arXiv: arXiv:1503.08393

SUPPLEMENT TO "SLOPE IS ADAPTIVE TO UNKNOWN SPARSITY AND ASYMPTOTICALLY MINIMAX"

By Weijie Su* and Emmanuel Candès†

Stanford University

APPENDIX A: PROOFS OF TECHNICAL RESULTS

As is standard, we write $a_n \asymp b_n$ for two positive sequences a_n and b_n if there exist two constants C_1 and C_2 (possibly depending on q) such that $C_1 a_n \le b_n \le C_2 a_n$ for all n. Also, we write $a_n \sim b_n$ if $a_n/b_n \to 1$.

A.1. Proofs for Section 2. We remind the reader that the proofs in this subsection rely on some lemmas to be stated later in the Appendix.

Proof of (2.1). For simplicity, denote by $\widehat{\boldsymbol{\beta}}$ the (full) Lasso solution $\widehat{\boldsymbol{\beta}}_{\text{Lasso}}$, and $\widehat{\boldsymbol{b}}_S$ the solution to the reduced Lasso problem

$$\underset{\boldsymbol{b}\in\mathbb{R}^k}{\text{minimize}}\ \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}_S\boldsymbol{b}\|^2+\lambda\|\boldsymbol{b}\|_1,$$

where S is the support of the ground truth $\boldsymbol{\beta}$. We show that (i)

$$\left\|\boldsymbol{X}_{\overline{S}}'\boldsymbol{z}\right\|_\infty \le (1+c/2)\sqrt{2\log p} \tag{A.1}$$

and (ii)

$$\left\|\boldsymbol{X}_{\overline{S}}'\boldsymbol{X}_S(\boldsymbol{\beta}_S-\widehat{\boldsymbol{b}}_S)\right\|_\infty < C\sqrt{(k\log^2 p)/n} \tag{A.2}$$

for some constant C, both happen with probability tending to one. Now observe that $\boldsymbol{X}_{\overline{S}}'(\boldsymbol{y}-\boldsymbol{X}_S\widehat{\boldsymbol{b}}_S)=\boldsymbol{X}_{\overline{S}}'\boldsymbol{z}+\boldsymbol{X}_{\overline{S}}'\boldsymbol{X}_S(\boldsymbol{\beta}_S-\widehat{\boldsymbol{b}}_S)$. Hence, combining (A.1) and (A.2) and using the fact that $(k\log p)/n\to 0$ give

$$\begin{aligned}\left\|\boldsymbol{X}_{\overline{S}}'(\boldsymbol{y}-\boldsymbol{X}_S\widehat{\boldsymbol{b}}_S)\right\|_\infty &\le \left\|\boldsymbol{X}_{\overline{S}}'\boldsymbol{X}_S(\boldsymbol{\beta}_S-\widehat{\boldsymbol{b}}_S)\right\|_\infty+\left\|\boldsymbol{X}_{\overline{S}}'\boldsymbol{z}\right\|_\infty\\ &\le C\sqrt{(k\log^2 p)/n}+(1+c/2)\sqrt{2\log p}\\ &= o(\sqrt{2\log p})+(1+c/2)\sqrt{2\log p}\\ &< (1+c)\sqrt{2\log p}\end{aligned}$$

*Supported in part by a General Wang Yaowu Stanford Graduate Fellowship.

†Supported in part by NSF under grant CCF-0963835 and by the Math + X Award from the Simons Foundation.

with probability approaching one. This last inequality together with the fact that $\widehat{\boldsymbol{b}}_S$ obeys the KKT conditions for the reduced Lasso problem imply that padding $\widehat{\boldsymbol{b}}_S$ with zeros on $\overline{S}$ obeys the KKT conditions for the full Lasso problem and is, therefore, solution.

We need to justify (A.1) and (A.2). First, Lemmas A.6 and A.5 imply (A.1). Next, to show (A.2), we rewrite the left-hand side in (A.2) as

$$\boldsymbol{X}_{\overline{S}}'\boldsymbol{X}_S(\boldsymbol{\beta}_S - \widehat{\boldsymbol{b}}_S) = \boldsymbol{X}_{\overline{S}}'\boldsymbol{X}_S(\boldsymbol{X}_S'\boldsymbol{X}_S)^{-1}(\boldsymbol{X}_S'(\boldsymbol{y} - \boldsymbol{X}_S\widehat{\boldsymbol{b}}_S) - \boldsymbol{X}_S'\boldsymbol{z}).$$

By Lemma A.7, we have that

$$\left\|\boldsymbol{X}_S'(\boldsymbol{y} - \boldsymbol{X}_S\widehat{\boldsymbol{b}}_S) - \boldsymbol{X}_S'\boldsymbol{z}\right\| \le \sqrt{k}\lambda + \left\|\boldsymbol{X}_S'\boldsymbol{z}\right\| \le \sqrt{k}\lambda + \sqrt{32k\log(p/k)} \le C'\sqrt{k\log p}$$

holds with probability at least $1 - \mathrm{e}^{-n/2} - (\sqrt{2}\mathrm{e}k/p)^k \to 1$. In addition, Lemma A.11 with $t = 1/2$ gives

$$\left\|\boldsymbol{X}_S(\boldsymbol{X}_S'\boldsymbol{X}_S)^{-1}\right\| \le \frac{1}{\sqrt{1 - 1/n} - \sqrt{k^\star/n} - 1/2} < 3$$

with probability at least $1 - \mathrm{e}^{-n/8} \to 1$. Hence, from the last two inequalities it follows that

$$\left\|\boldsymbol{X}_S(\boldsymbol{X}_S'\boldsymbol{X}_S)^{-1}(\boldsymbol{X}_S'(\boldsymbol{y} - \boldsymbol{X}_S\widehat{\boldsymbol{b}}_S) - \boldsymbol{X}_S'\boldsymbol{z})\right\| \le C''\sqrt{k\log p} \tag{A.3}$$

with probability at least $1 - \mathrm{e}^{-n/2} - (\sqrt{2}\mathrm{e}k/p)^k - \mathrm{e}^{-n/8} \to 1$. Since $\boldsymbol{X}_{\overline{S}}'$ is independent of $\boldsymbol{X}_S(\boldsymbol{X}_S'\boldsymbol{X}_S)^{-1}(\boldsymbol{X}_S'(\boldsymbol{y} - \boldsymbol{X}_S\widehat{\boldsymbol{b}}_S) - \boldsymbol{X}_S'\boldsymbol{z})$, Lemma A.6 gives

$$\begin{aligned}&\left\|\boldsymbol{X}_{\overline{S}}'\boldsymbol{X}_S(\boldsymbol{X}_S'\boldsymbol{X}_S)^{-1}(\boldsymbol{X}_S'(\boldsymbol{y} - \boldsymbol{X}_S\widehat{\boldsymbol{b}}_S) - \boldsymbol{X}_S'\boldsymbol{z})\right\|_\infty \\ &\qquad \le \sqrt{\frac{2\log p}{n}}\left\|\boldsymbol{X}_S(\boldsymbol{X}_S'\boldsymbol{X}_S)^{-1}(\boldsymbol{X}_S'(\boldsymbol{y} - \boldsymbol{X}_S\widehat{\boldsymbol{b}}_S) - \boldsymbol{X}_S'\boldsymbol{z})\right\|\end{aligned}$$

with probability approaching one. Combining this with (A.3) gives (A.2).

Let $\widetilde{\boldsymbol{b}}_S$ be the solution to

$$\underset{\boldsymbol{b}\in\mathbb{R}^k}{\text{minimize}}\ \frac{1}{2}\left\|\boldsymbol{\beta}_S + \boldsymbol{X}_S'\boldsymbol{z} - \boldsymbol{b}\right\|^2 + \lambda\|\boldsymbol{b}\|_1.$$

To complete the proof of (2.1), it suffices to establish (i) that for any constant $\delta > 0$,

$$\sup_{\|\boldsymbol{\beta}\|_0\le k}\mathbb{P}\left(\frac{\|\widetilde{\boldsymbol{b}}_S - \boldsymbol{\beta}_S\|^2}{2(1+c)^2k\log p} > 1 - \delta\right) \to 1, \tag{A.4}$$

and (ii)

$$\|\widetilde{\boldsymbol{b}}_S - \widehat{\boldsymbol{b}}_S\| = o_{\mathbb{P}}\left(\|\widetilde{\boldsymbol{b}}_S - \boldsymbol{\beta}_S\|\right) \tag{A.5}$$

since (A.4) and (A.5) give

$$\sup_{\|\boldsymbol{\beta}\|_0 \le k} \mathbb{P}\left(\frac{\|\widehat{\boldsymbol{b}}_S - \boldsymbol{\beta}_S\|^2}{2(1+c)^2 k \log p} > 1 - \delta\right) \to 1 \tag{A.6}$$

for each $\delta > 0$. Note that taking $\delta = 1 - 1/(1+c)^2$ in (A.6) and using the fact that $\widehat{\boldsymbol{b}}_S$ is solution to Lasso with probability approaching one finish the proof

Proof of (A.4). Let $\beta_i = \infty$ if $i \in S$ and otherwise zero (treat ∞ as a sufficiently large positive constant). For each $i \in S$, $\widetilde{b}_{S,i} = \beta_i + \boldsymbol{X}_i'\boldsymbol{z} - \lambda$, and

$$|\widetilde{b}_{S,i} - \beta_i| = |\boldsymbol{X}_i'\boldsymbol{z} - \lambda| \ge \lambda - |\boldsymbol{X}_i'\boldsymbol{z}|.$$

On the event $\{\max_{i\in S} |\boldsymbol{X}_i'\boldsymbol{z}| \le \lambda\}$, which happens with probability tending to one, this inequality gives

$$\begin{aligned}\|\widetilde{\boldsymbol{b}}_S - \boldsymbol{\beta}_S\|^2 \ge \sum_{i\in S}(\lambda - |\boldsymbol{X}_i'\boldsymbol{z}|)^2 &= k\lambda^2 - 2\lambda\sum_{i\in S}|\boldsymbol{X}_i'\boldsymbol{z}| + \sum_{i\in S}(\boldsymbol{X}_i'\boldsymbol{z})^2 \\ &= (1 + o_{\mathbb{P}}(1))2(1+c)^2 k \log p,\end{aligned}$$

where we have used that both $\sum_{i\in S}(\boldsymbol{X}_i'\boldsymbol{z})^2$ and $\sum_{i\in S}|\boldsymbol{X}_i'\boldsymbol{z}|$ are $O_{\mathbb{P}}(k)$. This proves the claim.

