The Annals of Statistics
arXiv: arXiv:1511.01957

SUPPLEMENT TO "FALSE DISCOVERIES OCCUR EARLY ON THE LASSO PATH"

BY WEIJIE SU* AND MAŁGORZATA BOGDAN† AND EMMANUEL CANDÈS‡

University of Pennsylvania, University of Wroclaw, and Stanford University

APPENDIX A: ROAD MAP TO THE PROOFS

In this section, we provide an overview of the proof of Theorem 2.1, presenting all the key steps and ingredients. Detailed proofs are distributed in Appendices B–D. At a high level, the proof structure has the following three elements:

1. Characterize the Lasso solution at a fixed λ asymptotically, predicting the (non-random) asymptotic values of the FDP and of the TPP denoted by $\mathrm{fdp}^\infty(\lambda)$ and $\mathrm{tpp}^\infty(\lambda)$, respectively. These limits depend depend on Π, δ, ϵ and σ.
2. Exhibit uniform convergence over λ in the sense that

$$\sup_{\lambda_{\min} \le \lambda \le \lambda_{\max}} |\mathrm{FDP}(\lambda) - \mathrm{fdp}^\infty(\lambda)| \xrightarrow{\mathbb{P}} 0,$$

   and similarly for $\mathrm{TPP}(\lambda)$. A consequence is that in the limit, the asymptotic trade-off between true and false positive rates is given by the λ-parameterized curve $(\mathrm{tpp}^\infty(\lambda), \mathrm{fdp}^\infty(\lambda))$.
3. The trade-off curve from the last step depends on the prior Π. The last step optimizes it by varying λ and Π.

Whereas the last two steps are new and present some technical challenges, the first step is accomplished largely by resorting to off-the-shelf AMP theory. We now present each step in turn. Throughout this section we work under our working hypothesis and take the noise level σ to be positive.

*Supported in part by NSF under grant CCF-0963835.

†Supported in part by the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement No. 602552 and co-financed by the Polish Ministry of Science and Higher Education under Grant Agreement 2932/7.PR/2013/2

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

Step 1. First, Lemma A.1 below accurately predicts the asymptotic limits of FDP and TPP at a fixed λ . This lemma is borrowed from [2], which follows from Theorem 1.5 in [1] in a natural way, albeit with some effort spent in resolving a continuity issue near the origin. Recall that $\eta_t(\cdot)$ is the soft-thresholding operator defined as $\eta_t(x) = \mathrm{sgn}(x)(|x| - t)_+$, and $\Pi^\star$ is the distribution of Π conditionally on being nonzero;

$$\Pi = \begin{cases} \Pi^\star, & \text{w.p. } \epsilon, \\ 0, & \text{w.p. } 1 - \epsilon. \end{cases}$$

Denote by α_0 the unique root of $(1 + t^2)\Phi(-t) - t\phi(t) = \delta/2$.

LEMMA A.1 (Theorem 1 in [2]; see also Theorem 1.5 in [1]). *The Lasso solution with a fixed $\lambda > 0$ obeys*

$$\frac{V(\lambda)}{p} \xrightarrow{\mathbb{P}} 2(1-\epsilon)\Phi(-\alpha), \qquad \frac{T(\lambda)}{p} \xrightarrow{\mathbb{P}} \epsilon \cdot \mathbb{P}(|\Pi^\star + \tau W| > \alpha\tau),$$

where W is $\mathcal{N}(0,1)$ independent of Π, and $\tau > 0, \alpha > \max\{\alpha_0, 0\}$ is the unique solution to

$$\begin{aligned} \tau^2 &= \sigma^2 + \frac{1}{\delta}\mathbb{E}\left(\eta_{\alpha\tau}(\Pi + \tau W) - \Pi\right)^2 \\ \lambda &= \left(1 - \frac{1}{\delta}\mathbb{P}(|\Pi + \tau W| > \alpha\tau)\right)\alpha\tau. \end{aligned} \tag{A.1}$$

Note that both τ and α depend on λ.

We pause to briefly discuss how Lemma A.1 follows from Theorem 1.5 in [1]. There, it is rigorously proven that the joint distribution of $(\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})$ is, in some sense, asymptotically the same as that of $(\boldsymbol{\beta}, \eta_{\alpha\tau}(\boldsymbol{\beta} + \tau\boldsymbol{W}))$, where $\boldsymbol{W}$ is a p-dimensional vector of i.i.d. standard normals independent of $\boldsymbol{\beta}$, and where the soft-thresholding operation acts in a componentwise fashion. Roughly speaking, the Lasso estimate $\widehat{\beta}_j$ looks like $\eta_{\alpha\tau}(\beta_j + \tau W_j)$, so that we are applying soft thresholding at level $\alpha\tau$ rather than λ and the noise level is τ rather than σ. With these results in place, we informally obtain

$$\begin{aligned} V(\lambda)/p = \#\{j : \widehat{\beta}_j \neq 0, \beta_j = 0\}/p &\approx \mathbb{P}(\eta_{\alpha\tau}(\Pi + \tau W) \neq 0, \Pi = 0) \\ &= (1-\epsilon)\, \mathbb{P}(|\tau W| > \alpha\tau) \\ &= 2(1-\epsilon)\, \Phi(-\alpha). \end{aligned}$$

Similarly, $T(\lambda)/p \approx \epsilon\, \mathbb{P}(|\Pi^\star + \tau W| > \alpha\tau)$. For details, we refer the reader to Theorem 1 in [2].

Step 2. Our interest is to extend this convergence result uniformly over a range of λ's. The proof of this step is the subject of Section B.

LEMMA A.2. *For any fixed $0 < \lambda_{\min} < \lambda_{\max}$, the convergence of $V(\lambda)/p$ and $T(\lambda)/p$ in Lemma A.1 is uniform over $\lambda \in [\lambda_{\min}, \lambda_{\max}]$.*

Hence, setting

$$\text{fd}^{\infty}(\lambda) := 2(1-\epsilon)\Phi(-\alpha), \quad \text{td}^{\infty}(\lambda) := \epsilon\, \mathbb{P}(|\Pi^{\star} + \tau W| > \alpha\tau) \tag{A.2}$$

we have

$$\sup_{\lambda_{\min} \le \lambda \le \lambda_{\max}} \left| \frac{V(\lambda)}{p} - \text{fd}^{\infty}(\lambda) \right| \xrightarrow{\mathbb{P}} 0,$$

and

$$\sup_{\lambda_{\min} \le \lambda \le \lambda_{\max}} \left| \frac{T(\lambda)}{p} - \text{td}^{\infty}(\lambda) \right| \xrightarrow{\mathbb{P}} 0.$$

To exhibit the trade-off between FDP and TPP, we can therefore focus on the far more amenable quantities $\text{fd}^{\infty}(\lambda)$ and $\text{td}^{\infty}(\lambda)$ instead of $V(\lambda)$ and $T(\lambda)$. Since $\text{FDP}(\lambda) = V(\lambda)/(V(\lambda) + T(\lambda))$ and $\text{TPP}(\lambda) = T(\lambda)/|\{j : \beta_j \neq 0\}|$, this gives

$$\sup_{\lambda_{\min} \le \lambda \le \lambda_{\max}} |\text{FDP}(\lambda) - \text{fdp}^{\infty}(\lambda)| \xrightarrow{\mathbb{P}} 0, \quad \text{fdp}^{\infty}(\lambda) = \frac{\text{fd}^{\infty}(\lambda)}{\text{fd}^{\infty}(\lambda) + \text{td}^{\infty}(\lambda)},$$

and

$$\sup_{\lambda_{\min} \le \lambda \le \lambda_{\max}} |\text{TPP}(\lambda) - \text{tpp}^{\infty}(\lambda)| \xrightarrow{\mathbb{P}} 0, \quad \text{tpp}^{\infty}(\lambda) = \frac{\text{td}^{\infty}(\lambda)}{\epsilon},$$

so that $\text{fdp}^{\infty}(\lambda)$ and $\text{tpp}^{\infty}(\lambda)$ are the predicted FDP and TPP. (We shall often hide the dependence on λ.)

Step 3. As remarked earlier, both $\text{tpp}^{\infty}(\lambda)$ and $\text{fdp}^{\infty}(\lambda)$ depend on Π, δ, ϵ and σ. In Appendix C, we will see that we can parameterize the trade-off curve $(\text{tpp}^{\infty}(\lambda), \text{fdp}^{\infty}(\lambda))$ by the true positive rate so that there is a function q^{Π} obeying $q^{\Pi}(\text{tpp}^{\infty}) = \text{fdp}^{\infty}$; furthermore, this function depends on Π and σ only through Π/σ. Therefore, realizations of the FDP-TPP pair fall asymptotically arbitrarily close to q^{Π}. It remains to optimize the curve q^{Π} over Π/σ. Specifically, the last step in Appendix C characterizes the envelope $q^{\star}$ formally given as

$$q^{\star}(u; \delta, \epsilon) = \inf q^{\Pi}(u; \delta, \epsilon),$$

where the infimum is taken over all feasible priors Π. The key ingredient in optimizing the trade-off is given by Lemma C.1.

Taken together, these three steps sketch the basic strategy for proving Theorem 2.1, and the remainder of the proof is finally carried out in Appendix D. In particular, we also establish the noiseless result ($\sigma = 0$) by using a sequence of approximating problems with noise levels approaching zero.

APPENDIX B: FOR ALL VALUES OF λ SIMULTANEOUSY

In this section we aim to prove Lemma A.2 and, for the moment, take $\sigma > 0$. Also, we shall frequently use results in [1], notably, Theorem 1.5, Lemma 3.1, Lemma 3.2, and Proposition 3.6 therein. Having said this, most of our proofs are rather self-contained, and the strategies accessible to readers who have not yet read [1]. We start by stating two auxiliary lemmas below, whose proofs are deferred to Section B.1.

Lemma B.1. *For any $c > 0$, there exists a constant $r_c > 0$ such that for any arbitrary $r > r_c$,*

$$\sup_{\|\boldsymbol{u}\|=1} \# \left\{ 1 \le j \le p : |\boldsymbol{X}_j^\top \boldsymbol{u}| > \frac{r}{\sqrt{n}} \right\} \le cp$$

holds with probability tending to one.

A key ingredient in the proof of Lemma A.2 is, in a certain sense, the uniform continuity of the support of $\widehat{\boldsymbol{\beta}}(\lambda)$. This step is justified by the auxiliary lemma below which demonstrates that the Lasso estimates are uniformly continuous in ℓ_2 norm.

Lemma B.2. *Fixe $0 < \lambda_{\min} < \lambda_{\max}$. Then there is a constant c such for any $\lambda^- < \lambda^+$ in $[\lambda_{\min}, \lambda_{\max}]$,*

$$\sup_{\lambda^- \le \lambda \le \lambda^+} \left\| \widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-) \right\| \le c\sqrt{(\lambda^+ - \lambda^-)p}$$

holds with probability tending to one.

Proof of Lemma A.2. We prove the uniform convergence of $V(\lambda)/p$ and similar arguments apply to $T(\lambda)/p$. To begin with, let $\lambda_{\min} = \lambda_0 < \lambda_1 < \cdots < \lambda_m = \lambda_{\max}$ be equally spaced points and set $\Delta := \lambda_{i+1} - \lambda_i = (\lambda_{\max} - \lambda_{\min})/m$; the number of knots m shall be specified later. It follows from Lemma A.1 that

$$\max_{0 \le i \le m} |V(\lambda_i)/p - \mathrm{fd}^\infty(\lambda_i)| \xrightarrow{\mathbb{P}} 0 \tag{B.1}$$

by a union bound. Now, according to Corollary 1.7 from [1], the solution α to equation (A.1) is continuous in λ and, therefore, $\mathrm{fd}^\infty(\lambda)$ is also continuous on $[\lambda_{\min}, \lambda_{\max}]$. Thus, for any constant $\omega > 0$, the equation

$$|\mathrm{fd}^\infty(\lambda) - \mathrm{fd}^\infty(\lambda')| \le \omega \tag{B.2}$$

holds for all $\lambda_{\min} \le \lambda, \lambda' \le \lambda_{\max}$ satisfying $|\lambda - \lambda'| \le 1/m$ provided m is sufficiently large. We now aim to show that if m is sufficiently large (but fixed), then

$$\max_{0 \le i < m} \sup_{\lambda_i \le \lambda \le \lambda_{i+1}} |V(\lambda)/p - V(\lambda_i)/p| \le \omega \tag{B.3}$$

holds with probability approaching one as $p \to \infty$. Since ω is arbitrary small, combining (B.1), (B.2), and (B.3) gives uniform convergence by applying the triangle inequality.

