum_{i=1}^n \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\{1 - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}^T_{i,\mathcal{M}_\alpha} - \frac{1}{n\{1+\exp(2\omega^*)\}^2}\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}^T_{i,\mathcal{M}_\alpha},$$

with probability tending to 1. Therefore, we have

$$\begin{aligned}\lambda_{\min}&\left(\frac{1}{n}\sum_{i=1}^n \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\{1 - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^{**})\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}^T_{i,\mathcal{M}_\alpha}\right) \quad (0.4)\\ &\geq \lambda_{\min}\left(\frac{1}{n\{1+\exp(2\omega^*)\}^2}\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}^T_{i,\mathcal{M}_\alpha}\right),\end{aligned}$$

with probability tending to 1. The matrix $\boldsymbol{\Sigma}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha} - \boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}$ is positive definite. By Condition (A11), we have

$$\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}) \geq \lambda_{\min}(\boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}) \geq c_3. \quad (0.5)$$

Similar to (6.25), we can show that

$$\Pr\left(\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T - \boldsymbol{\Sigma}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}\right\|_2 = O\left(\frac{s_\alpha\sqrt{\log n}}{\sqrt{n}}\right)\right) \to 1. \tag{0.6}$$

In view of (0.5), under the condition $s_\alpha = o(n^{1/3})$, we have

$$\Pr\left(\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T\right) \geq \frac{c_3}{2}\right) \to 1.$$

This together with (0.4) implies that

$$\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^{**})\{1-\pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^{**})\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T\right) \geq \frac{c_3}{2\{1+\exp(2\omega^*)\}^2},$$

with probability tending to 1. Let $c_3^* = c_3/[3\{1+\exp(2\omega^*)\}^2]$. By (0.3), we have

$$\lambda_{\min}(\boldsymbol{D}_\pi) \geq c_3^*, \tag{0.7}$$

with probability tending to 1.

Under the event defined in (0.7), we have for any $\boldsymbol{\alpha} \in \partial N_{\pi,\tau}$,

$$\frac{1}{2}(\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha}^*)^T\boldsymbol{D}_\pi(\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha}^*) \geq \frac{c_3^*}{2}\|\boldsymbol{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha}^*\|_2^2 = \frac{c_3^* s_\alpha \tau^2}{2n}. \tag{0.8}$$

Besides, under Condition (C1) and (C2), we have

$$\|\lambda_1\bar{\rho}_1(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}^*)\|_2 \leq \sqrt{s_\alpha}\|\lambda_1\rho_1'(d_{n,\alpha})\|_2 = o(n^{-1/2}).$$

Therefore, for any sufficiently small $\varepsilon_0 > 0$, we have

$$\|\lambda_1\bar{\rho}_1(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}^*)\|_2 \leq \varepsilon_0 n^{-1/2}. \tag{0.9}$$

Moreover, it follows from (4.1) that

$$\mathrm{E}\boldsymbol{X}_{0,\mathcal{M}_\alpha}\{A_0 - \pi^*(\boldsymbol{X}_0,\boldsymbol{\alpha}^*)\} = 0.$$

Therefore, we have

$$\begin{aligned} \frac{1}{n}\mathrm{E}\left\|\sum_i \boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\right\|_2^2 &= \mathrm{trace}\left(\frac{1}{n}\sum_i \mathrm{E}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}^2\boldsymbol{X}_{0,\mathcal{M}_\alpha}^T\right) \\ \leq \quad \mathrm{trace}\left(\mathrm{E}\boldsymbol{X}_{0,\mathcal{M}_\alpha}\boldsymbol{X}_{0,\mathcal{M}_\alpha}^T\right) &\leq s_\alpha\lambda_{\max}\left(\mathrm{E}\boldsymbol{X}_{0,\mathcal{M}_\alpha}\boldsymbol{X}_{0,\mathcal{M}_\alpha}^T\right) = c_4 s_\alpha. \end{aligned}$$