Proof of (A.5). Apply Lemma 4.2 with T replaced by S (here each of $\widehat{\boldsymbol{b}}_S, \widetilde{\boldsymbol{b}}_S$ and $\boldsymbol{\beta}$ is supported on S). Since $k/p \to 0$, for any constant $\delta' > 0$, all the singular values of $\boldsymbol{X}_S$ lie in $(1-\delta', 1+\delta')$ with overwhelming probability (see, for example, [5]). Consequently, Lemma 4.2 ensures (A.5).

□

Proof of (2.2). We assume $\sigma = 1$ and put $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathrm{BH}}$. As in the proof of Theorem 1.1, we decompose the total loss as

$$|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq}} - \boldsymbol{\beta}\|^2 = \|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq},S} - \boldsymbol{\beta}_S\|^2 + \|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq},\overline{S}} - \boldsymbol{\beta}_{\overline{S}}\|^2 = \|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq},S} - \boldsymbol{\beta}_S\|^2 + \|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq},\overline{S}}\|^2.$$

The largest possible value of the loss off support is achieved when $\boldsymbol{y}_{\overline{S}}$ is sequentially soft-thresholded by $\boldsymbol{\lambda}^{-[k]}$. Hence, by the proof of Lemma 3.3, we obtain

$$\mathbb{E}\,\|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq},\overline{S}}\|^2 = o\left(2k\log(p/k)\right)$$

for all k-sparse $\boldsymbol{\beta}$.

Now, we turn to consider the loss on support. For any $i \in S$, the loss is at most

$$\left(|z_i| + \lambda_{r(i)}\right)^2 = \lambda_{r(i)}^2 + z_i^2 + 2|z_i|\lambda_{r(i)}.$$

Summing the above equalities over all $i \in S$ gives

$$\mathbb{E}\,\|\widehat{\boldsymbol{\beta}}_{\mathrm{Seq},S} - \boldsymbol{\beta}_S\|^2 \le \sum_{i=1}^{k} \lambda_i^2 + \sum_{i \in S} z_i^2 + 2\sum_{i \in S} |z_i|\lambda_{r(i)}.$$

Note that the first term $\sum_{i=1}^{k} \lambda_i^2 = (1 + o(1))\, 2k\log(p/k)$, and the second term has expectation $\mathbb{E}\sum_{i \in S} z_i^2 = k = o(2k\log(p/k))$, so that it suffices to show that

$$\mathbb{E}\left[2\sum_{i \in S} |z_i|\lambda_{r(i)}\right] = o\left(2k\log(p/k)\right). \tag{A.7}$$

We emphasize that both z_i and $r(i)$ are random so that $\{\lambda_{r(i)}\}_{i \in S}$ and $\{z_i\}_{i \in S}$ may not be independent. Without loss of generality, assume $S = \{1, \ldots, k\}$ and for $1 \le i \le k$, let $r'(i)$ be the rank of the ith observation among the first k. Since $\boldsymbol{\lambda}$ is nonincreasing and $r'(i) \le r(i)$, we have

$$\sum_{1 \le i \le k} |z_i|\lambda_{r(i)} \le \sum_{1 \le i \le k} |z_i|\lambda_{r'(i)} \le \sum_{1 \le i \le k} |z|_{(i)}\lambda_i,$$

where $|z|_{(1)} \ge \cdots \ge |z|_{(k)}$ are the order statistics of $z_1, \ldots, z_k$. The second inequality follows from the fact that for any nonnegative sequences $\{a_i\}$ and $\{b_i\}$, $\sum_i a_i b_i \le \sum_i a_{(i)} b_{(i)}$. Therefore, letting $\zeta_1, \ldots, \zeta_k$ be i.i.d. $\mathcal{N}(0,1)$, (A.7) follows from the estimate

$$\sum_{i=1}^{k} \lambda_i\, \mathbb{E}\,|\zeta|_{(i)} = o\left(2k\log(p/k)\right). \tag{A.8}$$

To argue about (A.8), we work with the approximations $\lambda_i \sim \sqrt{2\log(p/i)}$ and $\mathbb{E}\,|\zeta|_{(i)} = O\left(\sqrt{2\log(2k/i)}\right)$ (see e.g. (A.15)), so that the claim is a consequence of

$$\sum_{i=1}^{k} \sqrt{\log\frac{p}{i}\log\frac{2k}{i}} = o\left(2k\log(p/k)\right),$$

which is justified as follows:

$$\begin{aligned}\sum_{i=1}^{k}\sqrt{\log\frac{p}{i}\log\frac{2k}{i}} &\le k\int_0^1\sqrt{\log\frac{p/k}{x}\log\frac{2}{x}}\mathrm{d}x\\ &\le k\int_0^1\sqrt{\log\frac{p}{k}}\sqrt{\log\frac{2}{x}}+\frac{\log\frac{1}{x}\sqrt{\log\frac{2}{x}}}{2\sqrt{\log(p/k)}}\mathrm{d}x\\ &= C_1k\sqrt{\log\frac{p}{k}}+\frac{C_2k}{\sqrt{\log(p/k)}}\end{aligned}$$

for some absolute constants C_1, C_2. Since $\log(p/k)\to\infty$, it is clear that the right-hand side of the above display is of $o(2k\log(p/k))$. □

A.2. Proofs for Section 3. To begin with, we derive a dual formulation of the SLOPE program (1.6), which provides a nice geometrical interpretation. This dual formulation will also be used in the proof of Lemma 4.3. Our exposition largely borrows from [2].

Rewrite (1.6) as

$$\underset{\boldsymbol{b},\boldsymbol{r}}{\text{minimize}}\quad \frac{1}{2}\|\boldsymbol{r}\|^2+\sum_i\lambda_i|b|_{(i)}\quad\text{subject to}\quad \boldsymbol{X}\boldsymbol{b}+\boldsymbol{r}=\boldsymbol{y}, \tag{A.9}$$

whose Lagrangian is

$$\mathcal{L}(\boldsymbol{b},\boldsymbol{r},\boldsymbol{\nu}) := \frac{1}{2}\|\boldsymbol{r}\|^2+\sum_i\lambda_i|b|_{(i)}-\boldsymbol{\nu}'(\boldsymbol{X}\boldsymbol{b}+\boldsymbol{r}-\boldsymbol{y}).$$

Hence, the dual objective is given by

$$\begin{aligned}\inf_{\boldsymbol{b},\boldsymbol{r}}\ \mathcal{L}(\boldsymbol{b},\boldsymbol{r},\boldsymbol{\nu}) &= \boldsymbol{\nu}'\boldsymbol{y}-\sup_{\boldsymbol{r}}\left\{\boldsymbol{\nu}'\boldsymbol{r}-\frac{1}{2}\|\boldsymbol{r}\|^2\right\}-\sup_{\boldsymbol{b}}\left\{(\boldsymbol{X}'\boldsymbol{\nu})'\boldsymbol{b}-\sum_i\lambda_i|b|_{(i)}\right\}\\ &= \boldsymbol{\nu}'\boldsymbol{y}-\frac{1}{2}\|\boldsymbol{\nu}\|^2-\begin{cases}0 & \boldsymbol{\nu}\in C_{\boldsymbol{\lambda},\boldsymbol{X}}\\ +\infty & \text{otherwise},\end{cases}\end{aligned}$$

where $C_{\boldsymbol{\lambda},\boldsymbol{X}} := \{\boldsymbol{\nu} : \boldsymbol{X}'\boldsymbol{\nu} \text{ is majorized by } \boldsymbol{\lambda}\}$ is a (convex) polytope. It thus follows that the dual reads

$$\underset{\boldsymbol{\nu}}{\text{maximize}}\quad \boldsymbol{\nu}'\boldsymbol{y}-\frac{1}{2}\|\boldsymbol{\nu}\|^2\quad\text{subject to } \boldsymbol{\nu}\in C_{\boldsymbol{\lambda},\boldsymbol{X}}. \tag{A.10}$$

The equality $\boldsymbol{\nu}'\boldsymbol{y}-\|\boldsymbol{\nu}\|^2/2 = -\|\boldsymbol{y}-\boldsymbol{\nu}\|^2/2+\|\boldsymbol{y}\|^2/2$ reveals that the dual solution $\widehat{\boldsymbol{\nu}}$ is indeed the projection of $\boldsymbol{y}$ onto $C_{\boldsymbol{\lambda},\boldsymbol{X}}$. The minimization of the Lagrangian over $\boldsymbol{r}$ is attained at $\boldsymbol{r} = \boldsymbol{\nu}$. This implies that the primal solution $\widehat{\boldsymbol{\beta}}$ and the dual solution $\widehat{\boldsymbol{\nu}}$ obey

$$\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\nu}}. \tag{A.11}$$

We turn to proving the facts.