Let $\mathcal{S}(\lambda)$ be a short-hand for $\operatorname{supp}(\widehat{\boldsymbol{\beta}}(\lambda))$. Fix $0 \le i < m$ and put $\lambda^- = \lambda_i$ and $\lambda^+ = \lambda_{i+1}$. For any $\lambda \in [\lambda^-, \lambda^+]$,

$$|V(\lambda) - V(\lambda^-)| \le |\mathcal{S}(\lambda) \setminus \mathcal{S}(\lambda^-)| + |\mathcal{S}(\lambda^-) \setminus \mathcal{S}(\lambda)|. \tag{B.4}$$

Hence, it suffices to give upper bound about the sizes of $\mathcal{S}(\lambda) \setminus \mathcal{S}(\lambda^-)$ and $\mathcal{S}(\lambda^-) \setminus \mathcal{S}(\lambda)$. We start with $|\mathcal{S}(\lambda) \setminus \mathcal{S}(\lambda^-)|$.

The KKT optimality conditions for the Lasso solution state that there exists a subgradient $\boldsymbol{g}(\lambda) \in \partial \|\widehat{\boldsymbol{\beta}}(\lambda)\|_1$ obeying

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda)) = \lambda \boldsymbol{g}(\lambda)$$

for each λ. Note that $g_j(\lambda) = \pm 1$ if $j \in \mathcal{S}(\lambda)$. As mentioned earlier, our strategy is to establish some sort of continuity of the KKT conditions with respect to λ. To this end, let

$$\boldsymbol{u} = \frac{\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-)\right)}{\left\|\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-)\right)\right\|}$$

be a point in $\mathbb{R}^n$ with unit ℓ_2 norm. Then for each $j \in \mathcal{S}(\lambda) \setminus \mathcal{S}(\lambda^-)$, we have

$$|\boldsymbol{X}_j^\top \boldsymbol{u}| = \frac{\left|\boldsymbol{X}_j^\top \boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-))\right|}{\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-))\|} = \frac{|\lambda g_j(\lambda) - \lambda^- g_j(\lambda^-)|}{\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-))\|} \ge \frac{\lambda - \lambda^- |g_j(\lambda^-)|}{\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-))\|}.$$

Now, given an arbitrary constant $a > 0$ to be determined later, either $|g_j(\lambda^-)| \in [1-a, 1)$ or $|g_j(\lambda^-)| \in [0, 1-a)$. In the first case ((a) below) note that we exclude $|g_j(\lambda^-)| = 1$ because for random designs, when $j \notin \mathcal{S}(\lambda)$ the equality $|\boldsymbol{X}_j^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda^-))| = \lambda^-$ can only hold with zero probability (see e.g. [3]). Hence, at least one of the following statements hold:

(a) $|\boldsymbol{X}_j^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda^-))| = \lambda^-|g_j(\lambda^-)| \in [(1-a)\lambda^-, \lambda^-)$;

(b) $|\boldsymbol{X}_j^\top \boldsymbol{u}| \ge \frac{\lambda-(1-a)\lambda^-}{\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda)-\widehat{\boldsymbol{\beta}}(\lambda^-))\|} > \frac{a\lambda^-}{\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda)-\widehat{\boldsymbol{\beta}}(\lambda^-))\|}$.

In the second case, since the spectral norm $\sigma_{\max}(\boldsymbol{X})$ is bounded in probability (see e.g. [4]), we make use of Lemma B.2 to conclude that

$$\frac{a\lambda^-}{\|\boldsymbol{X}(\widehat{\boldsymbol{\beta}}(\lambda)-\widehat{\boldsymbol{\beta}}(\lambda^-))\|} \ge \frac{a\lambda^-}{\sigma_{\max}(\boldsymbol{X})\|\widehat{\boldsymbol{\beta}}(\lambda)-\widehat{\boldsymbol{\beta}}(\lambda^-)\|} \ge \frac{a\lambda^-}{c\sigma_{\max}(\boldsymbol{X})\sqrt{(\lambda^+-\lambda^-)p}} \ge c'a\sqrt{\frac{m}{n}}$$

holds for all $\lambda^- \le \lambda \le \lambda^+$ with probability tending to one. Above, the constant c' only depends on $\lambda_{\min}, \lambda_{\max}, \delta$ and Π. Consequently, we see that

$$\sup_{\lambda^-\le\lambda\le\lambda^+} \left|\mathcal{S}(\lambda) \setminus \mathcal{S}(\lambda^-)\right| \le \#\left\{j : (1-a)\lambda^- \le |\boldsymbol{X}_j^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda^-))| < \lambda^-\right\} + \#\left\{j : |\boldsymbol{X}_j^\top \boldsymbol{u}| > c'a\sqrt{m/n}\right\}.$$

Equality (3.21) of [1] guarantees the existence of a constant a such that the event[1]

$$\#\left\{j : (1-a)\lambda^- \le |\boldsymbol{X}_j^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda^-))| < \lambda^-\right\} \le \frac{\omega p}{4} \tag{B.5}$$

happens with probability approaching one. Since $\lambda^- = \lambda_i$ is always in the interval $[\lambda_{\min}, \lambda_{\max}]$, the constant a can be made to be independent of the index i. For the second term, it follows from Lemma B.1 that for sufficiently large m, the event

$$\#\left\{j : |\boldsymbol{X}_j^\top \boldsymbol{u}| > c'a\sqrt{m/n}\right\} \le \frac{\omega p}{4} \tag{B.6}$$

also holds with probability approaching one. Combining (B.5) and (B.6), we get

$$\sup_{\lambda^-\le\lambda\le\lambda^+} \left|\mathcal{S}(\lambda) \setminus \mathcal{S}(\lambda^-)\right| \le \frac{\omega p}{2} \tag{B.7}$$

holds with probability tending to one.

Next, we bound $|\mathcal{S}(\lambda^-) \setminus \mathcal{S}(\lambda)|$. Applying Theorem 1.5 in [1], we can find a constant $\nu > 0$ independent of $\lambda^- \in [\lambda_{\min}, \lambda_{\max}]$ such that

$$\#\left\{j : 0 < |\widehat{\beta}_j(\lambda^-)| < \nu\right\} \le \frac{\omega p}{4} \tag{B.8}$$

[1] Apply Theorem 1.8 to carry over the results for AMP iterates to Lasso solution.

happens with probability approaching one. Furthermore, the simple inequality

$$\|\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-)\|^2 \geq \nu^2 \# \left\{ j : j \in \mathcal{S}(\lambda^-) \setminus \mathcal{S}(\lambda), |\widehat{\beta}_j(\lambda^-)| \geq \nu \right\},$$

together with Lemma B.2, give

$$\# \left\{ j : j \in \mathcal{S}(\lambda^-) \setminus \mathcal{S}(\lambda), |\widehat{\beta}_j(\lambda^-)| \geq \nu \right\} \leq \frac{\|\widehat{\boldsymbol{\beta}}(\lambda) - \widehat{\boldsymbol{\beta}}(\lambda^-)\|^2}{\nu^2} \leq \frac{c^2(\lambda^+ - \lambda^-)p}{\nu^2} \tag{B.9}$$

for all $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ with probability converging to one. Taking m sufficiently large such that $\lambda^+ - \lambda^- = (\lambda_{\max} - \lambda_{\min})/m \leq \omega\nu^2/4c^2$ in (B.9) and combining this with (B.8) gives that

$$\sup_{\lambda^- \leq \lambda \leq \lambda^+} \left|\mathcal{S}(\lambda^-) \setminus \mathcal{S}(\lambda)\right| \leq \frac{\omega p}{2} \tag{B.10}$$

holds with probability tending to one.

To conclude the proof, note that both (B.7) and (B.10) hold for a large but fixed m. Substituting these two inequalities into (B.4) confirms (B.3) by taking a union bound.

As far as the true discovery number $T(\lambda)$ is concerned, all the arguments seamlessly apply and we do not repeat them. This terminates the proof. □

B.1. Proofs of auxiliary lemmas. In this section, we prove Lemmas B.1 and B.2. While the proof of the first is straightforward, the second crucially relies on Lemma B.5, whose proof makes use of Lemmas B.3 and B.4. Hereafter, we denote by $o_{\mathbb{P}}(1)$ any random variable which tends to zero in probability.

Proof of Lemma B.1. Since

$$\|\boldsymbol{u}^\top \boldsymbol{X}\|^2 \geq \frac{r^2}{n} \# \left\{ 1 \leq j \leq p : |\boldsymbol{X}_j^\top \boldsymbol{u}| > \frac{r}{\sqrt{n}} \right\},$$

we have

$$\begin{aligned} \# \left\{ 1 \leq j \leq p : |\boldsymbol{X}_j^\top \boldsymbol{u}| > \frac{r}{\sqrt{n}} \right\} &\leq \frac{n}{r^2} \|\boldsymbol{u}^\top \boldsymbol{X}\|^2 \leq \frac{n\sigma_{\max}(\boldsymbol{X})^2 \|\boldsymbol{u}\|^2}{r^2} \\ &= (1 + o_{\mathbb{P}}(1)) \frac{(1+\sqrt{\delta})^2 p}{r^2}, \end{aligned}$$

where we make use of $\lim n/p = \delta$ and $\sigma_{\max}(\boldsymbol{X}) = 1 + \delta^{-1/2} + o_{\mathbb{P}}(1)$. To complete the proof, take any $r_c > 0$ such that $(1+\sqrt{\delta})/r_c < \sqrt{c}$. □

LEMMA B.3. *Take a sequence $a_1 \geq a_2 \geq \cdots \geq a_p \geq 0$ with at least one strict inequality, and suppose that*

$$\frac{p \sum_{i=1}^p a_i^2}{\left(\sum_{i=1}^p a_i\right)^2} \geq M$$

for some $M > 1$. Then for any $1 \leq s \leq p$,

$$\frac{\sum_{i=1}^s a_i^2}{\sum_{i=1}^p a_i^2} \geq 1 - \frac{p^3}{M s^3}.$$

PROOF OF LEMMA B.3. By the monotonicity of $\boldsymbol{a}$,

$$\frac{\sum_{i=1}^s a_i^2}{s} \geq \frac{\sum_{i=1}^p a_i^2}{p},$$

which implies

$$\frac{p \sum_{i=1}^s a_i^2}{\left(\sum_{i=1}^s a_i\right)^2} \geq \frac{s \sum_{i=1}^p a_i^2}{\left(\sum_{i=1}^p a_i\right)^2} \geq \frac{sM}{p}. \tag{B.11}$$

Similarly,

$$\sum_{i=s+1}^p a_i^2 \leq (p-s) \left(\frac{\sum_{i=1}^s a_i}{s} \right)^2 \tag{B.12}$$

and it follows from (B.11) and (B.12) that

$$\frac{\sum_{i=1}^s a_i^2}{\sum_{i=1}^p a_i^2} \geq \frac{\sum_{i=1}^s a_i^2}{\sum_{i=1}^s a_i^2 + (p-s)\left(\frac{\sum_{i=1}^s a_i}{s}\right)^2} \geq \frac{\frac{sM}{p^2}}{\frac{sM}{p^2} + \frac{p-s}{s^2}} \geq 1 - \frac{p^3}{M s^3}.$$

□

LEMMA B.4. *Assume $n/p \to 1$, i.e. $\delta = 1$. Suppose s obeys $s/p \to 0.01$. Then with probability tending to one, the smallest singular value obeys*

$$\min_{|\mathcal{S}|=s} \sigma_{\min}(\boldsymbol{X}_{\mathcal{S}}) \geq \frac{1}{2},$$

where the minimization is over all subsets of $\{1, \ldots, p\}$ of cardinality s.

PROOF OF LEMMA B.4. For a fixed $\mathcal{S}$, we have

$$\mathbb{P}\left(\sigma_{\min}(\boldsymbol{X}_{\mathcal{S}}) < 1 - \sqrt{s/n} - t\right) \leq \mathrm{e}^{-nt^2/2}$$

for all $t \geq 0$, please see [4]. The claim follows from plugging $t = 0.399$ and a union bound over $\binom{p}{s} \leq \exp(pH(s/p))$ subsets, where $H(q) = -q \log q - (1-q)\log(1-q)$. We omit the details. □

The next lemma bounds the Lasso solution in ℓ_2 norm uniformly over λ. We use some ideas from the proof of Lemma 3.2 in [1].