This together with (0.9) implies that

$$\mathrm{E}\|\boldsymbol{v}_\pi\|^2 \leq \mathrm{E}\left\|\frac{1}{n}\sum_i \boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\} + \lambda_1\bar{\rho}_1(\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})\right\|_2^2 \tag{0.10}$$
$$\leq 2\mathrm{E}\left\|\frac{1}{n}\sum_i \boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\right\|_2^2 + \frac{2\varepsilon_0^2}{n} = \frac{2}{n}(\varepsilon_0^2 + c_4 s_\alpha).$$

By (0.8), (0.10) and Markov's inequality, we have

$$\begin{aligned}
& \Pr\left(\inf_{\boldsymbol{\alpha}\in\partial N_{\pi,\tau}}\left\{\frac{1}{2}(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T\boldsymbol{D}_\pi(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}) - (\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T\boldsymbol{v}_\pi\right\} > 0\right)\\
\geq\ & \Pr\left(\inf_{\boldsymbol{\alpha}\in\partial N_{\pi,\tau}}\left\{\frac{1}{2}(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T\boldsymbol{D}_\pi(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}) - (\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T\boldsymbol{v}_\pi\right\} > 0, \lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*\right)\\
\geq\ & \Pr\left(\inf_{\boldsymbol{\alpha}\in\partial N_{\pi,\tau}}\left\{\frac{c_3^*}{2}\|\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2^2 - \|\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2\|\boldsymbol{v}_\pi\|_2\right\} > 0, \lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*\right)\\
=\ & \Pr\left(\inf_{\boldsymbol{\alpha}\in\partial N_{\pi,\tau}}\left\{\frac{c_3^*}{2}\|\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2 - \|\boldsymbol{v}_\pi\|_2\right\} > 0, \lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*\right)\\
=\ & \Pr\left(\|\boldsymbol{v}_\pi\|_2 \leq \frac{c_3^*\sqrt{s_\alpha}\tau}{2\sqrt{n}}, \lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*\right) \geq \Pr\left(\|\boldsymbol{v}_\pi\|_2 \leq \frac{c_3^*\sqrt{s_\alpha}\tau}{2\sqrt{n}}\right) - \Pr(\lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*)\\
\geq\ & 1 - \frac{4n\mathrm{E}\|\boldsymbol{v}_\pi\|_2^2}{(c_3^*)^2\tau^2 s_\alpha} - \Pr(\lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*) \geq 1 - \frac{8(\varepsilon_0^2 + c_4 s_\alpha)}{(c_3^*)^2\tau^2 s_\alpha} - \Pr(\lambda_{\min}(\boldsymbol{D}_\pi)\geq c_3^*).
\end{aligned}$$

Set $\varepsilon_0 = \sqrt{c_4 s_\alpha}$. For any $\varepsilon > 0$, take $\tau = 4\sqrt{c_4}/(c_3^*\sqrt{\varepsilon})$. By (0.7), we have

$$\begin{aligned}
& \liminf_n \Pr\left(\inf_{\boldsymbol{\alpha}\in\partial N_{\pi,\tau}}\left\{\frac{1}{2}(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T\boldsymbol{D}_\pi(\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}) - (\boldsymbol{\alpha}_{\mathcal{M}_\alpha}-\boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})^T\boldsymbol{v}_\pi\right\} > 0\right)\\
=\ & 1 - \frac{8(\varepsilon_0^2 + c_4 s_\alpha)}{(c_3^*)^2\tau^2 s_\alpha} = 1 - \varepsilon.
\end{aligned}$$

This completes the first step of the proof.

Step 2: Denoted by $\widehat{\boldsymbol{\alpha}}$ the local maximizer of $\overline{Q}_\pi(\boldsymbol{\alpha})$ constrained on the s_α-dimensional subspace $\{\boldsymbol{\alpha}\in\mathbb{R}^p : \boldsymbol{\alpha}_{\mathcal{M}_\alpha^c} = 0\}$. We have

$$\|\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2 = O_p\left(\frac{\sqrt{s_\alpha}}{\sqrt{n}}\right). \tag{0.11}$$

In the following, we show (0.1) is satisfied. By Condition (C1), $\rho_1'(0+) \triangleq \partial\rho(0+,\lambda_1)/\partial t$ is independent of λ_1. Hence, it suffices to show

$$\max_{j\in\mathcal{M}_\alpha^c}\left|\sum_{i=1}^n X_{i,j}\{A_i - \pi^*(\boldsymbol{X}_i,\widehat{\boldsymbol{\alpha}})\}\right| = o_p(n\lambda_1). \tag{0.12}$$