Proof of Fact 3.1. Without loss of generality, suppose both $\boldsymbol{a}$ and $\boldsymbol{b}$ are nonnegative and arranged in nonincreasing order. Denote by $T_k^{\boldsymbol{a}}$ the sum of the first k terms of $\boldsymbol{a}$ with $T_0^{\boldsymbol{a}} \triangleq 0$, and similarly for $\boldsymbol{b}$. We have

$$\|\boldsymbol{a}\|^2 = \sum_{k=1}^{p} a_k(T_k^{\boldsymbol{a}} - T_{k-1}^{\boldsymbol{a}}) = \sum_{k=1}^{p-1} T_k^{\boldsymbol{a}}(a_k - a_{k+1}) + a_p T_p^{\boldsymbol{a}} \geq \sum_{k=1}^{p-1} T_k^{\boldsymbol{b}}(a_k - a_{k+1}) + a_p T_p^{\boldsymbol{b}} = \sum_{k=1}^{p} a_k b_k.$$

Similarly,

$$\|\boldsymbol{b}\|^2 = \sum_{k=1}^{p-1} T_k^{\boldsymbol{b}}(b_k - b_{k+1}) + b_p T_p^{\boldsymbol{b}} \leq \sum_{k=1}^{p-1} T_k^{\boldsymbol{a}}(b_k - b_{k+1}) + b_p T_p^{\boldsymbol{a}} = \sum_{k=1}^{p} b_k(T_k^{\boldsymbol{a}} - T_{k-1}^{\boldsymbol{a}}) = \sum_{k=1}^{p} a_k b_k,$$

which proves the claim. $\square$

Proofs of Facts 3.2 and 3.3. Taking $\boldsymbol{X} = \boldsymbol{I}_p$ in the dual formulation, (A.11) immediately implies that $\boldsymbol{a} - \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})$ is the projection of $\boldsymbol{a}$ onto the polytope $C_{\boldsymbol{\lambda}, \boldsymbol{I}_p}$. By definition, $C_{\boldsymbol{\lambda}, \boldsymbol{I}_p}$ consists of all vectors majorized by $\boldsymbol{\lambda}$. Hence, $\boldsymbol{a} - \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})$ is always majorized by $\boldsymbol{\lambda}$. In particular, if $\boldsymbol{a}$ is majorized by $\boldsymbol{\lambda}$, then the projection $\boldsymbol{a} - \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})$ of $\boldsymbol{a}$ is identical to $\boldsymbol{a}$ itself. This gives $\operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a}) = \boldsymbol{0}$. $\square$

Proof of Fact 3.4. Assume $\boldsymbol{a}$ is nonnegative without loss of generality. It is intuitively obvious that

$$\boldsymbol{b} \geq \boldsymbol{a} \quad \Longrightarrow \quad \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{b}) \geq \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a}),$$

where as usual $\boldsymbol{b} \geq \boldsymbol{a}$ means that $\boldsymbol{b} - \boldsymbol{a} \in \mathbb{R}_+^p$. In other words, if the observations increase, the fitted values do not decrease. To save time, we directly verify this claim by using Algorithm 3 (FastProxSL1) from [2]. By the averaging step of that algorithm, we can see that for each $1 \leq i, j \leq p$,

$$\frac{\partial\,[\operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})]_i}{\partial a_j} = \begin{cases} \frac{1}{\#\{1 \leq k \leq p\colon\ [\operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})]_k = [\operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})]_j\}}, & \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})_j = \operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})_i > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This holds for all $\boldsymbol{a} \in \mathbb{R}^p$ except for a set of measure zero. The nonnegativity of $\partial\,[\operatorname{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})]_i / \partial a_j$ along with the Lipschitz continuity of the prox imply the monotonicity property. A consequence is that $\| [\mathrm{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})]_{\overline{T}} \|$ does not decrease as we let $a_i \to \infty$ for all $i \in T$. In the limit, $\| [\mathrm{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})]_{\overline{T}} \|$ monotonically converges to $\| \mathrm{prox}_{\boldsymbol{\lambda}^{-|T|}}(\boldsymbol{a}_{\overline{T}}) \|$. This gives the desired inequality. ☐

As a remark, we point out that the proofs of Facts 3.2 and 3.3 suggest a very simple proof of Lemma 3.1. Since $\boldsymbol{a} - \mathrm{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a})$ is the projection of $\boldsymbol{a}$ onto $C_{\boldsymbol{\lambda},\boldsymbol{I}_p}$, $\| \mathrm{prox}_{\boldsymbol{\lambda}}(\boldsymbol{a}) \|$ is thus the distance between $\boldsymbol{a}$ and the polytope $C_{\boldsymbol{\lambda},\boldsymbol{I}_p}$. Hence, it suffices to find a point in the polytope at a distance of $\|(|\boldsymbol{a}| - \boldsymbol{\lambda})_+\|$ away from $\boldsymbol{a}$. The point $\boldsymbol{b}$ defined as $b_i = \min\{|a_i|, \lambda_i\}$ does the job.

Now, we proceed to prove the preparatory lemmas for Theorem 1.1, namely, Lemmas A.3 and A.4. The first two lemmas below can be found in [1].

LEMMA A.1. *Let U be a* $\mathrm{Beta}(a,b)$ *random variable. Then*

$$\mathbb{E}\log U = (\log\Gamma(a))' - (\log\Gamma(a+b))',$$

where Γ denotes the Gamma function and $(\log\Gamma(x))'$ is the derivative with respect to x.

LEMMA A.2. *For any integer $m \geq 1$,*

$$(\log\Gamma(m))' = -\gamma + \sum_{j=1}^{m-1}\frac{1}{j} = \log m + O\Big(\frac{1}{m}\Big),$$

where $\gamma = 0.577215\cdots$ is the Euler constant.

LEMMA A.3. *Let $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{p-k})$. Under the assumptions of Theorem 1.1, for any constant $A > 0$,*

$$\frac{1}{2k\log(p/k)}\sum_{i=1}^{\lfloor Ak \rfloor}\mathbb{E}\left(|\zeta|_{(i)} - \lambda^{\mathrm{BH}}_{k+i}\right)_+^2 \to 0.$$

PROOF OF LEMMA A.3. Write $\lambda_i = \lambda_i^{\mathrm{BH}}$ for simplicity. It is sufficient to prove a stronger version in which the order statistics $|\zeta|_{(i)}$ come from p i.i.d. $\mathcal{N}(0,1)$. The reason is that the order statistics will be stochastically larger, thus enlarging $\mathbb{E}\left(|\zeta|_{(i)} - \lambda^{\mathrm{BH}}_{k+i}\right)_+^2$, since $(\zeta - \lambda)_+^2$ is nondecreasing in ζ. Applying the bias-variance decomposition, we get

(A.12)
$$\mathbb{E}\left(|\zeta|_{(i)} - \lambda_{k+i}\right)_+^2 \leq \mathbb{E}\left(|\zeta|_{(i)} - \lambda_{k+i}\right)^2 = \mathrm{Var}(|\zeta_{(i)}|) + \left(\mathbb{E}\,|\zeta_{(i)}| - \lambda_{k+i}\right)^2.$$

We proceed to control each term separately.

For the variance, a direct application of Proposition 4.2 in [3] gives

$$\operatorname{Var}(|\zeta_{(i)}|) = O\Big(\frac{1}{i\log(p/i)}\Big) \tag{A.13}$$

for all $i \le p/2$. Hence,

$$\sum_{i=1}^{\lfloor Ak \rfloor} \operatorname{Var}(|\zeta_{(i)}|) = O\left(\sum_{i=1}^{\lfloor Ak \rfloor} \frac{1}{i\log(p/i)}\right) = o(2k\log(p/k)),$$

where the last step makes use of $\log(p/k) \to \infty$. It remains to show that

$$\sum_{i=1}^{\lfloor Ak \rfloor} \big(\mathbb{E}\,|\zeta_{(i)}| - \lambda_{k+i}\big)^2 = o(2k\log(p/k)). \tag{A.14}$$

Let $U_1, \ldots, U_p$ be i.i.d. uniform random variables on $(0,1)$ and $U_{(i)}$ be the i^{th} smallest—please note that for a change, the U_i's are sorted in increasing order. We know that $U_{(i)}$ is distributed as $\operatorname{Beta}(i, p+1-i)$ and that $|\zeta|_{(i)}$ has the same distribution as $\Phi^{-1}(1 - U_{(i)}/2)$. Making use of Lemmas A.1 and A.2 then gives

$$\mathbb{E}\,|\zeta|_{(i)}^2 = \mathbb{E}\left[\Phi^{-1}(1-U_{(i)}/2)^2\right] \sim \mathbb{E}\left[2\log(2/U_{(i)})\right] = 2\log 2 + 2\sum_{j=i}^{p}\frac{1}{j} = (1+o(1))2\log(p/i),$$

where the second step follows from $(1+o_{\mathbb{P}}(1))2\log(2/U_{(i)}) \le \Phi^{-1}(1 - U_{(i)}/2)^2 \le 2\log(2/U_{(i)})$ for $i = o(p)$. As a result,

$$\begin{aligned} \mathbb{E}\,|\zeta_{(i)}| &\le \sqrt{\mathbb{E}\,|\zeta|_{(i)}^2} = (1+o(1))\sqrt{2\log(p/i)} \\ \mathbb{E}\,|\zeta_{(i)}| &= \sqrt{\mathbb{E}\,|\zeta|_{(i)}^2 - \operatorname{Var}(|\zeta|_{(i)})} = (1+o(1))\sqrt{2\log(p/i)}. \end{aligned} \tag{A.15}$$

Similarly, since $k + i = o(p)$ and q is constant, we have the approximation

$$\lambda_{k+i} = (1+o(1))\sqrt{2\log(p/(k+i))},$$

which together with (A.15) reveals that

$$\big(\mathbb{E}\,|\zeta_{(i)}| - \lambda_{k+i}\big)^2 \le (1+o(1))\,2\left[\sqrt{\log(p/i)} - \sqrt{\log(p/(k+i))}\right]^2 + o(1)\log(p/i). \tag{A.16}$$

The second term in the right-hand side contributes at most $o(1)\,Ak\log(p/(Ak)) = o(1)\,2k\log(p/k)$ in the sum (A.14). For the first term, we get

$$\left[\sqrt{\log(p/i)} - \sqrt{\log(p/(k+i))}\right]^2 = \frac{\log^2(1+k/i)}{\left[\sqrt{\log(p/i)} + \sqrt{\log(p/(k+i))}\right]^2} = o(1)\log^2(1+k/i).$$

Hence, it contributes at most

(A.17)
$$\begin{aligned} o(1)\sum_{i=1}^{\lfloor Ak\rfloor}\log^2(1+k/i) &\le o(1)\sum_{i=1}^{\lfloor Ak\rfloor} k\int_{\frac{i-1}{k}}^{\frac{i}{k}}\log^2(1+1/x)\mathrm{d}x \\ &= o(1)k\int_0^A \log^2(1+1/x)\mathrm{d}x = o(2k\log(p/k)). \end{aligned}$$

Combining (A.17), (A.16) and (A.14) concludes the proof. ☐

LEMMA A.4. *Let $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{p-k})$ and $A > 0$ be any constant satisfying $q(1+A)/A < 1$. Then, under the assumptions of Theorem 1.1,*

$$\frac{1}{2k\log(p/k)}\sum_{i=\lceil Ak\rceil}^{p-k}\mathbb{E}\left(|\zeta|_{(i)} - \lambda^{\mathrm{BH}}_{k+i}\right)_+^2 \to 0.$$