LEMMA B.5. *Given any positive constants $\lambda_{\min} < \lambda_{\max}$, there exists a constant C such that*

$$\mathbb{P}\left(\sup_{\lambda_{\min}\le\lambda\le\lambda_{\max}} \|\widehat{\boldsymbol{\beta}}(\lambda)\| \le C\sqrt{p}\right) \to 1.$$

PROOF OF LEMMA B.5. For simplicity, we omit the dependency of $\widehat{\boldsymbol{\beta}}$ on λ when clear from context. We first consider the case where $\delta < 1$. Write $\widehat{\boldsymbol{\beta}} = \mathcal{P}_1(\widehat{\boldsymbol{\beta}}) + \mathcal{P}_2(\widehat{\boldsymbol{\beta}})$, where $\mathcal{P}_1(\widehat{\boldsymbol{\beta}})$ is the projection of $\widehat{\boldsymbol{\beta}}$ onto the null space of $\boldsymbol{X}$ and $\mathcal{P}_2(\widehat{\boldsymbol{\beta}})$ is the projection of $\widehat{\boldsymbol{\beta}}$ onto the row space of $\boldsymbol{X}$. By the rotational invariance of i.i.d. Gaussian vectors, the null space of $\boldsymbol{X}$ is a random subspace of dimension $p-n = (1-\delta+o(1))p$ with uniform orientation. Since $\mathcal{P}_1(\widehat{\boldsymbol{\beta}})$ belongs to the null space, Kashin's Theorem (see Theorem F.1 in [1]) gives that with probability at least $1-2^{-p}$,

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}\|^2 &= \|\mathcal{P}_1(\widehat{\boldsymbol{\beta}})\|^2 + \|\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2 \\
&\le c_1\frac{\|\mathcal{P}_1(\widehat{\boldsymbol{\beta}})\|_1^2}{p} + \|\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2 \\
&\le 2c_1\frac{\|\widehat{\boldsymbol{\beta}}\|_1^2 + \|\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|_1^2}{p} + \|\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2 \\
&\le \frac{2c_1\|\widehat{\boldsymbol{\beta}}\|_1^2}{p} + (1+2c_1)\|\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2
\end{aligned} \tag{B.13}$$

for some constant c_1 depending only on δ; the first step uses Kashin's theorem, the second the triangle inequality, and the third Cauchy-Schwarz inequality. The smallest nonzero singular value of the Wishart matrix $\boldsymbol{X}^\top\boldsymbol{X}$ is concentrated at $(1/\sqrt{\delta}-1)^2$ with probability tending to one (see e.g. [4]). In addition, since $\mathcal{P}_2(\widehat{\boldsymbol{\beta}})$ belongs to the row space of $\boldsymbol{X}$, we have

$$\|\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2 \le c_2\|\boldsymbol{X}\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2$$

with probability approaching one. Above, c_2 can be chosen to be $(1/\sqrt{\delta} -$

$1)^{-2} + o(1)$. Set $c_3 = c_2(1 + 2c_1)$. Continuing (B.13) yields

$$
\begin{aligned}
\|\widehat{\boldsymbol{\beta}}\|^2 &\le \frac{2c_1\|\widehat{\boldsymbol{\beta}}\|_1^2}{p} + c_2(1+2c_1)\|\boldsymbol{X}\mathcal{P}_2(\widehat{\boldsymbol{\beta}})\|^2 \\
&= \frac{2c_1\|\widehat{\boldsymbol{\beta}}\|_1^2}{p} + c_3\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 \\
&\le \frac{2c_1\|\widehat{\boldsymbol{\beta}}\|_1^2}{p} + 2c_3\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + 2c_3\|\boldsymbol{y}\|^2 \\
&\le \frac{2c_1\left(\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}\|_1\right)^2}{\lambda^2 p} + 4c_3\left(\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}\|_1\right) + 2c_3\|\boldsymbol{y}\|^2 \\
&\le \frac{c_1\|\boldsymbol{y}\|^4}{2\lambda^2 p} + 4c_3\|\boldsymbol{y}\|^2,
\end{aligned}
$$

where in the last inequality we use the fact $\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}\|_1 \le \frac{1}{2}\|\boldsymbol{y}\|^2$. Thus, it suffices to bound $\|\boldsymbol{y}\|^2$. The largest singular value of $\boldsymbol{X}^\top\boldsymbol{X}$ is bounded above by $(1/\sqrt{\delta} + 1)^2 + o_{\mathbb{P}}(1)$. Therefore,

$$
\|\boldsymbol{y}\|^2 = \|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{z}\|^2 \le 2\|\boldsymbol{X}\boldsymbol{\beta}\|^2 + 2\|\boldsymbol{z}\|^2 \le c_4\|\boldsymbol{\beta}\|^2 + 2\|\boldsymbol{z}\|^2.
$$

Since both β_i and z_i have bounded second moments, the law of large numbers claims that there exists a constant c_5 such that $c_4\|\boldsymbol{\beta}\|^2 + 2\|\boldsymbol{z}\|^2 \le c_5 p$ with probability approaching one. Combining all the inequalities above gives

$$
\sup_{\lambda_{\min} \le \lambda \le \lambda_{\max}} \|\widehat{\boldsymbol{\beta}}(\lambda)\|^2 \le \frac{c_1 c_5^2 p}{2\lambda^2} + 2c_3 c_5 p \le \left(\frac{c_1 c_5^2}{2\lambda_{\min}^2} + 2c_3 c_5\right) p
$$

with probability converging to one.

In the case where $\delta > 1$, the null space of $\boldsymbol{X}$ reduces to $\mathbf{0}$, hence $\mathcal{P}_1(\widehat{\boldsymbol{\beta}}) = 0$. Therefore, this reduces to a special case of the above argument.

Now, we turn to work on the case where $\delta = 1$. We start with

$$
\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 \le 2\|\boldsymbol{y}\|^2 + 2\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2
$$

and

$$
\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}\|_1 \le \frac{1}{2}\|\boldsymbol{y}\|^2.
$$

These two inequalities give that simultaneously over all λ,

$$
\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda)\|^2 \le 4\|\boldsymbol{y}\|^2 \le 4c_5 p \tag{B.14a}
$$

$$
\|\widehat{\boldsymbol{\beta}}(\lambda)\|_1 \le \frac{1}{2\lambda_{\min}}\|\boldsymbol{y}\|^2 \le \frac{c_5 p}{2\lambda_{\min}} \tag{B.14b}
$$

with probability converging to one. Let M be any constant larger than 1.7×10^7. If $p\|\widehat{\boldsymbol{\beta}}\|^2/\|\widehat{\boldsymbol{\beta}}\|_1^2 < M$, by (B.14b), we get

$$\|\widehat{\boldsymbol{\beta}}\| \le \frac{\sqrt{M}c_5}{2\lambda_{\min}}\sqrt{p}. \tag{B.15}$$

Otherwise, denoting by $\mathcal{T}$ the set of indices $1 \le i \le p$ that correspond to the $s := \lceil p/100 \rceil$ largest $|\widehat{\beta}|_i$, from Lemma B.3 we have

$$\frac{\|\widehat{\boldsymbol{\beta}}_{\mathcal{T}}\|^2}{\|\widehat{\boldsymbol{\beta}}\|^2} \ge 1 - \frac{p^3}{Ms^3} \ge 1 - \frac{10^6}{M}. \tag{B.16}$$

To proceed, note that

$$\begin{aligned}
\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}\| &= \|\boldsymbol{X}_{\mathcal{T}}\widehat{\boldsymbol{\beta}}_{\mathcal{T}} + \boldsymbol{X}_{\overline{\mathcal{T}}}\widehat{\boldsymbol{\beta}}_{\overline{\mathcal{T}}}\| \\
&\ge \|\boldsymbol{X}_{\mathcal{T}}\widehat{\boldsymbol{\beta}}_{\mathcal{T}}\| - \|\boldsymbol{X}_{\overline{\mathcal{T}}}\widehat{\boldsymbol{\beta}}_{\overline{\mathcal{T}}}\| \\
&\ge \|\boldsymbol{X}_{\mathcal{T}}\widehat{\boldsymbol{\beta}}_{\mathcal{T}}\| - \sigma_{\max}(\boldsymbol{X})\|\widehat{\boldsymbol{\beta}}_{\overline{\mathcal{T}}}\|.
\end{aligned}$$

By Lemma B.4, we get $\|\boldsymbol{X}_{\mathcal{T}}\widehat{\boldsymbol{\beta}}_{\mathcal{T}}\| \ge \frac{1}{2}\|\boldsymbol{\beta}_{\mathcal{T}}\|$, and it is also clear that $\sigma_{\max}(\boldsymbol{X}) = 2 + o_{\mathbb{P}}(1)$. Thus, by (B.16) we obtain

$$\begin{aligned}
\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}\| \ge \|\boldsymbol{X}_{\mathcal{T}}\widehat{\boldsymbol{\beta}}_{\mathcal{T}}\| - \sigma_{\max}(\boldsymbol{X})\|\widehat{\boldsymbol{\beta}}_{\overline{\mathcal{T}}}\| &\ge \frac{1}{2}\|\widehat{\boldsymbol{\beta}}_{\mathcal{T}}\| - (2+o_{\mathbb{P}}(1))\|\widehat{\boldsymbol{\beta}}_{\overline{\mathcal{T}}}\| \\
&\ge \frac{1}{2}\sqrt{1-\frac{10^6}{M}}\|\widehat{\boldsymbol{\beta}}\| - (2+o_{\mathbb{P}}(1))\sqrt{\frac{10^6}{M}}\|\widehat{\boldsymbol{\beta}}\| \\
&= (c + o_{\mathbb{P}}(1))\|\widehat{\boldsymbol{\beta}}\|,
\end{aligned}$$

where $c = \frac{1}{2}\sqrt{1-10^6/M} - 2\sqrt{10^6/M} > 0$. Hence, owing to (B.14a),

$$\|\widehat{\boldsymbol{\beta}}\| \le \frac{(2+o_{\mathbb{P}}(1))\sqrt{c_5}}{c}\sqrt{p} \tag{B.17}$$

In summary, with probability tending to one, in either case, namely, (B.15) or (B.17),

$$\|\widehat{\boldsymbol{\beta}}\| \le C\sqrt{p}$$

for come constant C. This holds uniformly for all $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, and thus completes the proof.

□

Now, we conclude this section by proving our main lemma.

PROOF OF LEMMA B.2. The proof extensively applies Lemma 3.1[2] in [1] and Lemma B.5. Let $\boldsymbol{x}+\boldsymbol{r}=\widehat{\boldsymbol{\beta}}(\lambda)$ and $\boldsymbol{x}=\widehat{\boldsymbol{\beta}}(\lambda^-)$ be the notations in the statement of Lemma 3.1 in [1]. Among the fives assumptions needed in that lemma, it suffices to verify the first, third and fourth. Lemma B.5 asserts that

$$\sup_{\lambda^-\le\lambda\le\lambda^+}\|\boldsymbol{r}(\lambda)\|=\sup_{\lambda^-\le\lambda\le\lambda^+}\|\widehat{\boldsymbol{\beta}}(\lambda)-\widehat{\boldsymbol{\beta}}(\lambda^-)\|\le 2A\sqrt{p}$$

with probability approaching one. This fulfills the first assumption by taking $c_1=2A$. Next, let $\boldsymbol{g}(\lambda^-)\in\partial\|\widehat{\boldsymbol{\beta}}(\lambda^-)\|_1$ obey

$$\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda^-))=\lambda^-\boldsymbol{g}(\lambda^-).$$

Hence,

$$\left\|-\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda^-))+\lambda\boldsymbol{g}(\lambda^-)\right\|=(\lambda-\lambda^-)\|\boldsymbol{g}(\lambda^-)\|\le(\lambda^+-\lambda^-)\sqrt{p}$$

which certifies the third assumption. To verify the fourth assumption, taking $t\to\infty$ in Proposition 3.6 of [1] ensures the existence of constants c_2,c_3 and c_4 such that, with probability tending to one, $\sigma_{\min}(\boldsymbol{X}_{T\cup T'})\ge c_4$ for $T=\{j:|g_j(\lambda^-)|\ge 1-c_2\}$ and arbitrary $T'\subset\{1,\ldots,p\}$ with $|T'|\le c_3p$. Further, these constants can be independent of λ^- since $\lambda^-\in[\lambda_{\min},\lambda_{\max}]$ belongs to a compact interval. Therefore, this lemma concludes that, with probability approaching one,

$$\begin{aligned}\sup_{\lambda^-\le\lambda\le\lambda^+}\|\widehat{\boldsymbol{\beta}}(\lambda)-\widehat{\boldsymbol{\beta}}(\lambda^-)\| &\le \sup_{\lambda^-\le\lambda\le\lambda^+}\left(\sqrt{\lambda-\lambda^-}+\frac{\lambda-\lambda^-}{\lambda}\right)\xi\sqrt{p}\\ &\le\left(\sqrt{\lambda^+-\lambda^-}+\frac{\lambda^+-\lambda^-}{\lambda_{\min}}\right)\xi\sqrt{p}\\ &=O\left(\sqrt{(\lambda^+-\lambda^-)p}\right).\end{aligned}$$

This finishes the proof. □

APPENDIX C: OPTIMIZING THE TRADE-OFF

In this section, we still work under our working hypothesis and $\sigma>0$. Fixing δ and ϵ, we aim to show that no pairs below the boundary curve can be realized. Owing to the uniform convergence established in Appendix B, it is sufficient to study the range of $(\mathrm{tpp}^\infty(\lambda),\mathrm{fdp}^\infty(\lambda))$ introduced in Appendix A by varying $\Pi^\star$ and λ. To this end, we introduce a useful trick based on the following lemma.