By Taylor's theorem, we have

$$\sum_{i=1}^n X_{i,j}\{A_i - \pi^*(\boldsymbol{X}_i, \widehat{\boldsymbol{\alpha}})\} = \underbrace{\sum_{i=1}^n X_{i,j}\{A_i - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^*)\}}_{I_{1,j}}$$

$$\underbrace{-\sum_{i=1}^n X_{i,j}\pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}_j^*)\{1-\pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}_j^*)\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T(\widehat{\alpha}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})}_{I_{2,j}},$$

for some $\boldsymbol{\alpha}_j^*$ lying on the line segment between $\boldsymbol{\alpha}^*$ and $\widehat{\boldsymbol{\alpha}}$. In the following, we show $\max_j |I_{1,j}| = o_p(n\lambda_1)$ and $\max_j |I_{2,j}| = o_p(n\lambda_1)$. The assertion (0.12) is thus satisfied.

Consider $\max_j |I_{1,j}|$. By Condition (A5) and the definition of the Orlicz norm, we have

$$\max_j \|X_{0,j}\{A_0 - \pi^*(\boldsymbol{X}_0, \boldsymbol{\alpha}^*)\}\|_{\psi_1} \le \max_j \|X_{0,j}\|_{\psi_1} \le \omega_0.$$

By (4.1), for any $j \in \{1, \dots, p\}$, we have

$$\mathrm{E}X_{0,j}\{A_0 - \pi^*(\boldsymbol{X}_0, \boldsymbol{\alpha}^*) = 0.$$

Similar to (6.24), we can show

$$\Pr\left(\max_j |I_{1,j}| = O(\sqrt{n\log p})\right) \to 1.$$

By Condition (C2), we have $n\lambda_1 \gg \sqrt{n \log p}$. This yields $\max_j |I_{1,j}| = o_p(n\lambda_1)$.

It remains to show $\max_j |I_{2,j}| = o_p(n\lambda_1)$. By Condition (A5) and Cauchy-Schwarz inequality, we have

$$\max_j |I_{2,j}| \le \max_j \left|\sum_{i=1}^n |X_{i,j}||\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})|\right| \tag{0.13}$$

$$\le \quad \omega_0 \sum_{i=1}^n |\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})| \le \omega_0 \sqrt{n \sum_{i=1}^n |\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})|^2}$$

$$\le \quad \omega_0 \sqrt{n\lambda_{\max}\left(\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T\right)\|\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2^2}.$$

Similar to (6.26), we can show

$$\lambda_{\max}\left(\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T\right) = O_p(n).$$

Combining this together with (0.11) and (0.13) yields

$$\max_j |I_{2,j}| = O_p(\sqrt{ns_\alpha}).$$

Under the condition $\lambda_1 \gg \sqrt{s_\alpha/n}$, we have $\max_j |I_{2,j}| = o_p(n\lambda_1)$. This proves (0.1).

Step 3: In Step 2, we've shown

$$\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha^c} = 0, \tag{0.14}$$

with probability tending to 1.

Since $\widehat{\boldsymbol{\alpha}}$ is local minimizer of (4.3), we have

$$\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}^T \{A_i - \pi^*(\boldsymbol{X}_i, \widehat{\boldsymbol{\alpha}})\} + n\lambda_1 \bar{\rho}_1(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha}) = 0. \tag{0.15}$$

By (0.11) and the condition $d_{n,\alpha} \gg \sqrt{s_\alpha/n}$, we have with probability tending to 1,

$$\min_{j\in\mathcal{M}_\alpha} |\widehat{\alpha}_j| \geq \min_{j\in\mathcal{M}_\alpha} |\alpha_j^*| - \max_{j\in\mathcal{M}_\alpha} |\widehat{\alpha}_j - \alpha_j^*| = 2d_{n,\alpha} - \|\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}\|_2 \gg 2d_{n,\alpha} - d_{n,\alpha} = d_{n,\alpha}.$$