PROOF OF LEMMA A.4. Again, write $\lambda_i = \lambda_i^{\mathrm{BH}}$ for simplicity. As in the proof of Lemma A.3 we work on a stronger version by assuming $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p)$. Denote by $q' = q(1+A)/A$. For any $u \ge 0$, let $\alpha_u := \mathbb{P}(|\mathcal{N}(0,1)| > \lambda_{k+i} + u) = 2\Phi(-\lambda_{k+i} - u)$. Then $\mathbb{P}(|\zeta|_{(i)} > \lambda_{k+i} + u)$ is just the tail probability of the binomial distribution with p trials and success probability α_u. By the Chernoff bound, this probability is bounded as

$$\mathbb{P}(|\zeta|_{(i)} > \lambda_{k+i} + u) \le \mathrm{e}^{-p\,\mathrm{KL}(i/p\|\alpha_u)}, \tag{A.18}$$

where $\mathrm{KL}(a\|b) := a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$ is the Kullback-Leibler divergence. Note that

$$\frac{\partial\,\mathrm{KL}(i/p\|b)}{\partial b} = -\frac{i/p}{b} + \frac{1-i/p}{1-b} \le -\frac{i}{pb} + 1 \tag{A.19}$$

for all $0 < b < i/p$. Hence, from (A.19) it follows that

(A.20)
$$\begin{aligned} \mathrm{KL}\left(i/p\|\alpha_u\right) - \mathrm{KL}\left(i/p\|\alpha_0\right) = -\int_{\alpha_u}^{\alpha_0}\frac{\partial\,\mathrm{KL}}{\partial b}\mathrm{d}b &\ge \int_{\alpha_u}^{\alpha_0}\frac{i}{pb} - 1\mathrm{d}b \\ &\ge \int_{\mathrm{e}^{-u\lambda_{k+i}}\alpha_0}^{\alpha_0}\frac{i}{pb} - 1\mathrm{d}b \\ &= \frac{iu\lambda_{k+i}}{p} - \alpha_0\left(1 - \mathrm{e}^{-u\lambda_{k+i}}\right), \end{aligned}$$

where the second inequality makes use of $\alpha_u \le \mathrm{e}^{-u\lambda_{k+i}}\alpha_0$. With the proviso that $q(1+A)/A < 1$ and $i \ge Ak$, it follows that

$$\alpha_0 = q(k+i)/p \le q' i/p. \tag{A.21}$$

Hence, substituting (A.21) into (A.20), we see that (A.18) yields

$$\begin{aligned}
\mathbb{P}(|\zeta|_{(i)} > \lambda_{k+i} + u) &\le \mathrm{e}^{-p\left(\mathrm{KL}(\frac{i}{p}\|\alpha_u) - \mathrm{KL}(\frac{i}{p}\|\alpha_0)\right)} \mathrm{e}^{-p\,\mathrm{KL}(\frac{i}{p}\|\alpha_0)} \\
&\le \mathrm{e}^{-p\left(\mathrm{KL}(\frac{i}{p}\|\alpha_u) - \mathrm{KL}(\frac{i}{p}\|\alpha_0)\right)} \\
&\le \exp\left(-iu\lambda_{k+i} + q' i \big(1 - \exp\big(-u\lambda_{k+i}\big)\big)\right).
\end{aligned} \tag{A.22}$$

With this preparation, we conlude the proof of our lemma as follows:

$$\begin{aligned}
\mathbb{E}\left(|\zeta|_{(i)} - \lambda_{k+i}\right)_+^2 &= \int_0^\infty \mathbb{P}\left((|\zeta|_{(i)} - \lambda_{k+i})_+^2 > x\right) \mathrm{d}x \\
&= \int_0^\infty \mathbb{P}(|\zeta|_{(i)} > \lambda_{k+i} + \sqrt{x}) \mathrm{d}x \\
&= 2\int_0^\infty u \mathbb{P}(|\zeta|_{(i)} > \lambda_{k+i} + u) \mathrm{d}u,
\end{aligned}$$

and plugging (A.22) gives

$$\begin{aligned}
\mathbb{E}\left(|\zeta|_{(i)} - \lambda_{k+i}\right)_+^2 &\le 2\int_0^\infty u \exp\Big(-iu\lambda_{k+i} + q' i \big(1 - \exp\big(-u\lambda_{k+i}\big)\big)\Big) \mathrm{d}u \\
&= \frac{2}{\lambda_{k+i}^2} \int_0^\infty x \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))i} \mathrm{d}x \\
&\le \frac{2}{\lambda_p^2} \int_0^\infty x \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))i} \mathrm{d}x.
\end{aligned}$$

This yields the upper bound

$$\begin{aligned}
\sum_{i=\lceil Ak \rceil}^{p-k} \mathbb{E}\left(|\zeta|_{(i)} - \lambda_{k+i}\right)_+^2 &\le \frac{2}{\lambda_p^2} \sum_{i=\lceil Ak \rceil}^{p-k} \int_0^\infty x \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))i} \mathrm{d}x \\
&\le \frac{2}{\Phi^{-1}(1-q/2)^2} \sum_{i=1}^{\infty} \int_0^\infty x \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))i} \mathrm{d}x \\
&= \frac{2}{\Phi^{-1}(1-q/2)^2} \int_0^\infty \frac{x \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))}}{1 - \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))}} \mathrm{d}x.
\end{aligned}$$

Since the integrand obeys

$$\lim_{x \to 0} \frac{x \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))}}{1 - \mathrm{e}^{-(x - q'(1-\mathrm{e}^{-x}))}} = \frac{1}{1-q'}$$

and decays exponentially fast as $x \to \infty$, we conclude that $\sum_{i=\lceil Ak \rceil}^{p-k} \mathbb{E}(|\zeta|_{(i)} - \lambda_{k+i})_+^2$ is bounded by a constant. This is a bit more than we need since $2k \log(p/k) \to \infty$. $\square$

A.3. Proofs for Section 4. In this paper, we often use the Borell inequality to show that $\mathbb{P}(\|\mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)\| > \sqrt{n} + t) \leq \exp(-t^2/2)$.

LEMMA A.5 (Borell's inequality). *Let $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)$ and f be an L-Lipschitz continuous function in $\mathbb{R}^n$. Then*

$$\mathbb{P}\left(f(\boldsymbol{\zeta}) > \mathbb{E}\, f(\boldsymbol{\zeta}) + t\right) \leq \mathrm{e}^{-\frac{t^2}{2L^2}}$$

for every $t > 0$.

LEMMA A.6. *Let $\zeta_1, \ldots, \zeta_p$ be i.i.d. $\mathcal{N}(0,1)$. Then*

$$\max_i |\zeta_i| \leq \sqrt{2 \log p}$$

holds with probability approaching one.

The latter classical result can be proved in many different ways. Suffices to say that it follows from a more subtle fact, namely, that

$$\sqrt{2\log p}\left(\max_i \zeta_i - \sqrt{2\log p} + \frac{\log\log p + \log 4\pi}{2\sqrt{2\log p}}\right)$$

converges weakly to a Gumbel distribution [4].

PROOF OF LEMMA 4.3. Let $\widehat{\boldsymbol{\beta}}^{\mathrm{lift}}$ be the lift of $\widehat{\boldsymbol{b}}_T$ in the sense that $\widehat{\boldsymbol{\beta}}^{\mathrm{lift}}_T = \widehat{\boldsymbol{b}}_T$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{lift}}_{\overline{T}} = \mathbf{0}$ and let $|T| = m$. Further, set $\widetilde{\boldsymbol{\nu}} := \boldsymbol{y} - \boldsymbol{X}_T \widehat{\boldsymbol{b}}_T = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{lift}}$. Applying (A.10) and (A.11) to the reduced SLOPE program, we get that

$$\boldsymbol{X}_T' \widetilde{\boldsymbol{\nu}} \preceq \boldsymbol{\lambda}^{[m]}.$$

By the assumption, $\boldsymbol{X}_{\overline{T}}' \widetilde{\boldsymbol{\nu}}$ is majorized by $\boldsymbol{\lambda}^{-[m]}$. Hence, $\boldsymbol{X}'\widetilde{\boldsymbol{\nu}}$—the concatenation of $\boldsymbol{X}_T' \widetilde{\boldsymbol{\nu}}$ and $\boldsymbol{X}_{\overline{T}}' \widetilde{\boldsymbol{\nu}}$—is majorized by $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{[m]}, \boldsymbol{\lambda}^{-[m]})$. This confirms that $\widetilde{\boldsymbol{\nu}}$ is dual feasible with respect to the full SLOPE program. If additionally we show that

$$\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{lift}}\|^2 + \sum_i \lambda_i |\widehat{\beta}^{\mathrm{lift}}|_{(i)} = \widetilde{\boldsymbol{\nu}}'\boldsymbol{y} - \frac{1}{2}\|\widetilde{\boldsymbol{\nu}}\|^2, \tag{A.23}$$

then the strong duality claims that $\widehat{\boldsymbol{\beta}}^{\text{lift}}$ and $\widetilde{\boldsymbol{\nu}}$ must, respectively, be the optimal solutions to the full primal and dual.

In fact, (A.23) is self-evident. The right-hand side is the optimal value of the reduced dual (i.e., replacing $\boldsymbol{X}$ and $\boldsymbol{\lambda}$ by $\boldsymbol{X}_T$ and $\boldsymbol{\lambda}^{[m]}$ in (A.10)), while the left-hand side agrees with the optimal value of the reduced primal since

$$\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\text{lift}}\|^2=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{b}}_T\|^2 \text{ and } \sum_{i=1}^{p}\lambda_i|\widehat{\beta}^{\text{lift}}|_{(i)}=\sum_{i=1}^{m}\lambda_i|\widehat{b}_T|_{(i)}.$$

Since the reduced primal only has linear equality constraints and is clearly feasible, strong duality holds, and (A.23) follows from this. $\square$

LEMMA A.7. *Let $1\le k^\star<p$ be any (deterministic) integer, then*

$$\sup_{|T|=k^\star}\|\boldsymbol{X}_T'\boldsymbol{z}\|\le\sqrt{32k^\star\log(p/k^\star)}$$

with probability at least $1-\mathrm{e}^{-n/2}-(\sqrt{2}\mathrm{e}k^\star/p)^{k^\star}$. Above, the supremum is taken over all the subsets of $\{1,\ldots,p\}$ with cardinality $k^\star$.