[2]The conclusion of this lemma, $\|\boldsymbol{r}\|\le\sqrt{p}\xi(\epsilon,c_1,\ldots,c_5)$ in our notation, can be effortlessly strengthened to $\|\boldsymbol{r}\|\le(\sqrt{\epsilon}+\epsilon/\lambda)\xi(c_1,\ldots,c_5)\sqrt{p}$.

Lemma C.1. *For any fixed $\alpha > 0$, define a function $y = f(x)$ in the parametric form:*

$$x(t) = \mathbb{P}(|t + W| > \alpha)$$
$$y(t) = \mathbb{E}\left(\eta_\alpha(t + W) - t\right)^2$$

for $t \geq 0$, where W is a standard normal. Then f is strictly concave.

We use this to simplify the problem of detecting feasible pairs $(\mathrm{tpp}^\infty, \mathrm{fdp}^\infty)$. Denote by $\pi^\star := \Pi^\star/\tau$. Then (A.1) implies

$$\text{(C.1)} \qquad \begin{aligned} (1-\epsilon)\,\mathbb{E}\,\eta_\alpha(W)^2 + \epsilon\,\mathbb{E}(\eta_\alpha(\pi^\star + W) - \pi^\star)^2 &< \delta, \\ (1-\epsilon)\,\mathbb{P}(|W| > \alpha) + \epsilon\,\mathbb{P}(|\pi^\star + W| > \alpha) &< \min\{\delta, 1\}. \end{aligned}$$

We emphasize that (C.1) is not only necessary, but also sufficient in the following sense: given $0 < \delta_1 < \delta, 0 < \delta_2 < \min\{\delta, 1\}$ and $\pi^\star$, we can solve for τ by setting

$$(1-\epsilon)\,\mathbb{E}\,\eta_\alpha(W)^2 + \epsilon\,\mathbb{E}(\eta_\alpha(\pi^\star + W) - \pi^\star)^2 = \delta_1$$

and making use of the first line of (A.1), which can be alternatively written as

$$\tau^2 = \sigma^2 + \frac{\tau^2}{\delta}\left[(1-\epsilon)\,\mathbb{E}\,\eta_\alpha(W)^2 + \epsilon\,\mathbb{E}(\eta_\alpha(\pi^\star + W) - \pi^\star)^2\right].$$

($\Pi^\star = \tau\pi^\star$ is also determined, so does Π),[3] and along with

$$(1-\epsilon)\,\mathbb{P}(|W| > \alpha) + \epsilon\,\mathbb{P}(|\pi^\star + W| > \alpha) = \delta_2,$$

λ is also uniquely determined.

Since (C.1) is invariant if $\pi^\star$ is replaced by $|\pi^\star|$, we assume $\pi^\star \geq 0$ without loss of generality. As a function of t, $\mathbb{P}(|t + W| > \alpha)$ attains the minimum $\mathbb{P}(|W| > \alpha) = 2\Phi(-\alpha)$ at $t = 0$, and the supremum equal to one at $t = \infty$. Hence, there must exist $\epsilon' \in (0, 1)$ obeying

$$\text{(C.2)} \qquad \mathbb{P}(|\pi^\star + W| > \alpha) = (1-\epsilon')\,\mathbb{P}(|W| > \alpha) + \epsilon'.$$

As a consequence, the predicted TPP and FDP can be alternatively expressed as

$$\text{(C.3)} \qquad \mathrm{tpp}^\infty = 2(1-\epsilon')\Phi(-\alpha) + \epsilon', \quad \mathrm{fdp}^\infty = \frac{2(1-\epsilon)\Phi(-\alpha)}{2(1-\epsilon\epsilon')\Phi(-\alpha) + \epsilon\epsilon'}.$$

[3]Not every pair $(\delta_1, \delta_2) \in (0, \delta) \times (0, \min\{\delta, 1\})$ is feasible below the Donoho-Tanner phase transition. Nevertheless, this does not affect our discussion.

Compared to the original formulation, this expression is preferred since it only involves scalars.

Now, we seek equations that govern ϵ' and α, given δ and ϵ. Since both $\mathbb{E}(\eta_\alpha(t+W)-t)^2$ and $\mathbb{P}(|t+W|>\alpha)$ are monotonically increasing with respect to $t \ge 0$, there exists a function f obeying

$$\mathbb{E}(\eta_\alpha(t+W)-t)^2 = f\left(\mathbb{P}(|t+W|>\alpha)\right).$$

Lemma C.1 states that f is concave. Then

$$\mathbb{E}(\eta_\alpha(\pi^\star+W)-\pi^\star)^2 = f\left(\mathbb{P}(|\pi^\star+W|>\alpha\right)$$

and (C.2) allows us to view the argument of f in the right-hand side as an average of a random variable taking value $\mathbb{P}(|W|>\alpha)$ with probability $1-\epsilon'$ and value one with probability ϵ'. Therefore, Jensen's inequality states that

$$\mathbb{E}(\eta_\alpha(\pi^\star+W)-\pi^\star)^2 \ge (1-\epsilon')f\left(\mathbb{P}(|W|>\alpha)\right)+\epsilon' f(1) = (1-\epsilon')\,\mathbb{E}\,\eta_\alpha(W)^2+\epsilon'(\alpha^2+1).$$

Combining this with (C.1) gives

$$(1-\epsilon\epsilon')\,\mathbb{E}\,\eta_\alpha(W)^2+\epsilon\epsilon'(\alpha^2+1) < \delta, \tag{C.4a}$$

$$(1-\epsilon\epsilon')\,\mathbb{P}(|W|>\alpha)+\epsilon\epsilon' < \min\{\delta,1\}. \tag{C.4b}$$

Similar to (C.1), (C.4) is also sufficient in the same sense, and (C.4b) is automatically satisfied if $\delta > 1$.

The remaining part of this section studies the range of $(\mathrm{tpp}^\infty, \mathrm{fdp}^\infty)$ given by (C.3) under the constraints (C.4). Before delving into the details, we remark that this reduction of $\pi^\star$ to a two-point prior is realized by setting $\pi^\star = \infty$ (equivalently $\Pi^\star = \infty$) with probability ϵ' and otherwise $+0$, where $+0$ is considered to be nonzero. Though this prior is not valid since the working hypothesis requires a finite second moment, it can nevertheless be approximated by a sequence of instances, please see the example given in Section 2.

The lemma below recognizes that for certain (δ,ϵ) pairs, the TPP is asymptotically bounded above away from 1.

LEMMA C.2. *Put*

$$u^\star(\delta,\epsilon) := \begin{cases} 1-\frac{(1-\delta)(\epsilon-\epsilon^\star)}{\epsilon(1-\epsilon^\star)}, & \delta<1 \text{ and } \epsilon>\epsilon^\star(\delta),\\ 1, & \text{otherwise.}\end{cases}$$

Then

$$\mathrm{tpp}^\infty < u^\star(\delta,\epsilon).$$

Moreover, tpp^∞ *can be arbitrarily close to* $u^\star$.

This lemma directly implies that above the Donoho-Tanner phase transition (i.e. $\delta < 1$ and $\epsilon > \epsilon^\star(\delta)$), there is a fundamental limit on the TPP for arbitrarily strong signals. Consider

$$2(1-\epsilon)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon(1+t^2) = \delta. \tag{C.5}$$

For $\delta < 1$, $\epsilon^\star$ is the only positive constant in $(0,1)$ such that (C.5) with $\epsilon = \epsilon^\star$ has a unique positive root. Alternatively, the function $\epsilon^\star = \epsilon^\star(\delta)$ is implicitly given in the following parametric form:

$$\delta = \frac{2\phi(t)}{2\phi(t) + t(2\Phi(t) - 1)}$$
$$\epsilon^\star = \frac{2\phi(t) - 2t\Phi(-t)}{2\phi(t) + t(2\Phi(t) - 1)}$$

for $t > 0$, from which we see that $\epsilon^\star < \delta < 1$. Take the sparsity level k such as $\epsilon^\star p < k < \delta p = n$, from which we have $u^\star < 1$. As a result, the Lasso is unable to select all the k true signals even when the signal strength is arbitrarily high. This is happening even though the Lasso has the chance to select up to $n > k$ variables.

Any u between 0 and $u^\star$ (non-inclusive) can be realized as tpp^∞. Recall that we denote by $t^\star(u)$ the unique root in (α_0, ∞) (α_0 is the root of $(1+t^2)\Phi(-t) - t\phi(t) = \delta/2$) to the equation

$$\frac{2(1-\epsilon)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon(1+t^2) - \delta}{\epsilon\left[(1+t^2)(1-2\Phi(-t)) + 2t\phi(t)\right]} = \frac{1-u}{1-2\Phi(-t)}. \tag{C.6}$$

For a proof of this fact, we refer to Lemma C.4. Last, recall that

$$q^\star(u; \delta, \epsilon) = \frac{2(1-\epsilon)\Phi(-t^\star(u))}{2(1-\epsilon)\Phi(-t^\star(u)) + \epsilon u}.$$

We can now state the fundamental trade-off between fdp^∞ and tpp^∞.

LEMMA C.3. *If* $\mathrm{tpp}^\infty \geq u$ *for* $u \in (0, u^\star)$, *then*

$$\mathrm{fdp}^\infty > q^\star(u).$$

In addition, fdp^∞ *can be arbitrarily close to* $q^\star(u)$.

C.1. Proofs of Lemmas C.1, C.2 and C.3.

Proof of Lemma C.1. First of all, f is well-defined since both $x(t)$ and $y(t)$ are strictly increasing functions of t. Note that

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \phi(\alpha - t) - \phi(-\alpha - t), \quad \frac{\mathrm{d}y}{\mathrm{d}t} = 2t\left[\Phi(\alpha - t) - \Phi(-\alpha - t)\right].$$

Applying the chain rule gives

$$\begin{aligned} f'(t) &= \frac{\mathrm{d}y}{\mathrm{d}t}\Big/\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{2t\left[\Phi(\alpha - t) - \Phi(-\alpha - t)\right]}{\phi(\alpha - t) - \phi(-\alpha - t)} \\ &= \frac{2t\int_{-\alpha-t}^{\alpha-t} \mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u}{\mathrm{e}^{-\frac{(\alpha-t)^2}{2}} - \mathrm{e}^{-\frac{(-\alpha-t)^2}{2}}} = \frac{2t\int_{-\alpha}^{\alpha} \mathrm{e}^{-\frac{(u-t)^2}{2}}\mathrm{d}u}{\mathrm{e}^{-\frac{\alpha^2+t^2-2\alpha t}{2}} - \mathrm{e}^{-\frac{\alpha^2+t^2+2\alpha t}{2}}} \\ &= \frac{2t\mathrm{e}^{\frac{\alpha^2}{2}}\int_{-\alpha}^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}\mathrm{e}^{tu}\mathrm{d}u}{\mathrm{e}^{\alpha t} - \mathrm{e}^{-\alpha t}} = \frac{2\mathrm{e}^{\frac{\alpha^2}{2}}\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}(\mathrm{e}^{tu} + \mathrm{e}^{-tu})\mathrm{d}u}{\int_0^{\alpha} \mathrm{e}^{tu} + \mathrm{e}^{-tu}\mathrm{d}u}. \end{aligned}$$