By the monotonicity of $\rho'(\cdot,\lambda)$ in (C1) and the condition $p'_{\lambda_1}(d_{n,\alpha}) = o(s_\alpha^{-1/2} n^{-1/2})$ in (C2), we have

$$\|\lambda_1 \bar{\rho}_1(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha})\|_2 \leq \sqrt{s_\alpha} \|\lambda_1 \bar{\rho}_1(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha})\|_\infty \leq \sqrt{s_\alpha} \lambda_1 \rho_1(d_{n,\alpha}) = o(n^{-1/2}),$$

with probability tending to 1. By (0.15), we have

$$\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}^T \{A_i - \pi^*(\boldsymbol{X}_i, \widehat{\boldsymbol{\alpha}})\} = o_p(n^{-1/2}). \tag{0.16}$$

Besides, under the event defined in (0.14), it follows from Taylor's theorem that

$$\begin{aligned} \sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha} \{A_i - \pi^*(\boldsymbol{X}_i, \widehat{\boldsymbol{\alpha}})\} &= \sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha} \{A_i - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^*)\} \qquad (0.17) \\ - \quad \sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha} \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^*) &\{1 - \pi^*(\boldsymbol{X}_i, \boldsymbol{\alpha}^*)\} \boldsymbol{X}_{i,\mathcal{M}_\alpha}^T (\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha}) + \boldsymbol{R}_\pi, \end{aligned}$$

where the remainder term $\boldsymbol{R}_\pi$ satisfies

$$\|\boldsymbol{R}_\pi\|_\infty \leq \max_{j\in\mathcal{M}_\alpha} \sum_{i=1}^n |X_{i,j}| \|\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T (\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})\|_2^2,$$

under the event defined in (0.14). Similar to (6.20), we can show

$$\|\boldsymbol{R}_\pi\|_\infty \leq \omega_0 \sum_{i=1}^n \|\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T (\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}^*_{\mathcal{M}_\alpha})\|_2^2 = O_p(s_\alpha).$$

Hence,

$$\|\boldsymbol{R}_\pi\|_2 \le \sqrt{s_\alpha}\|\boldsymbol{R}_\pi\|_\infty = O_p(s_\alpha^{3/2}) = o_p(n^{1/2}), \tag{0.18}$$

under the condition $s_\alpha = o(n^{1/3})$.

Moreover, similar to (0.6), we have

$$\Pr\left(\left\|\frac{1}{n}\sum_{i=1}^n \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\{1-\pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T - \boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}\right\|_2 = O\left(\frac{s_\alpha\sqrt{\log n}}{\sqrt{n}}\right)\right) \to 1.$$

By (0.11) and the condition $s_\alpha = o(n^{1/3})$, this further yields

$$\begin{aligned}
&\left|\left(\sum_{i=1}^n \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\{1-\pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T - n\boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}\right)(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha})\right| \\
\le\ & \left\|\sum_{i=1}^n \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\{1-\pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\}\boldsymbol{X}_{i,\mathcal{M}_\alpha}\boldsymbol{X}_{i,\mathcal{M}_\alpha}^T - n\boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}\right\|_2 \|\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha}\|_2 \\
=\ & O_p\left(\sqrt{s_\alpha n\log n}\right) O_p\left(\frac{\sqrt{s_\alpha}}{\sqrt{n}}\right) = o_p(n^{1/2}).
\end{aligned}$$

Combining this together with (0.17) and (0.18), we've shown

$$\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\} = n\boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha}) + o_p(n^{1/2}).$$

Under the condition $\liminf \lambda_{\min}(\boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}) \ge c_3$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha} - \boldsymbol{\alpha}_{\mathcal{M}_\alpha}) = \frac{1}{\sqrt{n}}\boldsymbol{\Omega}_{\mathcal{M}_\alpha,\mathcal{M}_\alpha}^{-1}\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}_\alpha}\{A_i - \pi^*(\boldsymbol{X}_i,\boldsymbol{\alpha}^*)\} + o_p(1).$$

The asymptotic linearity of $\widehat{\boldsymbol{\alpha}}_{\mathcal{M}_\alpha}$ is thus proven. □

References

[1] Jianqing Fan and Jinchi Lv. Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans. Inform. Theory*, 57(8):5467–5484, 2011.