PROOF OF LEMMA A.7. Conditional on $\boldsymbol{z}$, it is easy to see that $\boldsymbol{X}'\boldsymbol{z}$ is distributed as i.i.d. centered Gaussian random variables with variance $\|\boldsymbol{z}\|^2/n$. This observation enables us to write

$$\boldsymbol{X}'\boldsymbol{z}\overset{d}{=}\frac{\|\boldsymbol{z}\|}{\sqrt{n}}(\zeta_1,\ldots,\zeta_p),$$

where $\boldsymbol{\zeta}:=(\zeta_1,\ldots,\zeta_p)$ consists of i.i.d. $\mathcal{N}(0,1)$ independent of $\|\boldsymbol{z}\|$. Hence, it is sufficient to prove that

$$\|\boldsymbol{z}\|\le 2\sqrt{n},\quad |\zeta|_{(1)}^2+\cdots+|\zeta|_{(k^\star)}^2\le 8k^\star\log(p/k^\star)$$

simultaneously with probability at least $1-\mathrm{e}^{-n/2}-\left(\sqrt{2}\mathrm{e}k^\star/p\right)^{k^\star}$. From Lemma A.5, we know that $\mathbb{P}(\|\boldsymbol{z}\|>2\sqrt{n})\le\mathrm{e}^{-n/2}$ so we just need to establish the other inequality. To this end, observe that

$$\begin{aligned}\mathbb{P}\left(|\zeta|_{(1)}^2+\cdots+|\zeta|_{(k^\star)}^2>8k^\star\log(p/k^\star)\right)&\le\frac{\mathbb{E}\,\mathrm{e}^{\frac{1}{4}\left(|\zeta|_{(1)}^2+\cdots+|\zeta|_{(k^\star)}^2\right)}}{\mathrm{e}^{2k^\star\log\frac{p}{k^\star}}}\\&\le\frac{\sum_{i_1<\cdots<i_{k^\star}}\mathbb{E}\,\mathrm{e}^{\frac{1}{4}\left(|\zeta|_{i_1}^2+\cdots+|\zeta|_{i_{k^\star}}^2\right)}}{\mathrm{e}^{2k^\star\log\frac{p}{k^\star}}}\\&=\frac{\binom{p}{k^\star}2^{k^\star/2}}{\mathrm{e}^{2k^\star\log\frac{p}{k^\star}}}\\&\le\left(\frac{\sqrt{2}\mathrm{e}k^\star}{p}\right)^{k^\star}.\end{aligned}$$

□

We record an elementary result which simply follows from $\Phi^{-1}(1-c/2) \le \sqrt{2\log 1/c}$ for each $0 < c < 1$.

Lemma A.8. *Fix $0 < q < 1$. Then for all $1 \le k \le p/2$,*

$$\sum_{i=1}^{k} (\lambda_i^{\mathrm{BH}})^2 \le C_q \cdot k \log(p/k),$$

for some constant $C_q > 0$.

In the next two lemmas, we use the BHq critical values $\boldsymbol{\lambda}^{\mathrm{BH}}$ to majorize sequences of Gaussian order statistics. Again, $\boldsymbol{a} \preceq \boldsymbol{b}$ means that $\boldsymbol{b}$ majorizes $\boldsymbol{a}$.

Lemma A.9. *Given any constant $c > 1/(1-q)$, suppose $\max\{ck, k+d\} \le k^\star < p$ for any (deterministic) sequence d that diverges to ∞. Let $\zeta_1, \ldots, \zeta_{p-k}$ be i.i.d. $\mathcal{N}(0,1)$. Then*

$$\left(|\zeta|_{(k^\star-k+1)}, |\zeta|_{(k^\star-k+2)}, \ldots, |\zeta|_{(p-k)}\right) \preceq \left(\lambda^{\mathrm{BH}}_{k^\star+1}, \lambda^{\mathrm{BH}}_{k^\star+2}, \ldots, \lambda^{\mathrm{BH}}_{p}\right)$$

with probability approaching one.

Proof of Lemma A.9. It suffices to prove the stronger case where $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_p)$. Let $U_1, \ldots, U_p$ be i.i.d. uniform random variables on $[0,1]$ and $U_{(1)} \le \cdots \le U_{(p)}$ the corresponding order statistics. Since

$$(|\zeta|_{(k^\star-k+1)}, \ldots, |\zeta|_{(p-k)}) \overset{d}{=} \left(\Phi^{-1}(1-U_{(k^\star-k+1)}/2), \ldots, \Phi^{-1}(1-U_{(p-k)}/2)\right),$$

the conclusion would follow from

$$\mathbb{P}\left(U_{(k^\star-k+j)} \ge q(k^\star+j)/p,\ \forall j \in \{1, \ldots, p-k^\star\}\right) \to 1.$$

Let $E_1, \ldots, E_{p+1}$ be i.i.d. exponential random variables with mean 1 and denote by $T_i = E_1 + \cdots + E_i$. Then the order statistics $U_{(i)}$ have the same joint distribution with T_i/T_{p+1}. Fixing an arbitrary constant $q' \in (q, 1-1/c)$, we have

$$\mathbb{P}\left(U_{(k^\star-k+j)} \ge q(k^\star+j)/p, \forall j\right) \ge \mathbb{P}\left(T_{k^\star-k+j} \ge q'(k^\star+j), \forall j\right) - \mathbb{P}\left(T_{p+1} > q'p/q\right).$$

Since $\mathbb{P}(T_{p+1} > q'p/q) \to 0$ by the law of large numbers, it is sufficient to prove

$$\mathbb{P}\left(T_{k^\star-k+j} \ge q'(k^\star+j), \forall j \in \{1, \ldots, p-k^\star\}\right) \to 1. \tag{A.24}$$

This event can be rewritten as

$$T_{k^\star-k+j} - T_{k^\star-k} - q'j \geq q'k^\star - T_{k^\star-k}$$

for all $1 \leq j \leq p - k^\star$. Hence, (A.24) is reduced to proving

$$\mathbb{P}\left(\min_{1\leq j\leq p-k^\star} T_{k^\star-k+j} - T_{k^\star-k} - q'j \geq q'k^\star - T_{k^\star-k}\right) \to 1. \tag{A.25}$$

As a random walk, $T_{k^\star-k+j} - T_{k^\star-k} - q'j$ has i.i.d. increments with mean $1-q' > 0$ and variance 1. Thus $\min_{1\leq j\leq p-k^\star} T_{k^\star-k+j} - T_{k^\star-k} - q'j$ converges weakly to a bounded random variable in distribution. Consequently, (A.25) holds if one can demonstrate that $q'k^\star - T_{k^\star-k}$ diverges to $-\infty$ as $p \to \infty$ in probability. To see this, observe that

$$q'k^\star - T_{k^\star-k} = \frac{q'k^\star}{k^\star - k}(k^\star - k) - T_{k^\star-k} \leq \frac{q'c}{c-1}(k^\star - k) - T_{k^\star-k},$$

where we use the fact $k^\star \geq ck$. Under our hypothesis $q'c/(c-1) < 1$, the process $\{q'ct/(c-1) - T_t : t \in \mathbb{N}\}$ is a random walk drifting towards $-\infty$. Recognizing that $k^\star - k \geq d \to \infty$, we see that $q'c(k^\star - k)/(c-1) - T_{k^\star-k}$ (weakly) diverges to $-\infty$ since it corresponds to a position of the preceding random walk at $t \to \infty$. This concludes the proof.

□

LEMMA A.10. *Let $\zeta_1, \ldots, \zeta_{p-k}$ be i.i.d. $\mathcal{N}(0,1)$. Then there exists a constant C_q only depending on q such that*

$$(\zeta_1, \ldots, \zeta_{p-k}) \preceq C_q \cdot \sqrt{\frac{\log p}{\log(p/k)}}\left(\lambda^{\mathrm{BH}}_{k+1}, \ldots, \lambda^{\mathrm{BH}}_p\right)$$

with probability tending to one as $p \to \infty$ and $k/p \to 0$.

PROOF OF LEMMA A.10. Let $U_1, \ldots, U_{p-k}$ be i.i.d. uniform random variables on $[0,1]$ and replace ζ_i by $\Phi^{-1}(1 - U_i/2)$. Note that

$$\Phi^{-1}(1 - U_i/2) \leq \sqrt{2\log\frac{2}{U_i}}, \quad \lambda^{\mathrm{BH}}_{k+i} \asymp \sqrt{2\log\frac{2p}{k+i}};$$

Hence, it suffices to prove that for some constant κ'_q,

$$\log(2/U_{(i)})\log(p/k) \leq \kappa'_q \cdot \log p \cdot \log(2p/(k+i)) \tag{A.26}$$

holds for all $i = 1, \ldots, p-k$ with probability approaching one. Applying the representation given in the proof of Lemma A.9 and noting that $T_{p+1} = (1 + o_{\mathbb{P}}(1))p$, we see that (A.26) is implied by

$$\log(3p/T_i)\log(p/k) \le \kappa_q' \cdot \log p \cdot \log(2p/(k+i)). \tag{A.27}$$

We consider $i \le 4\sqrt{p}$ and $i > 4\sqrt{p}$ separately.

Suppose first that $i \le 4\sqrt{p}$. In this case,

$$\log(2p/(k+i)) = (1 + o(1))\log(p/k).$$

Thus (A.27) would follow from

$$\log(3p/T_i) = O(\log p)$$

for all such i. This is, however, self-evident since $T_i \ge E_1 \ge 1/p$ with probability $1 - \mathrm{e}^{-1/p} = o(1)$.