Since $x(t)$ is strictly increasing in t, we see that $f''(t) \le 0$ is equivalent to saying that the function

$$g(t) := \frac{\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}(\mathrm{e}^{tu} + \mathrm{e}^{-tu})\mathrm{d}u}{\int_0^{\alpha} \mathrm{e}^{tu} + \mathrm{e}^{-tu}\mathrm{d}u} \equiv \frac{\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}\cosh(tu)\mathrm{d}u}{\int_0^{\alpha}\cosh(tu)\mathrm{d}u}$$

is decreasing in t. Hence, it suffices to show that

$$g'(t) = \frac{\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}u\sinh(tu)\mathrm{d}u\int_0^{\alpha}\cosh(tv)\mathrm{d}v - \int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}\cosh(tu)\mathrm{d}u\int_0^{\alpha} v\sinh(tv)\mathrm{d}v}{\left(\int_0^{\alpha}\cosh(tv)\mathrm{d}v\right)^2} \le 0.$$

Observe that the numerator is equal to

$$\begin{aligned} &\int_0^{\alpha}\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}u\sinh(tu)\cosh(tv)\mathrm{d}u\mathrm{d}v - \int_0^{\alpha}\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}v\cosh(tu)\sinh(tv)\mathrm{d}u\mathrm{d}v \\ &\qquad = \int_0^{\alpha}\int_0^{\alpha} \mathrm{e}^{-\frac{u^2}{2}}\left(u\sinh(tu)\cosh(tv) - v\cosh(tu)\sinh(tv)\right)\mathrm{d}u\mathrm{d}v \\ &\qquad \overset{u\leftrightarrow v}{=} \int_0^{\alpha}\int_0^{\alpha} \mathrm{e}^{-\frac{v^2}{2}}\left(v\sinh(tv)\cosh(tu) - u\cosh(tv)\sinh(tu)\right)\mathrm{d}v\mathrm{d}u \\ &\qquad = \frac{1}{2}\int_0^{\alpha}\int_0^{\alpha}(\mathrm{e}^{-\frac{u^2}{2}} - \mathrm{e}^{-\frac{v^2}{2}})\left(u\sinh(tu)\cosh(tv) - v\cosh(tu)\sinh(tv)\right)\mathrm{d}u\mathrm{d}v. \end{aligned}$$

Then it is sufficient to show that

$$(\mathrm{e}^{-\frac{u^2}{2}} - \mathrm{e}^{-\frac{v^2}{2}})\left(u\sinh(tu)\cosh(tv) - v\cosh(tu)\sinh(tv)\right) \le 0$$

for all $u, v, t \geq 0$. To see this, suppose $u \geq v$ without loss of generality so that $\mathrm{e}^{-\frac{u^2}{2}} - \mathrm{e}^{-\frac{v^2}{2}} \leq 0$ and

$$\begin{aligned} u \sinh(tu)\cosh(tv) - v\cosh(tu)\sinh(tv) &\geq v(\sinh(tu)\cosh(tv) - \cosh(tu)\sinh(tv)) \\ &= v\sinh(tu - tv) \geq 0. \end{aligned}$$

This analysis further reveals that $f''(t) < 0$ for $t > 0$. Hence, f is strictly concave. □

To prove the other two lemmas, we collect some useful facts about (C.5). This equation has (a) one positive root for $\delta \geq 1$ or $\delta < 1, \epsilon = \epsilon^\star$, (b) two positive roots for $\delta < 1$ and $\epsilon < \epsilon^\star$, and (c) no positive root if $\delta < 1$ and $\epsilon > \epsilon^\star$. In the case of (a) and (b), call $t(\epsilon, \delta)$ the positive root of (C.5) (choose the larger one if there are two). Then $t(\epsilon, \delta)$ is a decreasing function of ϵ. In particular, $t(\epsilon, \delta) \to \infty$ as $\epsilon \to 0$. In addition, $2(1-\epsilon)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon(1+t^2) > \delta$ if $t > t(\epsilon, \delta)$.

LEMMA C.4. *For any $0 < u < u^\star$, (C.6) has a unique root, denoted by $t^\star(u)$, in (α_0, ∞). In addition, $t^\star(u)$ strictly decreases as u increases, and it further obeys $0 < (1-u)/(1-2\Phi(-t^\star(u))) < 1$.*

LEMMA C.5. *As a function of u,*

$$q^\star(u) = \frac{2(1-\epsilon)\Phi(-t^\star(u))}{2(1-\epsilon)\Phi(-t^\star(u)) + \epsilon u}$$

is strictly increasing on $(0, u^\star)$.

PROOF OF LEMMA C.2. We first consider the regime: $\delta < 1, \epsilon > \epsilon^\star$. By (C.3), it is sufficient to show that $\mathrm{tpp}^\infty = 2(1-\epsilon')\Phi(-\alpha) + \epsilon' < u^\star$ under the constraints (C.4). From (C.4b) it follows that

$$\Phi(-\alpha) = \frac{1}{2}\mathbb{P}(|W| > \alpha) < \frac{\delta - \epsilon\epsilon'}{2(1-\epsilon\epsilon')},$$

which can be rearranged as

$$2(1-\epsilon')\Phi(-\alpha) + \epsilon' < \frac{(1-\epsilon')(\delta - \epsilon\epsilon')}{1-\epsilon\epsilon'} + \epsilon'.$$

The right-hand side is an increasing function of ϵ' because its derivative is equal to $(1-\epsilon)(1-\delta)/(1-\epsilon\epsilon')^2$ and is positive. Since the range of ϵ' is $(0, \epsilon^\star/\epsilon)$, we get

$$2(1-\epsilon')\Phi(-\alpha) + \epsilon' < \frac{(1-\epsilon^\star/\epsilon)(\delta - \epsilon \cdot \epsilon^\star/\epsilon)}{1 - \epsilon \cdot \epsilon^\star/\epsilon} + \epsilon^\star/\epsilon = u^\star.$$

This bound $u^\star$ can be arbitrarily approached: let $\epsilon' = \epsilon^\star/\epsilon$ in the example given in Section 2.5; then set $\lambda = \sqrt{M}$ and take $M \to \infty$.

We turn our attention to the easier case where $\delta \geq 1$, or $\delta < 1$ and $\epsilon \leq \epsilon^\star$. By definition, the upper limit $u^\star = 1$ trivially holds. It remains to argue that tpp^∞ can be arbitrarily close to 1. To see this, set $\Pi^\star = M$ almost surely, and take the same limits as before: then $\mathrm{tpp}^\infty \to 1$.

□

Proof of Lemma C.3. We begin by first considering the boundary case:

$$\mathrm{tpp}^\infty = u. \tag{C.7}$$

In view of (C.3), we can write

$$\mathrm{fdp}^\infty = \frac{2(1-\epsilon)\Phi(-\alpha)}{2(1-\epsilon)\Phi(-\alpha) + \epsilon\mathrm{tpp}^\infty} = \frac{2(1-\epsilon)\Phi(-\alpha)}{2(1-\epsilon)\Phi(-\alpha) + \epsilon u}.$$

Therefore, a lower bound on fdp^∞ is equivalent to maximizing α under the constraints (C.4) and (C.7).

Recall $\mathbb{E}\,\eta_\alpha(W)^2 = 2(1+\alpha^2)\Phi(-\alpha) - 2\alpha\phi(\alpha)$. Then from (C.7) and (C.4a) we obtain a sandwiching expression for $1-\epsilon'$:

$$\frac{(1-\epsilon)\left[2(1+\alpha^2)\Phi(-\alpha) - 2\alpha\phi(\alpha)\right] + \epsilon(1+\alpha^2) - \delta}{\epsilon\left[(1+\alpha^2)(1-2\Phi(-\alpha)) + 2\alpha\phi(\alpha)\right]} < 1-\epsilon' = \frac{1-u}{1-2\Phi(-\alpha)},$$

which implies

$$\frac{(1-\epsilon)\left[2(1+\alpha^2)\Phi(-\alpha) - 2\alpha\phi(\alpha)\right] + \epsilon(1+\alpha^2) - \delta}{\epsilon\left[(1+\alpha^2)(1-2\Phi(-\alpha)) + 2\alpha\phi(\alpha)\right]} - \frac{1-u}{1-2\Phi(-\alpha)} < 0.$$

The left-hand side of this display tends to $1-(1-u) = u > 0$ as $\alpha \to \infty$, and takes on the value 0 if $\alpha = t^\star(u)$. Hence, by the uniqueness of $t^\star(u)$ provided by Lemma C.4, we get $\alpha < t^\star(u)$. In conclusion,

$$\mathrm{fdp}^\infty = \frac{2(1-\epsilon)\Phi(-\alpha)}{2(1-\epsilon)\Phi(-\alpha) + \epsilon u} > \frac{2(1-\epsilon)\Phi(-t^\star(u))}{2(1-\epsilon)\Phi(-t^\star(u)) + \epsilon u} = q^\star(u). \tag{C.8}$$

It is easy to see that fdp^∞ can be arbitrarily close to $q^\star(u)$.

To finish the proof, we proceed to consider the general case $\mathrm{tpp}^\infty = u' > u$. The previous discussion clearly remains valid, and hence (C.8) holds if u is replaced by u'; that is, we have

$$\mathrm{fdp}^\infty > q^\star(u').$$

By Lemma C.5, it follows from the monotonicity of $q^\star(\cdot)$ that $q^\star(u') > q^\star(u)$. Hence, $\mathrm{fdp}^\infty > q^\star(u)$, as desired.

□

C.2. Proofs of auxiliary lemmas.

PROOF OF LEMMA C.4. Set

$$\zeta := 1 - \frac{2(1-\epsilon)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon(1+t^2) - \delta}{\epsilon\left[(1+t^2)(1-2\Phi(-t)) + 2t\phi(t)\right]}$$

or, equivalently,

$$2(1-\epsilon\zeta)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon\zeta(1+t^2) = \delta. \tag{C.9}$$

As in Section C.1, we abuse notation a little and let $t(\zeta) = t(\epsilon\zeta, \delta)$ denote the (larger) positive root of (C.9). Then the discussion about (C.5) in Section C.1 shows that $t(\zeta)$ decreases as ζ increases. Note that in the case where $\delta < 1$ and $\epsilon > \epsilon^\star(\delta)$, the range of ζ in (C.9) is assumed to be $(0, \epsilon^\star/\epsilon)$, since otherwise (C.9) does not have a positive root (by convention, set $\epsilon^\star(\delta) = 1$ if $\delta > 1$).

Note that (C.6) is equivalent to

$$u = 1 - \frac{2(1-\epsilon)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon(1+t^2) - \delta}{\epsilon\left[(1+t^2)(1-2\Phi(-t)) + 2t\phi(t)\right]/(1-2\Phi(-t))}. \tag{C.10}$$

Define

$$h(\zeta) = 1 - \frac{2(1-\epsilon)\left[(1+t(\zeta)^2)\Phi(-t(\zeta)) - t(\zeta)\phi(t(\zeta))\right] + \epsilon(1+t(\zeta)^2) - \delta}{\epsilon\left[(1+t(\zeta)^2)(1-2\Phi(-t(\zeta))) + 2t(\zeta)\phi(t(\zeta))\right]/(1-2\Phi(-t(\zeta)))}.$$

In view of (C.9) and (C.10), the proof of this lemma would be completed once we show the existence of ζ such that $h(\zeta) = u$. Now we prove this fact.

On the one hand, as $\zeta \searrow 0$, we see $t(\zeta) \nearrow \infty$. This leads to

$$h(\zeta) \to 0.$$

On the other hand, if $\zeta \nearrow \min\{1, \epsilon^\star/\epsilon\}$, then $t(\zeta)$ converges to $t^\star(u^\star) > \alpha_0$, which satisfies

$$2(1-\min\{\epsilon,\epsilon^\star\})\left[(1+t^{\star 2})\Phi(-t^\star) - t^\star\phi(t^\star)\right] + \min\{\epsilon,\epsilon^\star\}(1+t^{\star 2}) = \delta.$$

Consequently, we get

$$h(\zeta) \to u^\star.$$

Therefore, by the continuity of $h(\zeta)$, for any $u \in (0, u^\star)$ we can find $0 < \epsilon' < \min\{1, \epsilon^\star/\epsilon\}$ such that $h(\epsilon') = u$. Put $t^\star(u) = t(\epsilon')$. We have

$$\frac{1-u}{1-2\Phi(-t^\star(u))} = 1 - \epsilon' < 1.$$

Last, to prove the uniqueness of $t^\star(u)$ and its monotonically decreasing dependence on u, it suffices to ensure that (a) $t(\zeta)$ is a decreasing function of ζ, and (b) $h(\zeta)$ is an increasing function of ζ. As seen above, (a) is true, and (b) is also true as can be seen from writing h as $h(\zeta) = 2(1-\zeta)\Phi(-t(\zeta)) + \zeta$, which is an increasing function of ζ.

□

PROOF OF LEMMA C.5. Write

$$q^\star(u) = \frac{2(1-\epsilon)}{2(1-\epsilon) + \epsilon u/\Phi(-t^\star(u))}.$$

This suggests that the lemma amounts to saying that $u/\Phi(-t^\star(u))$ is a decreasing function of u. From (C.10), we see that this function is equal to

$$\frac{1}{\Phi(-t^\star(u))} - \frac{(1-2\Phi(-t^\star(u)))\,\{2(1-\epsilon)\left[(1+(t^\star(u))^2)\Phi(-t^\star(u)) - t^\star(u)\phi(t^\star(u))\right] + \epsilon(1+(t^\star(u))^2) - \delta}{\epsilon\Phi(-t^\star(u))\left[(1+(t^\star(u))^2)(1-2\Phi(-t^\star(u))) + 2t^\star(u)\phi(t^\star(u))\right]}$$

With the proviso that $t^\star(u)$ is decreasing in u, it suffices to show that

$$\begin{aligned} &\frac{1}{\Phi(-t)} - \frac{(1-2\Phi(-t))\left\{2(1-\epsilon)\left[(1+t^2)\Phi(-t) - t\phi(t)\right] + \epsilon(1+t^2) - \delta\right\}}{\epsilon\Phi(-t)\left[(1+t^2)(1-2\Phi(-t)) + 2t\phi(t)\right]} \\ &= \frac{\delta}{\epsilon}\cdot\underbrace{\frac{1-2\Phi(-t)}{\Phi(-t)\left[(1+t^2)(1-2\Phi(-t)) + 2t\phi(t)\right]}}_{f_1(t)} - \frac{2}{\epsilon}\cdot\underbrace{\frac{(1-2\Phi(-t))\left[(1+t^2)\Phi(-t) - t\phi(t)\right]}{\Phi(-t)\left[(1+t^2)(1-2\Phi(-t)) + 2t\phi(t)\right]}}_{f_2(t)} + 2 \end{aligned}$$

is an increasing function of $t > 0$. Simple calculations show that f_1 is increasing while f_2 is decreasing over $(0, \infty)$. This finishes the proof. □

APPENDIX D: PROOF OF THEOREM 2.1

With the results given in Appendices B and C in place, we are ready to characterize the optimal false/true positive rate trade-off. Up until now, the results hold for bounded λ, and we thus need to extend the results to arbitrarily large λ. It is intuitively easy to conceive that the support size of $\widehat{\boldsymbol{\beta}}$ will be small with a very large λ, resulting in low power. The following lemma, whose proof constitutes the subject of Section D.1, formalizes this point. In this section, $\sigma \geq 0$ may take on the value zero. Also, we work with $\lambda_0 = 0.01$ and $\eta = 0.001$ to carry fewer mathematical symbols; any other numerical values would clearly work.

LEMMA D.1. *For any $c > 0$, there exists λ' such that*

$$\sup_{\lambda > \lambda'} \frac{\#\{j : \widehat{\beta}_j(\lambda) \neq 0\}}{p} \leq c$$

holds with probability converging to one.

Assuming the conclusion of Lemma D.1, we prove claim (b) in Theorem 1 (noisy case), and then (a) (noiseless case). (c) is a simple consequence of (a) and (b), and (d) follows from Appendix C.

Case $\sigma > 0$. Let c be sufficiently small such that $q^\star(c/\epsilon) < 0.001$. Pick a large enough λ' such that Lemma D.1 holds. Then with probability tending to one, for all $\lambda > \lambda'$, we have

$$\mathrm{TPP}(\lambda) = \frac{T(\lambda)}{k \vee 1} \leq (1 + o_{\mathbb{P}}(1)) \frac{\#\{j : \widehat{\beta}_j(\lambda) \neq 0\}}{\epsilon p} \leq (1 + o_{\mathbb{P}}(1)) \frac{c}{\epsilon}.$$

On this event, we get

$$q^\star(\mathrm{TPP}(\lambda)) - 0.001 \leq q^\star(c/\epsilon + o_{\mathbb{P}}(1)) - 0.001 \leq 0,$$

which implies that

$$\bigcap_{\lambda > \lambda'} \left\{ \mathrm{FDP}(\lambda) \geq q^\star\left(\mathrm{TPP}(\lambda)\right) - 0.001 \right\} \tag{D.1}$$

holds with probability approaching one.

Now we turn to work on the range $[0.01, \lambda']$. By Lemma A.2, we get that $V(\lambda)/p$ (resp. $T(\lambda)/p$) converges in probability to $\mathrm{fd}^\infty(\lambda)$ (resp. $\mathrm{td}^\infty(\lambda)$) uniformly over $[0.01, \lambda']$. As a consequence,

$$\mathrm{FDP}(\lambda) = \frac{V(\lambda)}{\max\{V(\lambda) + T(\lambda), 1\}} \xrightarrow{\mathbb{P}} \frac{\mathrm{fd}^\infty(\lambda)}{\mathrm{fd}^\infty(\lambda) + \mathrm{td}^\infty(\lambda)} = \mathrm{fdp}^\infty(\lambda) \tag{D.2}$$

uniformly over $\lambda \in [0.01, \lambda']$. The same reasoning also justifies that

$$\mathrm{TPP}(\lambda) \xrightarrow{\mathbb{P}} \mathrm{tpp}^\infty(\lambda) \tag{D.3}$$

uniformly over $\lambda \in [0.01, \lambda']$. From Lemma C.3 it follows that

$$\mathrm{fdp}^\infty(\lambda) > q^\star(\mathrm{tpp}^\infty(\lambda)).$$

Hence, by the continuity of $q^\star(\cdot)$, combining (D.2) with (D.3) gives that

$$\mathrm{FDP}(\lambda) \geq q^\star(\mathrm{TPP}(\lambda)) - 0.001$$

holds simultaneously for all $\lambda \in [0.01, \lambda']$ with probability tending to one. This concludes the proof.

Case $\sigma = 0$. Fix λ and let $\sigma > 0$ be sufficiently small. We first prove that Lemma A.1 still holds for $\sigma = 0$ if α and τ are taken to be the limiting solution to (A.1) with $\sigma \to 0$, denoted by α' and τ'. Introduce $\widehat{\boldsymbol{\beta}}^\sigma$ to be the Lasso solution with data $\boldsymbol{y}^\sigma := \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{z} = \boldsymbol{y} + \boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ is independent of $\boldsymbol{X}$ and $\boldsymbol{\beta}$. Our proof strategy is based on the approximate equivalence between $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^\sigma$.

It is well known that the Lasso residuals $\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ are obtained by projecting the response $\boldsymbol{y}$ onto the polytope $\{\boldsymbol{r} : \|\boldsymbol{X}^\top \boldsymbol{r}\|_\infty \le \lambda\}$. The non-expansive property of projections onto convex sets gives

$$\left\|(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma) - (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\right\| \le \|\boldsymbol{y}^\sigma - \boldsymbol{y}\| = \|\boldsymbol{z}\|.$$

If $\boldsymbol{P}(\cdot)$ is the projection onto the polytope, then $\boldsymbol{I} - \boldsymbol{P}$ is also non-expansive and, therefore,

$$\left\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\| \le \|\boldsymbol{z}\|. \tag{D.4}$$

Hence, from Lemma B.1 and $\|\boldsymbol{z}\| = (1 + o_{\mathbb{P}}(1))\sigma\sqrt{n}$ it follows that, for any $c > 0$ and r_c depending on c,

$$\#\{1 \le j \le p : |\boldsymbol{X}_j^\top(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma - \boldsymbol{y} + \boldsymbol{X}\widehat{\boldsymbol{\beta}})| > 2r_c\sigma\} \le cp \tag{D.5}$$

holds with probability converging to one. Let $\boldsymbol{g}$ and $\boldsymbol{g}^\sigma$ be subgradients certifying the KKT conditions for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^\sigma$. From

$$\boldsymbol{X}_j^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \lambda g_j,$$
$$\boldsymbol{X}_j^\top(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma) = \lambda g_j^\sigma,$$

we get a simple relationship:

$$\{j : |g_j| \ge 1 - a/2\} \setminus \{j : |g_j^\sigma| \ge 1 - a/2 - 2r_c\sigma/\lambda\} \subseteq \{j : |\boldsymbol{X}_j^\top(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma - \boldsymbol{y} + \boldsymbol{X}\widehat{\boldsymbol{\beta}})| > 2r_c\sigma\}.$$

Choose σ sufficiently small such that $2r_c\sigma/\lambda < a/2$, that is, $\sigma < a\lambda/(4r_c)$. Then

$$\{j : |g_j| \ge 1 - a/2\} \setminus \{j : |g_j^\sigma| \ge 1 - a\} \subseteq \{j : |\boldsymbol{X}_j^\top(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma - \boldsymbol{y} + \boldsymbol{X}\widehat{\boldsymbol{\beta}})| > 2r_c\sigma\}. \tag{D.6}$$

As earlier, denote by $\mathcal{S} = \operatorname{supp}(\widehat{\boldsymbol{\beta}})$ and $\mathcal{S}^\sigma = \operatorname{supp}(\widehat{\boldsymbol{\beta}}^\sigma)$. In addition, let $\mathcal{S}_v = \{j : |g_j| \ge 1 - v\}$ and similarly $\mathcal{S}_v^\sigma = \{j : |g_j^\sigma| \ge 1 - v\}$. Notice that we have dropped the dependence on λ since λ is fixed. Continuing, since $\mathcal{S} \subseteq \mathcal{S}_{\frac{a}{2}}$, from (D.6) we obtain

$$\mathcal{S} \setminus \mathcal{S}_a^\sigma \subseteq \{j : |\boldsymbol{X}_j^\top(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma - \boldsymbol{y} + \boldsymbol{X}\widehat{\boldsymbol{\beta}})| > 2r_c\sigma\}. \tag{D.7}$$

This suggests that we apply Proposition 3.6 of [1] that claims[4] the existence of positive constants $a_1 \in (0,1), a_2$, and a_3 such that with probability tending to one,

$$\sigma_{\min}(\boldsymbol{X}_{\mathcal{S}^\sigma_{a_1} \cup \mathcal{S}'}) \geq a_3 \tag{D.8}$$

for all $|\mathcal{S}'| \leq a_2 p$. These constants also have positive limits a_1', a_2', a_3', respectively, as $\sigma \to 0$. We take $a < a_1', c < a_2'$ (we will specify a, c later) and sufficiently small σ in (D.7), and $\mathcal{S}' = \{j : |\boldsymbol{X}_j^\top(\boldsymbol{y}^\sigma - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma - \boldsymbol{y} + \boldsymbol{X}\widehat{\boldsymbol{\beta}})| > 2r_c\sigma\}$. Hence, on this event, (D.5), (D.7), and (D.8) together give

$$\|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^\sigma\| = \left\|\boldsymbol{X}_{\mathcal{S}^\sigma_a \cup \mathcal{S}'}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}^\sigma_a \cup \mathcal{S}'} - \widehat{\boldsymbol{\beta}}^\sigma_{\mathcal{S}^\sigma_a \cup \mathcal{S}'})\right\| \geq a_3 \|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^\sigma\|$$

for sufficiently small σ, which together with (D.4) yields

$$\|\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}^\sigma\| \leq \frac{(1 + o_{\mathbb{P}}(1))\sigma\sqrt{n}}{a_3}. \tag{D.9}$$

Recall that the $\|\widehat{\boldsymbol{\beta}}\|_0$ is the number of nonzero entries in the vector $\widehat{\boldsymbol{\beta}}$. From (D.9), using the same argument outlined in (B.8), (B.9), and (B.10), we have

$$\|\widehat{\boldsymbol{\beta}}\|_0 \geq \|\widehat{\boldsymbol{\beta}}^\sigma\|_0 - c'p + o_{\mathbb{P}}(p) \tag{D.10}$$

for some constant $c' > 0$ that decreases to 0 as $\sigma/a_3 \to 0$.