Suppose now that $i > 4\sqrt{p}$. In this case, we make use of the fact that $T_i > i/2 - \sqrt{p}$ for all i with probability tending to one as $p \to \infty$. Then we prove a stronger result, namely,

$$\log \frac{3p}{i/2 - \sqrt{p}} \cdot \log \frac{p}{k} \le \kappa_q' \log p \cdot \log \frac{2p}{k+i}.$$

for all $i > 4\sqrt{p}$. This follows from the two observations below:

$$\log \frac{3p}{i/2 - \sqrt{p}} \asymp \log \frac{p}{i}, \ \ \log \frac{2p}{k+i} \ge \min\left\{\log \frac{p}{i}, \log \frac{p}{k}\right\}.$$

□

In the proofs of the next two lemmas, namely, Lemma A.11 and Lemma A.12, we introduce an orthogonal matrix $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ that obeys

$$\boldsymbol{Q}\boldsymbol{z} = (\|\boldsymbol{z}\|, 0, \ldots, 0).$$

In the proofs, $\boldsymbol{Q}$ is further set to be measurable with respect to $\boldsymbol{z}$. Hence, $\boldsymbol{Q}$ is independent of $\boldsymbol{X}$. There are many options available to construct such a $\boldsymbol{Q}$, including the Householder transformation. Set

$$\boldsymbol{W} = \begin{bmatrix} \widetilde{\boldsymbol{w}} \\ \widetilde{\boldsymbol{W}} \end{bmatrix} := \boldsymbol{Q}\boldsymbol{X},$$

where $\widetilde{\boldsymbol{w}} \in \mathbb{R}^{1\times p}$ and $\widetilde{\boldsymbol{W}} \in \mathbb{R}^{(n-1)\times p}$. The independence between $\boldsymbol{Q}$ and $\boldsymbol{X}$ suggests that $\boldsymbol{W}$ is still a Gaussian random matrix, consisting of i.i.d. $\mathcal{N}(0, 1/n)$ entries. Note that

$$\boldsymbol{X}_i' \boldsymbol{z} = (\boldsymbol{Q}\boldsymbol{X}_i)'(\boldsymbol{Q}\boldsymbol{z}) = \|\boldsymbol{z}\|(\boldsymbol{Q}\boldsymbol{X}_i)_1 = \|\boldsymbol{z}\|\widetilde{w}_i.$$

This implies that $S^\star$ is constructed as the union of S and the $k^\star - k$ indices in $\{1, \ldots, p\} \setminus S$ with the largest $|\widetilde{w}_i|$. Since $\widetilde{\boldsymbol{w}}$ and $\widetilde{\boldsymbol{W}}$ are independent, we see that both $\widetilde{\boldsymbol{W}}_{\overline{S^\star}}$ and $\widetilde{\boldsymbol{W}}_{S^\star}$ are also Gaussian random matrices. These points are crucial in the proof of these two lemmas.

LEMMA A.11. *Let $k < k^\star < \min\{n, p\}$ be any (deterministic) integer. Denote by $\sigma_{\min}$ and $\sigma_{\max}$, respectively, the smallest and the largest singular value of $\boldsymbol{X}_{S^\star}$. Then for any $t > 0$,*

$$\sigma_{\min} > \sqrt{1 - 1/n} - \sqrt{k^\star/n} - t$$

holds with probability at least $1 - \mathrm{e}^{-nt^2/2}$. Furthermore,

$$\sigma_{\max} < \sqrt{1 - 1/n} + \sqrt{k^\star/n} + \sqrt{8k^\star \log(p/k^\star)/n} + t$$

holds with probability at least $1 - \mathrm{e}^{-nt^2/2} - (\sqrt{2}\mathrm{e}k^\star/p)^{k^\star}$.

PROOF OF LEMMA A.11. Recall that $\widetilde{\boldsymbol{W}}_{S^\star} \in \mathbb{R}^{(n-1)\times k^\star}$ is a Gaussian design with i.i.d. $\mathcal{N}(0, 1/n)$ entries. Since $\boldsymbol{W}_{S^\star}$ and $\boldsymbol{X}_{S^\star}$ have the the same set of singular values, we consider $\boldsymbol{W}_{S^\star}$.

Classical theory on Wishart matrices (see [5], for example) asserts that (i) all the singular values of $\widetilde{\boldsymbol{W}}_{S^\star}$ are larger than $\sqrt{1 - 1/n} - \sqrt{k^\star/n} - t$ with probability at least $1 - \mathrm{e}^{-nt^2/2}$, and (ii) are all smaller than $\sqrt{1 - 1/n} + \sqrt{k^\star/n} + t$ with probability at least $1 - \mathrm{e}^{-nt^2/2}$. Clearly, all the singular values larger of $\boldsymbol{W}_{S^\star}$ are at least as large as $\sigma_{\min}(\widetilde{\boldsymbol{W}}_{S^\star})$. Thus, (i) yields the first claim. For the other, Lemma A.7 asserts that the event $\|\widetilde{\boldsymbol{w}}_{S^\star}\| \le \sqrt{8k^\star \log(p/k^\star)}$ happens with probability at least $1 - (\sqrt{2}\mathrm{e}k^\star/p)^{k^\star}$. On this event,

$$\|\boldsymbol{W}_{S^\star}\| \le \sqrt{\|\widetilde{\boldsymbol{W}}_{S^\star}\|^2 + 8k^\star \log(p/k^\star)},$$

where $\|\cdot\|$ denotes the spectral norm. Hence, (ii) gives

$$\|\boldsymbol{W}_{S^\star}\| \le \|\widetilde{\boldsymbol{W}}_{S^\star}\| + \sqrt{8k^\star \log(p/k^\star)} \le \sqrt{1 - 1/n} + \sqrt{k^\star/n} + t + \sqrt{8k^\star \log(p/k^\star)}$$

with probability at least $1 - \mathrm{e}^{-nt^2/2} - (\sqrt{2}\mathrm{e}k^\star/p)^{k^\star}$.

$\square$

Lemma A.12. *Denote by $\widehat{\boldsymbol{b}}_{S^\star}$ the solution to the reduced SLOPE problem* (4.6) *with $T = S^\star$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{\boldsymbol{\epsilon}}$. Keep the assumptions from Lemma A.9, and additionally assume $k^\star/\min\{n,p\} \to 0$. Then there exists a constant C_q only depending on q such that*

$$\boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{X}_{S^\star}(\boldsymbol{\beta}_{S^\star} - \widehat{\boldsymbol{b}}_{S^\star}) \preceq C_q \cdot \sqrt{\frac{k^\star \log p}{n}}\,(\lambda^{\mathrm{BH}}_{k^\star+1}, \ldots, \lambda^{\mathrm{BH}}_p)$$

with probability tending to one.

Proof of Lemma A.12. In this proof, C is a constant that only depends on q and whose value may change at each occurrence. Rearrange the objective term as

$$\begin{aligned}
\boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{X}_{S^\star}(\boldsymbol{\beta}_{S^\star} - \widehat{\boldsymbol{b}}_{S^\star}) &= \boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{X}_{S^\star}(\boldsymbol{X}'_{S^\star}\boldsymbol{X}_{S^\star})^{-1}(\boldsymbol{X}'_{S^\star}(\boldsymbol{y} - \boldsymbol{X}_{S^\star}\widehat{\boldsymbol{b}}_{S^\star}) - \boldsymbol{X}'_{S^\star}\boldsymbol{z}) \\
&= \boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{X}_{S^\star}(\boldsymbol{X}'_{S^\star}\boldsymbol{X}_{S^\star})^{-1}(\boldsymbol{X}'_{S^\star}(\boldsymbol{y} - \boldsymbol{X}_{S^\star}\widehat{\boldsymbol{b}}_{S^\star}) - \boldsymbol{X}'_{S^\star}\boldsymbol{z}) \\
&= \boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{Q}'\boldsymbol{\xi},
\end{aligned}$$

where

$$\boldsymbol{\xi} := \boldsymbol{Q}\boldsymbol{X}_{S^\star}(\boldsymbol{X}'_{S^\star}\boldsymbol{X}_{S^\star})^{-1}\left(\boldsymbol{X}'_{S^\star}(\boldsymbol{y} - \boldsymbol{X}_{S^\star}\widehat{\boldsymbol{b}}_{S^\star}) - \boldsymbol{X}'_{S^\star}\boldsymbol{z}\right).$$

For future usage, note that $\boldsymbol{\xi}$ only depends on $\widetilde{\boldsymbol{w}}$ and $\widetilde{\boldsymbol{W}}_{S^\star}$ and is, therefore, independent of $\widetilde{\boldsymbol{W}}_{\overline{S^\star}}$.

We begin by bounding $\|\boldsymbol{\xi}\|$. It follows from the KKT condition of SLOPE that $\boldsymbol{X}'_{S^\star}(\boldsymbol{y} - \boldsymbol{X}_{S^\star}\widehat{\boldsymbol{b}}_{S^\star})$ is majorized by $\boldsymbol{\lambda}^{[k^\star]}$. Hence, it follows from Fact 3.1 that

$$\left\|\boldsymbol{X}'_{S^\star}(\boldsymbol{y} - \boldsymbol{X}_{S^\star}\widehat{\boldsymbol{b}}_{S^\star})\right\| \le \|\boldsymbol{\lambda}^{[k^\star]}\|. \tag{A.28}$$

Lemma A.11 with $t = 1/2$ gives

$$\left\|\boldsymbol{X}_{S^\star}(\boldsymbol{X}'_{S^\star}\boldsymbol{X}_{S^\star})^{-1}\right\| \le \left(\sqrt{1 - 1/n} - \sqrt{k^\star/n} - 1/2\right)^{-1} < 2.01 \tag{A.29}$$

with probability at least $1 - \mathrm{e}^{-n/8}$ for sufficiently large p, where in the last step we have used $k^\star/n \to 0$. Hence, from (A.28) and (A.29) we get

$$\begin{aligned}
\|\boldsymbol{\xi}\| &\le \left\|\boldsymbol{X}_{S^\star}(\boldsymbol{X}'_{S^\star}\boldsymbol{X}_{S^\star})^{-1}\right\| \cdot \left\|\boldsymbol{X}'_{S^\star}(\boldsymbol{y} - \boldsymbol{X}_{S^\star}\widehat{\boldsymbol{b}}_{S^\star}) - \boldsymbol{X}'_{S^\star}\boldsymbol{z}\right\| \\
&\le 2.01\left(\|\boldsymbol{\lambda}^{[k^\star]}\| + 4\sqrt{2k^\star\log(p/k^\star)}\right) \\
&\le 2.01\left((1+\epsilon)\sqrt{C} + 4\sqrt{2}\right)\sqrt{k^\star\log(p/k^\star)} \\
&= C \cdot \sqrt{k^\star\log(p/k^\star)}
\end{aligned} \tag{A.30}$$

with probability at least $1 - \mathrm{e}^{-n/2} - (\sqrt{2}\mathrm{e}k^\star/p)^{k^\star} - \mathrm{e}^{-n/8} \to 1$; we used Lemma A.7 in the second line and Lemma A.8 in the third. (A.30) will help us in finishing the proof.