We now develop a tight upper bound on $\|\widehat{\boldsymbol{\beta}}\|_0$. Making use of (D.7) gives

$$\|\widehat{\boldsymbol{\beta}}\|_0 \leq \|\widehat{\boldsymbol{\beta}}^\sigma\|_0 + \#\{j : 1 - a \leq |g_j^\sigma| < 1\} + cp.$$

As in (B.5), (3.21) of [1] implies that

$$\#\left\{j : (1-a) \leq |g_j^\sigma| < 1\right\}/p \xrightarrow{\mathbb{P}} \mathbb{P}\left((1-a)\alpha\tau \leq |\Pi + \tau W| < \alpha\tau\right).$$

Note that both α and τ depend on σ, and as $\sigma \to 0$, α and τ converge to, respectively, $\alpha' > 0$ and $\tau' > 0$. Hence, we get

$$\|\widehat{\boldsymbol{\beta}}\|_0 \leq \|\widehat{\boldsymbol{\beta}}^\sigma\|_0 + c''p + o_{\mathbb{P}}(p) \tag{D.11}$$

for some constant $c'' > 0$ that can be made arbitrarily small if $\sigma \to 0$ by first taking a and c sufficiently small.

With some obvious notation, a combination of (D.10) and (D.11) gives

$$|V - V^\sigma| \leq c'''p, \quad |T - T^\sigma| \leq c'''p, \tag{D.12}$$

[4] Use a continuity argument to carry over the result of this proposition for finite t to ∞.

for some constant $c''' = c'''_\sigma \to 0$ as $\sigma \to 0$. As $\sigma \to 0$, observe the convergence,

$$\mathrm{fd}^{\infty,\sigma} = 2(1-\epsilon)\Phi(-\alpha) \to 2(1-\epsilon)\Phi(-\alpha') = \mathrm{fd}^{\infty,0},$$

and

$$\mathrm{td}^{\infty,\sigma} = \epsilon\,\mathbb{P}(|\Pi^\star + \tau W| > \alpha\tau) \to \epsilon\,\mathbb{P}(|\Pi^\star + \tau' W| > \alpha'\tau') = \mathrm{td}^{\infty,0}.$$

By applying Lemma A.1 to V^σ and T^σ and making use of (D.12), the conclusions

$$\frac{V}{p} \xrightarrow{\mathbb{P}} \mathrm{fd}^{\infty,0} \quad \text{and} \quad \frac{T}{p} \xrightarrow{\mathbb{P}} \mathrm{td}^{\infty,0}$$

follow.

Finally, the results for some fixed λ can be carried over to a bounded interval $[0.01, \lambda']$ in exactly the same way as in the case where $\sigma > 0$. Indeed, the key ingredients, namely, Lemmas A.2 and B.2 still hold. To extend the results to $\lambda > \lambda'$, we resort to Lemma D.1.

For a fixed a prior Π, our arguments immediately give an instance-specific trade-off. Let $q^\Pi(\cdot; \delta, \sigma)$ be the function defined as

$$q^\Pi\left(\mathrm{tpp}^{\infty,\sigma}(\lambda); \delta, \sigma\right) = \mathrm{fdp}^{\infty,\sigma}(\lambda)$$

for all $\lambda > 0$. It is worth pointing out that the sparsity parameter ϵ is implied by Π and that q^Π depends on Π and σ only through Π/σ (if $\sigma = 0$, q^Π is invariant by rescaling). By definition, we always have

$$q^\Pi(u; \delta, \sigma) > q^\star(u)$$

for any u in the domain of q^Π. As is implied by the proof, it is impossible to have a series of instances Π such that $q^\Pi(u)$ converges to $q^\star(u)$ at *two* different points. Now, we state the instance-specific version of Theorem 2.1.

THEOREM D.2. *Fix $\delta \in (0, \infty)$ and assume the working hypothesis. In either the noiseless or noisy case and for any arbitrary small constants $\lambda_0 > 0$ and $\eta > 0$, the event*

$$\bigcap_{\lambda \ge \lambda_0} \left\{ \mathrm{FDP}(\lambda) \ge q^\Pi\left(\mathrm{TPP}(\lambda)\right) - \eta \right\}$$

holds with probability tending to one.

D.1. Proof of Lemma D.1. Consider the KKT conditions restricted to $\mathcal{S}(\lambda)$:

$$\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\left(\boldsymbol{y}-\boldsymbol{X}_{\mathcal{S}(\lambda)}\widehat{\boldsymbol{\beta}}(\lambda)\right)=\lambda\boldsymbol{g}(\lambda).$$

Here we abuse the notation a bit by identifying both $\widehat{\boldsymbol{\beta}}(\lambda)$ and $\boldsymbol{g}(\lambda)$ as $|\mathcal{S}(\lambda)|$-dimensional vectors. As a consequence, we get

$$\widehat{\boldsymbol{\beta}}(\lambda)=(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{y}-\lambda\boldsymbol{g}(\lambda)). \tag{D.13}$$

Notice that $\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)}$ is invertible almost surely since the Lasso solution has at most n nonzero components for *generic* problems (see e.g. [3]). By definition, $\widehat{\boldsymbol{\beta}}(\lambda)$ obeys

$$\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}_{\mathcal{S}(\lambda)}\widehat{\boldsymbol{\beta}}(\lambda)\|^2+\lambda\|\widehat{\boldsymbol{\beta}}(\lambda)\|_1\le\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{X}_{\mathcal{S}(\lambda)}\cdot\mathbf{0}\|^2+\lambda\|\mathbf{0}\|_1=\frac{1}{2}\|\boldsymbol{y}\|^2. \tag{D.14}$$

Substituting (D.13) into (D.14) and applying the triangle inequality give

$$\frac{1}{2}\left(\|\lambda\boldsymbol{X}_{\mathcal{S}(\lambda)}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}\boldsymbol{g}(\lambda)\|-\|\boldsymbol{y}-\boldsymbol{X}_{\mathcal{S}(\lambda)}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{y}\|\right)^2+\lambda\|\widehat{\boldsymbol{\beta}}(\lambda)\|_1\le\frac{1}{2}\|\boldsymbol{y}\|^2.$$

Since $\boldsymbol{I}_{|\mathcal{S}(\lambda)|}-\boldsymbol{X}_{\mathcal{S}(\lambda)}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}$ is simply a projection, we get

$$\|\boldsymbol{y}-\boldsymbol{X}_{\mathcal{S}(\lambda)}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{y}\|\le\|\boldsymbol{y}\|.$$

Combining the last displays gives

$$\lambda\|\boldsymbol{X}_{\mathcal{S}(\lambda)}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}\boldsymbol{g}(\lambda)\|\le 2\|\boldsymbol{y}\|, \tag{D.15}$$

which is our key estimate.

Since $\sigma_{\max}(\boldsymbol{X})=(1+o_{\mathbb{P}}(1))(1+1/\sqrt{\delta})$,

$$\lambda\|\boldsymbol{X}_{\mathcal{S}(\lambda)}(\boldsymbol{X}_{\mathcal{S}(\lambda)}^{\top}\boldsymbol{X}_{\mathcal{S}(\lambda)})^{-1}\boldsymbol{g}(\lambda)\|\ge(1+o_{\mathbb{P}}(1))\frac{\lambda\|\boldsymbol{g}(\lambda)\|}{1+1/\sqrt{\delta}}=(1+o_{\mathbb{P}}(1))\frac{\lambda\sqrt{|\mathcal{S}(\lambda)|}}{1+1/\sqrt{\delta}}. \tag{D.16}$$

As for the right-hand side of (D.15), the law of large numbers reveals that

$$2\|\boldsymbol{y}\|=2\|\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{z}\|=(2+o_{\mathbb{P}}(1))\sqrt{n(\|\boldsymbol{\beta}\|^2/n+\sigma^2)}=(2+o_{\mathbb{P}}(1))\sqrt{p\,\mathbb{E}\,\Pi^2+n\sigma^2}.$$

Combining the last two displays, it follows from (D.15) and (D.16) that

$$\|\widehat{\boldsymbol{\beta}}(\lambda)\|_0\equiv|\mathcal{S}(\lambda)_\lambda|\le(4+o_{\mathbb{P}}(1))\frac{(p\,\mathbb{E}\,\Pi^2+n\sigma^2)(1+1/\sqrt{\delta})^2}{\lambda^2}=(1+o_{\mathbb{P}}(1))\frac{Cp}{\lambda^2}$$

for some constant C. It is worth emphasizing that the term $o_{\mathbb{P}}(1)$ is independent of any $\lambda>0$. Hence, we finish the proof by choosing any $\lambda'>\sqrt{C/c}$.

APPENDIX E: PROOF OF THEOREM 3.1

We propose two preparatory lemmas regarding the χ^2-distribution, which will be used in the proof of the theorem.

LEMMA E.1. *For any positive integer d and $t \ge 0$, we have*

$$\mathbb{P}(\chi_d \ge \sqrt{d} + t) \le \mathrm{e}^{-t^2/2}.$$

LEMMA E.2. *For any positive integer d and $t \ge 0$, we have*

$$\mathbb{P}(\chi_d^2 \le td) \le (\mathrm{e}t)^{\frac{d}{2}}.$$

The first lemma can be derived by the Gaussian concentration inequality, also known as the Borell's inequality. The second lemma has a simple proof:

$$\begin{aligned}\mathbb{P}(\chi_d^2 \le td) &= \int_0^{td} \frac{1}{2^{\frac{d}{2}}\Gamma(\frac{d}{2})} x^{\frac{d}{2}-1}\mathrm{e}^{-\frac{x}{2}}\mathrm{d}x \\ &\le \int_0^{td} \frac{1}{2^{\frac{d}{2}}\Gamma(\frac{d}{2})} x^{\frac{d}{2}-1}\mathrm{d}x = \frac{2(td)^{\frac{d}{2}}}{d2^{\frac{d}{2}}\Gamma(\frac{d}{2})}.\end{aligned}$$

Next, Stirling's formula gives

$$\mathbb{P}(\chi_d^2 \le td) \le \frac{2(td)^{\frac{d}{2}}}{d2^{\frac{d}{2}}\Gamma(\frac{d}{2})} \le \frac{2(td)^{\frac{d}{2}}}{d2^{\frac{d}{2}}\sqrt{\pi d}(\frac{d}{2})^{\frac{d}{2}}\mathrm{e}^{-\frac{d}{2}}} \le (\mathrm{e}t)^{\frac{d}{2}}.$$

Now, we turn to present the proof of Theorem 3.1. Denote by $\mathcal{S}$ a subset of $\{1, 2, \ldots, p\}$, and let $m_0 = |\mathcal{S} \cap \{j : \beta_j = 0\}|$ and $m_1 = |\mathcal{S} \cap \{j : \beta_j \neq 0\}|$. Certainly, both m_0 and m_1 depend on $\mathcal{S}$, but the dependency is often omitted for the sake of simplicity. As earlier, denote by $k = \#\{j : \beta_j \neq 0\}$, which obeys $k = (\epsilon + o_{\mathbb{P}}(1))p$. Write $\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{LS}}$ for the least-squares estimate obtained by regressing $\boldsymbol{y}$ onto $\boldsymbol{X}_{\mathcal{S}}$. Observe that (3.1) is equivalent to solving

$$\underset{\mathcal{S} \subset \{1,\ldots,p\}}{\operatorname{argmin}} \; \|\boldsymbol{y} - \boldsymbol{X}_{\mathcal{S}}\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{L}S}\|^2 + \lambda|\mathcal{S}|. \tag{E.1}$$

As is clear from (3.1), we only need to focus on $\mathcal{S}$ with cardinality no more than $\min\{n, p\}$. Denote by $\widehat{\mathcal{S}}$ the solution to (E.1), and define $\widehat{m}_0$ and $\widehat{m}_1$ as before. To prove Theorem 3.1 it suffices to show the following: for arbitrary small $c > 0$, we can find λ and M sufficiently large such that (3.1) gives

$$\mathbb{P}(\widehat{m}_0 > 2ck \text{ or } \widehat{m}_1 \le (1 - c)k) \to 0. \tag{E.2}$$

Assume this is true. Then from (E.2) we see that $\widehat{m}_0 \le 2ck$ and $\widehat{m}_1 > (1-c)k$ hold with probability tending to one. On this event, the TPP is

$$\frac{\widehat{m}_1}{k} > 1 - c,$$

and the FDP is

$$\frac{\widehat{m}_0}{\widehat{m}_0 + \widehat{m}_1} \le \frac{2ck}{2ck + (1-c)k} = \frac{2c}{1+c}.$$

Hence, we can have arbitrarily small FDP and almost full power by setting c arbitrarily small.