Write

$$\boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{X}_{S^\star}(\boldsymbol{\beta}_{S^\star} - \widehat{\boldsymbol{b}}_{S^\star}) = \boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{Q}'\boldsymbol{\xi} = \boldsymbol{W}'_{\overline{S^\star}}\boldsymbol{\xi} = \left(\widetilde{\boldsymbol{w}}'_{\overline{S^\star}}, \boldsymbol{0}\right)\boldsymbol{\xi} + \left(\boldsymbol{0}, \widetilde{\boldsymbol{W}}'_{\overline{S^\star}}\right)\boldsymbol{\xi}. \tag{A.31}$$

It follows from Lemma A.9 that $\widetilde{\boldsymbol{w}}_{\overline{S^\star}}$ is majorized by $\left(\lambda^{\mathrm{BH}}_{k^\star+1}, \lambda^{\mathrm{BH}}_{k^\star+2}, \ldots, \lambda^{\mathrm{BH}}_p\right)/\sqrt{n}$ in probability. As a result, the first term in the right-hand side obeys

$$\left(\widetilde{\boldsymbol{w}}'_{\overline{S^\star}}, \boldsymbol{0}\right)\boldsymbol{\xi} = \xi_1 \cdot \widetilde{\boldsymbol{w}}'_{\overline{S^\star}} \preceq \|\boldsymbol{\xi}\| \cdot \widetilde{\boldsymbol{w}}'_{\overline{S^\star}} \preceq C \cdot \sqrt{\frac{k^\star}{n}\log\frac{p}{k^\star}}\left(\lambda^{\mathrm{BH}}_{k^\star+1}, \lambda^{\mathrm{BH}}_{k^\star+2}, \ldots, \lambda^{\mathrm{BH}}_p\right) \tag{A.32}$$

with probability tending to one. For the second term, by exploiting the independence between $\boldsymbol{\xi}$ and $\widetilde{\boldsymbol{W}}_{\overline{S^\star}}$, we have

$$\left(\boldsymbol{0}, \widetilde{\boldsymbol{W}}'_{\overline{S^\star}}\right)\boldsymbol{\xi} \stackrel{d}{=} \sqrt{\frac{\xi_2^2 + \cdots + \xi_n^2}{n}}(\zeta_1, \ldots, \zeta_{p-k^\star}),$$

where $\zeta_1, \ldots, \zeta_{p-k^\star}$ are i.i.d. $\mathcal{N}(0, 1/n)$. Since $k^\star/p \to 0$, applying Lemma A.10 gives

$$(\zeta_1, \ldots, \zeta_{p-k^\star}) \preceq C \cdot \sqrt{\frac{\log p}{\log(p/k^\star)}}\left(\lambda^{\mathrm{BH}}_{k^\star+1}, \ldots, \lambda^{\mathrm{BH}}_p\right)$$

with probability approaching one. Hence, owing to (A.30),

$$\left(\boldsymbol{0}, \widetilde{\boldsymbol{W}}'_{\overline{S^\star}}\right)\boldsymbol{\xi} \preceq C \cdot \sqrt{\frac{k^\star \log p}{n}}\left(\lambda^{\mathrm{BH}}_{k^\star+1}, \ldots, \lambda^{\mathrm{BH}}_p\right) \tag{A.33}$$

holds with probability approaching one. Finally, combining (A.32) and (A.33) gives that

$$\begin{aligned}
\boldsymbol{X}'_{\overline{S^\star}}\boldsymbol{X}_{S^\star}(\boldsymbol{\beta}_{S^\star} - \widehat{\boldsymbol{b}}_{S^\star}) &= \left(\widetilde{\boldsymbol{w}}'_{\overline{S^\star}}, \boldsymbol{0}\right)\boldsymbol{\xi} + \left(\boldsymbol{0}, \widetilde{\boldsymbol{W}}'_{\overline{S^\star}}\right)\boldsymbol{\xi} \\
&\preceq C \cdot \left(\sqrt{\frac{k^\star}{n}\log\frac{p}{k^\star}} + \sqrt{\frac{k^\star \log p}{n}}\right) \cdot \left(\lambda^{\mathrm{BH}}_{k^\star+1}, \ldots, \lambda^{\mathrm{BH}}_p\right) \\
&\preceq C \cdot \sqrt{\frac{k^\star \log p}{n}}\left(\lambda^{\mathrm{BH}}_{k^\star+1}, \ldots, \lambda^{\mathrm{BH}}_p\right)
\end{aligned}$$

holds with probability tending to one. □

A.4. Proofs for Section 5.

LEMMA A.13. *Keep the assumptions from Lemma 5.1 and let* $\zeta_1, \ldots, \zeta_p$ *be i.i.d.* $\mathcal{N}(0,1)$. *Then*

$$\#\left\{2 \le i \le p : \zeta_i > \tau + \zeta_1\right\} \to \infty$$

in probability.

PROOF OF LEMMA A.13. With probability tending to one, $\tau' := \tau + \zeta_1$ also obeys $\tau'/\sqrt{2\log p} \to 1$ and $\sqrt{2\log p} - \tau' \to \infty$. This shows that we only need to prove a simpler version of this lemma, namely, $\#\left\{1 \le i \le p : \zeta_i > \tau\right\} \to \infty$ in probability.

Put $\Delta = \sqrt{2\log p} - \tau = o(\sqrt{2\log p})$ and $a = \mathbb{P}(\xi_1 > \tau)$. Then, $\#\left\{1 \le i \le p : \zeta_i > \tau\right\}$ is a binomial random variable with p trials and success probability a. Hence, it suffices to demonstrate that $ap \to \infty$. To this end, note that

$$\begin{aligned} a = 1 - \Phi(\tau) \sim \frac{1}{\tau}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\tau^2}{2}} &\asymp \frac{1}{\sqrt{2\log p}}\mathrm{e}^{-\log p - \Delta^2/2 + \Delta\sqrt{2\log p}} \\ &= \frac{1}{p\sqrt{2\log p}}\mathrm{e}^{(1+o(1))\Delta\sqrt{2\log p}}, \end{aligned}$$

which gives

$$ap \asymp \frac{1}{\sqrt{2\log p}}\mathrm{e}^{(1+o(1))\Delta\sqrt{2\log p}}.$$

Since $\Delta \to \infty$ (in fact, it is sufficient to have Δ bounded away from 0 from below), we have

$$\mathrm{e}^{(1+o(1))\Delta\sqrt{2\log p}}/\sqrt{2\log p} \to \infty,$$

as we wish. □

PROOF OF LEMMA 5.1. For sufficiently large p, $2(1-\epsilon)\log p \le (1 - \epsilon/2)\tau^2$. Hence, it is sufficient to show

$$\mathbb{P}_{\boldsymbol{\pi}}\left(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \le (1-\epsilon/2)\tau^2\right) \to 0$$

uniformly for all estimators $\widehat{\boldsymbol{\beta}}$. Letting I be the random coordinate,

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 = \sum_{j \ne I} \widehat{\beta}_j^2 + (\widehat{\beta}_I - \tau)^2 = \|\widehat{\boldsymbol{\beta}}\|^2 + \tau^2 - 2\tau\widehat{\beta}_I,$$

which is smaller than or equal to $(1-\epsilon/2)\tau^2$ if and only if

$$\widehat{\beta}_I \ge \frac{2\|\widehat{\boldsymbol{\beta}}\|^2 + \epsilon\tau^2}{4\tau}.$$

Denote by $A = A(\boldsymbol{y}; \widehat{\boldsymbol{\beta}})$ the set of all $i \in \{1, \ldots, p\}$ such that $\widehat{\beta}_i \ge (2\|\widehat{\boldsymbol{\beta}}\|^2 + \epsilon\tau^2)/(4\tau)$, and let $\widehat{b}$ be the minimum value of these $\widehat{\beta}_i$. Then

$$\widehat{b} \ge \frac{2\|\widehat{\boldsymbol{\beta}}\|^2 + \epsilon\tau^2}{4\tau} \ge \frac{2|A|\widehat{b}^2 + \epsilon\tau^2}{4\tau} \ge \frac{2\sqrt{2|A|\widehat{b}^2 \cdot \epsilon\tau^2}}{4\tau},$$

which gives

$$|A| \le 2/\epsilon. \tag{A.34}$$

Recall that $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \le (1 - \epsilon/2)\tau^2$ if and only if I is among these $|A|$ components. Hence,

$$\mathbb{P}_{\boldsymbol{\pi}}\left(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \le (1 - \epsilon/2)\tau^2 \middle| \boldsymbol{y}\right) = \mathbb{P}_{\boldsymbol{\pi}}(I \in A|\boldsymbol{y}) = \sum_{i \in A} \mathbb{P}_{\boldsymbol{\pi}}(I = i|\boldsymbol{y}) = \frac{\sum_{i \in A} \mathrm{e}^{\tau y_i}}{\sum_{i=1}^{p} \mathrm{e}^{\tau y_i}}, \tag{A.35}$$

where we use the fact that A is almost surely determined by $\boldsymbol{y}$. Since (A.35) is maximal if A is the set of indices with the largest y_i's, (A.35) and (A.34) together yield

$$\begin{aligned}&\mathbb{P}_{\boldsymbol{\pi}}\left(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \le (1 - \epsilon/2)\tau^2\right)\\ &\le \mathbb{P}_{\boldsymbol{\pi}}\left(y_I = \tau + z_I \text{ is at least the } \lceil 2/\epsilon \rceil^{\text{th}} \text{ largest among } y_1, \ldots, y_p\right) \to 0,\end{aligned}$$

where the last step is provided by Lemma A.13.