It remains to prove (E.2) with proper choices of λ and M. Since

$$\{\widehat{m}_0 > 2ck \text{ or } \widehat{m}_1 \le (1-c)k\} \subset \{\widehat{m}_0 + \widehat{m}_1 > (1+c)k\} \cup \{\widehat{m}_1 \le (1-c)k\},$$

we only need to prove

$$\mathbb{P}(\widehat{m}_1 \le (1-c)k) \to 0 \tag{E.3}$$

and

$$\mathbb{P}(\widehat{m}_0 + \widehat{m}_1 > (1+c)k) \to 0. \tag{E.4}$$

We first work on (E.3). Write

$$\boldsymbol{y} = \sum_{j \in \mathcal{S}, \beta_j = M} M\boldsymbol{X}_j + \sum_{j \in \overline{\mathcal{S}}, \beta_j = M} M\boldsymbol{X}_j + \boldsymbol{z}.$$

In this decomposition, the summand $\sum_{j \in \mathcal{S}, \beta_j = M} M\boldsymbol{X}_j$ is already in the span of $\boldsymbol{X}_{\mathcal{S}}$. This fact implies that the residual vector $\boldsymbol{y} - \boldsymbol{X}_{\mathcal{S}}\widehat{\boldsymbol{\beta}}^{\mathrm{L}S}_{\mathcal{S}}$ is the same as the projection of $\sum_{j \in \overline{\mathcal{S}}, \beta_j = M} M\boldsymbol{X}_j + \boldsymbol{z}$ onto the orthogonal complement of $\boldsymbol{X}_{\mathcal{S}}$. Thanks to the independence among $\boldsymbol{\beta}$, $\boldsymbol{X}$ and $\boldsymbol{z}$, our discussion proceeds by conditioning on the random support set of $\boldsymbol{\beta}$. A crucial but simple observation is that the orthogonal complement of $\boldsymbol{X}_{\mathcal{S}}$ of dimension $n - m_0 - m_1$ has uniform orientation, independent of $\sum_{j \in \overline{\mathcal{S}}, \beta_j = M} M\boldsymbol{X}_j + \boldsymbol{z}$. From this fact it follows that

$$L(\mathcal{S}) := \|\boldsymbol{y} - \boldsymbol{X}_{\mathcal{S}}\widehat{\boldsymbol{\beta}}^{\mathrm{L}S}_{\mathcal{S}}\|^2 + \lambda|\mathcal{S}| \overset{d}{=} \left(\sigma^2 + M^2(k - m_1)/n\right)\chi^2_{n-m_0-m_1} + \lambda(m_0 + m_1). \tag{E.5}$$

Call $E_{\mathcal{S},u}$ the event on which

$$L(\mathcal{S}) \le \sigma^2(n - k + 2u\sqrt{n-k} + u^2) + \lambda k$$

holds; here, $u > 0$ is a constant to be specified later. In the special case where $\mathcal{S} = \mathcal{T}$ and $\mathcal{T} \equiv \{j : \beta_j \neq 0\}$ is the true support, Lemma E.1 says that this event has probability bounded as

$$\begin{aligned}\mathbb{P}(E_{\mathcal{T},u}) &= \mathbb{P}\left(\sigma^2\chi^2_{n-k} + \lambda k \le \sigma^2(n-k+2u\sqrt{n-k}+u^2) + \lambda k\right)\\ &= \mathbb{P}\left(\chi^2_{n-k} \le n-k+2u\sqrt{n-k}+u^2\right)\\ &\ge 1 - \mathrm{e}^{-\frac{u^2}{2}}.\end{aligned} \tag{E.6}$$

By definition, $E_{\widehat{\mathcal{S}},u}$ is implied by $E_{\mathcal{T},u}$. Using this fact, we will show that $\widehat{m}_1$ is very close to k, thus validating (E.3). By making use of

$$\{\widehat{m}_1 \le (1-c)k\} \subset \{\widehat{m}_0+\widehat{m}_1 \ge (k+n)/2\} \cup \{\widehat{m}_1 \le (1-c)k, \widehat{m}_0+\widehat{m}_1 < (k+n)/2\},$$

we see that it suffices to establish that

$$\mathbb{P}(\widehat{m}_0 + \widehat{m}_1 \ge (k+n)/2) \to 0 \tag{E.7}$$

and

$$\mathbb{P}(\widehat{m}_1 \le (1-c)k, \widehat{m}_0 + \widehat{m}_1 < (k+n)/2) \to 0 \tag{E.8}$$

for some λ and sufficient large M. For (E.7), we have

$$\begin{aligned}\mathbb{P}(\widehat{m}_0 + \widehat{m}_1 \ge (k+n)/2) &\le \mathbb{P}(\overline{E}_{\mathcal{T},u}) + \mathbb{P}(E_{\mathcal{T},u} \cap \{\widehat{m}_0 + \widehat{m}_1 \ge (k+n)/2\})\\ &\le \mathbb{P}(\overline{E}_{\mathcal{T},u}) + \mathbb{P}(E_{\widehat{\mathcal{S}},u} \cap \{\widehat{m}_0 + \widehat{m}_1 \ge (k+n)/2\})\\ &\le \mathbb{P}(\overline{E}_{\mathcal{T},u}) + \sum_{m_0+m_1 \ge (k+n)/2} \mathbb{P}(E_{\mathcal{S},u})\\ &\le \mathrm{e}^{-\frac{u^2}{2}} + \sum_{m_0+m_1 \ge (k+n)/2} \mathbb{P}(E_{\mathcal{S},u}),\end{aligned} \tag{E.9}$$

where the last step makes use of (E.6), and the summation is over all $\mathcal{S}$ such that $m_0(\mathcal{S}) + m_1(\mathcal{S}) \ge (k+n)/2$. Due to (E.5), the event $\mathbb{E}_{\mathcal{S},u}$ has the same probability as

$$\begin{aligned}&\left(\sigma^2 + M^2(k-m_1)/n\right)\chi^2_{n-m_0-m_1} + \lambda(m_0+m_1) \le \sigma^2(n-k+2u\sqrt{n-k}+u^2) + \lambda k\\ &\Longleftrightarrow \chi^2_{n-m_0-m_1} \le \frac{\sigma^2(n-k+2u\sqrt{n-k}+u^2) + \lambda k - \lambda(m_0+m_1)}{\sigma^2 + M^2(k-m_1)/n}.\end{aligned} \tag{E.10}$$

Since $m_0 + m_1 \ge (k+n)/2$, we get

$$\sigma^2(n-k+2u\sqrt{n-k}+u^2)+\lambda k-\lambda(m_0+m_1) \le \sigma^2(n-k+2u\sqrt{n-k}+u^2)-\lambda(n-k)/2.$$

Requiring

$$\lambda > 2\sigma^2, \tag{E.11}$$

would yield $\sigma^2(n-k+2u\sqrt{n-k}+u^2) - \lambda(n-k)/2 < 0$ for sufficiently large n (depending on u) as $n-k\to\infty$. In this case, we have $\mathbb{P}(E_{\mathcal{S},u}) = 0$ whenever $m_0+m_1 \geq (k+n)/2$. Thus, taking $u\to\infty$ in (E.9) establishes (E.7).

Now we turn to (E.8). Observe that

$$\begin{aligned}
&\mathbb{P}\left(\widehat{m}_1 \leq (1-c)k \text{ and } \widehat{m}_0+\widehat{m}_1 < (k+n)/2\right) \\
&\quad\leq \mathbb{P}(\overline{E}_{\mathcal{T},u}) + \mathbb{P}(E_{\mathcal{T},u}\cap\{\widehat{m}_1 \leq (1-c)k \text{ and } \widehat{m}_0+\widehat{m}_1 < (k+n)/2\}) \\
&\quad\leq \mathrm{e}^{-\frac{u^2}{2}} + \mathbb{P}(E_{\widehat{\mathcal{S}},u}\cap\{\widehat{m}_1 \leq (1-c)k \text{ and } \widehat{m}_0+\widehat{m}_1 < (k+n)/2\}) \\
&\quad\leq \mathrm{e}^{-\frac{u^2}{2}} + \sum_{m_0+m_1<\frac{k+n}{2},\, m_1\leq(1-c)k} \mathbb{P}(E_{\mathcal{S},u}).
\end{aligned} \tag{E.12}$$

For $m_0+m_1 < (k+n)/2$ and $m_1 \leq (1-c)k$, notice that $n-m_0-m_1 > (n-k)/2 = (\delta-\epsilon+o_{\mathbb{P}}(1))p/2$, and $M^2(k-m_1)/n \geq cM^2k/n \sim (c\epsilon/\delta+o_{\mathbb{P}}(1))M^2$. Let $t_0 > 0$ be a constant obeying

$$\frac{\delta-\epsilon}{5}(1+\log t_0)+\log 2 < -1,$$

then choose M sufficiently large such that

$$\frac{2\sigma^2(\delta-\epsilon)+2\lambda\epsilon}{(\sigma^2+c\epsilon M^2/\delta)(\delta-\epsilon)} < t_0. \tag{E.13}$$

This gives

$$\frac{\sigma^2(n-k+2u\sqrt{n-k}+u^2)+\lambda k-\lambda(m_0+m_1)}{(\sigma^2+M^2(k-m_1)/n)(n-m_0-m_1)} < t_0$$

for sufficiently large n. Continuing (E.12) and applying Lemma E.2, we get

(E.14)

$$\begin{aligned}
&\mathbb{P}\left(\widehat{m}_1 \le (1-c)k \text{ and } \widehat{m}_0 + \widehat{m}_1 < (k+n)/2\right) \\
&\qquad \le \mathrm{e}^{-\frac{u^2}{2}} + \sum_{m_0+m_1<\frac{k+n}{2}, m_1 \le (1-c)k} \mathbb{P}(\chi^2_{n-m_0-m_1} \le t_0(n-m_0-m_1)) \\
&\qquad \le \mathrm{e}^{-\frac{u^2}{2}} + \sum_{m_0+m_1<\frac{k+n}{2}, m_1 \le (1-c)k} (\mathrm{e}t_0)^{\frac{n-m_0-m_1}{2}} \\
&\qquad \le \sum_{m_0+m_1<(k+n)/2, m_1 \le (1-c)k} (\mathrm{e}t_0)^{\frac{(\delta-\epsilon)p}{5}} \\
&\qquad \le \mathrm{e}^{-\frac{u^2}{2}} + 2^p(\mathrm{e}t_0)^{\frac{(\delta-\epsilon)p}{5}} \\
&\qquad \le \mathrm{e}^{-\frac{u^2}{2}} + \mathrm{e}^{-p}.
\end{aligned}$$

Taking $u \to \infty$ proves (E.8).

Having established (E.3), we proceed to prove (E.4). By definition,

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 + \lambda\|\widehat{\boldsymbol{\beta}}\|_0 &= \|\boldsymbol{y} - \boldsymbol{X}_{\widehat{S}}\widehat{\boldsymbol{\beta}}^{\mathrm{L}S}_{\widehat{S}}\|^2 + \lambda\|\boldsymbol{\beta}^{\mathrm{L}S}_{\widehat{S}}\|_0 \\
&\le \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_0 \\
&= \|\boldsymbol{z}\|^2 + \lambda k.
\end{aligned}$$

If

$$\lambda > \frac{\sigma^2\delta}{c\epsilon}, \tag{E.15}$$

then

$$\widehat{m}_0 + \widehat{m}_1 \le \frac{\|\boldsymbol{z}\|^2}{\lambda} + k = (1 + o_{\mathbb{P}}(1))\frac{\sigma^2 n}{\lambda} + k \le (1+c)k$$

holds with probability tending to one, whence (E.4).

To recapitulate, selecting λ obeying (E.11) and (E.15), and M sufficiently large such that (E.13) holds, imply that (E.2) holds. The proof of the theorem is complete.

REFERENCES

[1] M. Bayati and A. Montanari. The Lasso risk for Gaussian matrices. *IEEE Trans. Inform. Theory*, 58(4):1997–2017, 2012.

[2] M. Bogdan, E. v. d. Berg, W. Su, and E. J. Candès. Supplementary material to "Statistical estimation and testing via the ordered ℓ_1 norm". Available at `http://statweb.stanford.edu/~candes/papers/SortedL1_SM.pdf`, 2013.

[3] R. J. Tibshirani. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.

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

Departments of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
USA
E-mail: suw@wharton.upenn.edu

Institute of Mathematics
University of Wroclaw
50-137, Wroclaw
Poland
E-mail: malgorzata.bogdan@pwr.edu.pl

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