□

PROOF OF LEMMA 5.3. To closely follow the proof of Lemma 5.1, denote by $A = A(\boldsymbol{y}, \boldsymbol{X}; \widehat{\boldsymbol{\beta}})$ the set of all $i \in \{1, \ldots, p\}$ such that $\widehat{\beta}_i \ge (2\|\widehat{\boldsymbol{\beta}}\|^2 + \epsilon\alpha^2\tau^2)/(4\alpha\tau)$, and keep the same notation $\widehat{b}$ as before. Then $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \le (1 - \epsilon/2)\alpha^2\tau^2$ if and only if $I \in A$. Hence,

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\pi}}\left(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \le (1 - \epsilon/2)\alpha^2\tau^2 \middle| \boldsymbol{y}, \boldsymbol{X}\right) &= \mathbb{P}_{\boldsymbol{\pi}}(I \in A|\boldsymbol{y}, \boldsymbol{X})\\ &= \sum_{i \in A} \mathbb{P}_{\boldsymbol{\pi}}(I = i|\boldsymbol{y}, \boldsymbol{X})\\ &= \frac{\sum_{i \in A} \exp\left(\alpha\tau \boldsymbol{X}_i'\boldsymbol{y} - \alpha^2\tau^2\|\boldsymbol{X}_i\|^2/2\right)}{\sum_{i=1}^{p} \exp\left(\alpha\tau \boldsymbol{X}_i'\boldsymbol{y} - \alpha^2\tau^2\|\boldsymbol{X}_i\|^2/2\right)}\end{aligned} \tag{A.36}$$

and this quantity is maximal if A is the set of indices i with the largest values of $\boldsymbol{X}_i'\boldsymbol{y}/\alpha - \tau\|\boldsymbol{X}_i\|^2/2$. As shown in Lemma 5.1, $|A| \le 2/\epsilon$, which gives

$$\mathbb{P}_{\boldsymbol{\pi}}\left(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|^2 \le (1-\epsilon/2)\alpha^2\tau^2\right) \le \mathbb{P}_{\boldsymbol{\pi}}\left(\boldsymbol{X}_I'\boldsymbol{y}/\alpha-\tau\|\boldsymbol{X}_I\|^2/2 \text{ is at least the } \lceil 2/\epsilon\rceil^{\text{th}} \text{ largest}\right). \tag{A.37}$$

We complete the proof by showing that the probability in the right-hand side of (A.37) is negligible uniformly over all estimators $\widehat{\boldsymbol{\beta}}$ as $p\to\infty$. By the independence between I and $\boldsymbol{X},\boldsymbol{z}$, we can assume $I=1$ while evaluating this probability. With this in mind, we aim to show that there are sufficiently many i's such that

$$\boldsymbol{X}_i'\boldsymbol{y}/\alpha-\frac{\tau}{2}\|\boldsymbol{X}_i\|^2-\boldsymbol{X}_1'\boldsymbol{y}/\alpha+\frac{\tau}{2}\|\boldsymbol{X}_1\|^2=\boldsymbol{X}_i'\left(\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1\right)-\frac{\tau}{2}\|\boldsymbol{X}_i\|^2-\boldsymbol{X}_1'\boldsymbol{z}/\alpha-\frac{\tau}{2}\|\boldsymbol{X}_1\|^2$$

is positive. Since

$$\boldsymbol{X}_1'\boldsymbol{z}/\alpha+\frac{\tau}{2}\|\boldsymbol{X}_1\|^2=O_{\mathbb{P}}(1/\alpha)+\frac{\tau}{2}\left(1+O_{\mathbb{P}}(1/\sqrt{n})\right),$$

it suffices to show that

$$\#\left\{2\le i\le p:\boldsymbol{X}_i'\left(\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1\right)-\frac{\tau}{2}\|\boldsymbol{X}_i\|^2>\frac{C_1}{\alpha}+\frac{\tau}{2}+\frac{C_2\tau}{\sqrt{n}}\right\}\le\lceil 2/\epsilon\rceil-1 \tag{A.38}$$

holds with vanishing probability for all positive constants C_1, C_2. By the independence between $\boldsymbol{X}_i$ and $\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1$, we can replace $\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1$ by $(\|\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1\|,0,\ldots,0)$ in (A.38). That is,

$$\boldsymbol{X}_i'\left(\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1\right)-\frac{\tau}{2}\|\boldsymbol{X}_i\|^2\overset{d}{=}\|\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1\|\,X_{i,1}-\frac{\tau}{2}X_{i,1}^2-\frac{\tau}{2}\|\boldsymbol{X}_{i,-1}\|^2,$$

where $\boldsymbol{X}_{i,-1}\in\mathbb{R}^{n-1}$ is $\boldsymbol{X}_i$ without the first entry. To this end, we point out that the following three events all happen with probability tending to one:

$$\begin{gathered}\#\{2\le i\le p:\|\boldsymbol{X}_{i,-1}\|\le 1\}/p\to 1/2,\\ \max_i X_{i,1}^2\le\frac{2\log p}{n},\\ \|\boldsymbol{z}/\alpha+\tau\boldsymbol{X}_1\|\ge\left(\sqrt{n}-\sqrt{\log p}\right)/\alpha.\end{gathered} \tag{A.39}$$

Making use of this and (A.38), we only need to show that

$$N\triangleq\#\left\{2\le i\le 0.49p:\frac{1}{\alpha}\left(1-\sqrt{(\log p)/n}\right)\sqrt{n}X_{i,1}>\frac{\tau\log p}{n}+\frac{\tau}{2}+\frac{C_1}{\alpha}+\frac{\tau}{2}+\frac{C_2\tau}{\sqrt{n}}\right\}$$

$$\#\left\{2\le i\le 0.49p:\frac{1}{\alpha}\left(1-\sqrt{(\log p)/n}\right)\sqrt{n}X_{i,1}>\tau+\frac{\tau\log p}{n}+\frac{C_1}{\alpha}+\frac{C_2\tau}{\sqrt{n}}\right\}$$

obeys

$$N \le \lceil 2/\epsilon \rceil - 1 \tag{A.40}$$

with vanishing probability. The first line of (A.39) shows that there are at least $0.49p$ many i's such that $\|\boldsymbol{X}_{i,-1}\| \le 1$ and we assume they correspond to indices $2 \le i \le 0.49p$ without loss of generality. (Note that N is independent of all $\boldsymbol{X}_{i,-1}$'s.) Observe that

$$\tau' := \frac{\tau + \tau(\log p)/n + C_1/\alpha + C_2\tau/\sqrt{n}}{\left(1 - \sqrt{(\log p)/n}\right)/\alpha} = \alpha\left(1 + 2\sqrt{\frac{\log p}{n}}\right)\tau + O(1)$$

for sufficiently large p (to ensure $(\log p)/n$ is small). Hence, plugging the specific choice of τ and using $\alpha \le 1$, we obtain

$$\tau' \le \left(1 + 2\sqrt{(\log p)/n}\right)\tau + O(1) \le \sqrt{2\log p} - \log\sqrt{2\log p} + O(1),$$

which reveals that $\sqrt{2\log(0.49p)} - \tau' = \sqrt{2\log p} - \tau' + o(1) \to \infty$. Since $\sqrt{n}X_{n,i}$ are i.i.d. $\mathcal{N}(0,1)$, Lemma A.13 validates (A.40). $\square$

Proof of Corollary 1.5. Let $c > 0$ be a sufficiently small constant to be determined later. It is sufficient to prove the claim with p replaced by a possibly smaller value given by $p^\star := \min\{\lfloor cn \rfloor, p\}$ (if we knew that $\beta_i = 0$ for $p^\star + 1 \le i \le p$, the loss of any estimator $\boldsymbol{X}\widehat{\boldsymbol{\beta}}$ would not increase after projecting onto the linear space spanned by the first $p^\star$ columns). Hereafter, we assume $\boldsymbol{X} \in \mathbb{R}^{n\times p^\star}$ and $\boldsymbol{\beta} \in \mathbb{R}^{p^\star}$. Observe that $p = O(n)$ implies $p = O(p^\star)$ and, therefore,

$$\log(p^\star/k) \sim \log(p/k). \tag{A.41}$$

In particular, $k/p^\star \to 0$ and $n/\log(p^\star/k) \to \infty$. This suggests that we can apply Theorem 5.4 to our problem, obtaining

$$\inf_{\widehat{\boldsymbol{\beta}}} \sup_{\|\boldsymbol{\beta}\|_0 \le k} \mathbb{P}\left(\frac{\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2}{2k\log(p^\star/k)} > 1 - \epsilon'\right) \to 1.$$

for every constant $\epsilon' > 0$. Because of (A.41), we also have

$$\inf_{\widehat{\boldsymbol{\beta}}} \sup_{\|\boldsymbol{\beta}\|_0 \le k} \mathbb{P}\left(\frac{\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2}{2k\log(p/k)} > 1 - \epsilon'\right) \to 1 \tag{A.42}$$

for any $\epsilon' > 0$.

Since $p^\star/n \le c \le 1$, the smallest singular value of the Gaussian random matrix $\boldsymbol{X}$ is at least $1 - \sqrt{c} + o_{\mathbb{P}}(1)$ (see, for example, [5]). This result, together with (A.42), yields

$$\inf_{\widehat{\boldsymbol{\beta}}} \sup_{\|\boldsymbol{\beta}\|_0 \le k} \mathbb{P}\left(\frac{\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}\|^2}{2k\log(p/k)} > (1-\sqrt{c})^2(1-\epsilon') \right) \to 1$$

for each $\epsilon' > 0$. Finally, choose c and ϵ' sufficiently small such that $(1 - \sqrt{c})^2(1-\epsilon') > 1 - \epsilon$. □

REFERENCES

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables.* Courier Corporation, 2012.

[2] M. Bogdan, E. van den Berg, C. Sabatti, W. Su, and E. J. Candès. SLOPE–adaptive variable selection via convex optimization. *arXiv preprint arXiv:1407.3824*, 2014.

[3] S. Boucheron and M. Thomas. Concentration inequalities for order statistics. *Electronic Communications in Probability*, 17(51):1–12, 2012.

[4] L. de Haan and A. Ferreira. *Extreme value theory: An introduction.* Springer Science & Business Media, 2007.

[5] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing, Theory and Applications*, 2012.

Departments of Statistics
Stanford University
Stanford, CA 94305
USA
E-mail: wjsu@stanford.edu

Departments of Statistics and Mathematics
Stanford University
Stanford, CA 94305
USA
E-mail: candes@stanford.edu