On the asymptotic properties of SLOPE. Supplementary materials.

Michał Kos[a] and Małgorzata Bogdan[a,b]

[a] Department of Mathematics, University of Wroclaw
[b] Department of Statistics, Lund University

August 21, 2019

1 Proofs of Theorems 3.1 and 3.2

In this Section we present proofs of Theorems 3.1 and 3.2 and some additional interesting facts. The proofs are similar to each other on certain level. Therefore we organize them in a way that in authors opinion will be the most convenient to the reader. We would like to start with presenting some facts showing a specific connection between the score vector $U_i(\hat{b})$ and the estimator $\hat{b}$.
Straightforward, from the proof of the Theorem 3.2, we obtain following fact:

Corollary S.1.1. *Suppose assumptions of the Theorem 3.2 holds. When $\hat{b}_i \neq 0$ then $U_i(\hat{b})$ and $\hat{b}_i$ have the same sign (see inequality S.1.3).*

Proposition S.1.2. *Consider optimization problem given by generalized SLOPE (see: 3.1 in the Article) with an arbitrary sequence $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_p \geqslant 0$ and assume that $l(b)$ is a convex and differentiable function.*
Under above assumptions if $|\hat{b}_j| > |\hat{b}_k|$ then $|U_j(\hat{b})| \geqslant |U_k(\hat{b})|$.

Above Proposition enable further specification of the $(|b_1|, ..., |b_p|)'$ ordering in case when $|b|_{(j)} = |b|_{(j+1)}$ for some j. Let's introduce a notation $U_{(j)\#}(b)$ and $b_{(j)\#}$ on the score statistic and element of the vector b associated with the element $|b|_{(j)}$. Of course when $|b|_{(j)} = |b|_{(j+1)}$ the indexing is ambiguous and therefore we define it in a way that associated statistics have a following property:

$$|U_{(j)\#}(b)| \geqslant |U_{(j+1)\#}(b)|$$

Remark S.1.3. *By the above ordering and the Proposition S.1.2 we know that the vector* $(|U_{(1)\#}(\hat{b})|, ..., |U_{(p)\#}(\hat{b})|)'$ *associated with* $(|\hat{b}|_{(1)}, ..., |\hat{b}|_{(p)})'$ *is ordered* $(|U_{(1)\#}(\hat{b})| \geqslant ... \geqslant |U_{(p)\#}(\hat{b})|)$ *and in consequence we can write* $|U_{(j)\#}(\hat{b})| = |U(\hat{b})|_{(j)}$.

Corollary S.1.4. *Consider optimization problem given by generalized SLOPE (see:* 3.1 *in the Article) with an arbitrary sequence* $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_p \geqslant 0$ *and assume that* $l(b)$ *is a convex and differentiable function. Then for any arbitrary* $a > 0$ *and* $j = 1, ..., p$*, it holds:*

$$|U_{(j)\#}(\hat{b}) + a\hat{b}_{(j)\#}| = |U(\hat{b}) + a\hat{b}|_{(j)} = |T(a)|_{(j)}$$

Proof. Observe that:

$$|U_{(j)\#}(\hat{b}) + a\hat{b}_{(j)\#}| = |U_{(j)\#}(\hat{b})| + |a\hat{b}_{(j)\#}| = |U(\hat{b})|_{(j)} + |a\hat{b}|_{(j)} = |U(\hat{b}) + a\hat{b}|_{(j)} = |T(a)|_{(j)}$$

The first equality is a consequence of Corollary S.1.1. The second results from the Remark S.1.3. The third one comes from a fact that sum of two positive ordered sequences is also ordered and the last is simply a consequence of the $T(a)$ definition. ■

Through out this section we shall denote the objective function of generalized SLOPE (see: 3.1 in the Article) by:

$$f(b) = l(b) + \sum_{j=1}^{p} \lambda_j |b|_{(j)}$$

Proof. The proof of the Theorem 3.2.
We will start the proof by showing that:

$$\left(\hat{b}_i \neq 0\right) \Rightarrow (|T_i(a)| > \lambda_r)$$

and

$$\left(\hat{b}_i \neq 0\right) \Rightarrow \left(|U_i(\hat{b})| \geqslant \lambda_r\right)$$

From the optimality of $\hat{b}$ we have following inequality for any vector b:

$$0 \geqslant f(\hat{b}) - f(b) = l(\hat{b}) - l(b) + \sum_{j=1}^{p} \lambda_j |\hat{b}|_{(j)} - \sum_{j=1}^{p} \lambda_j |b|_{(j)} \tag{S.1.1}$$

In the first step we will prove that for a vector $b = \hat{b} + h(0, ..., 0, \underbrace{(-1)sgn(\hat{b}_i)}_{i-th\ \ pos.}, 0, ..., 0)^T = \hat{b} + hL_i$ and small enough, positive h we have:

$$\sum_{j=1}^{p} \lambda_j |\hat{b}|_{(j)} - \sum_{j=1}^{p} \lambda_j |b|_{(j)} \geqslant h\lambda_r \tag{S.1.2}$$

Let's consider first the simplest scenario where $|\hat{b}_i| \neq |\hat{b}_j|$ for any $i \neq j$. In this situation if we take

$$h < \min_{j,\ \ i \neq j} \left(\left| |\hat{b}_i| - |\hat{b}_j| \right| \right)$$

then the ordering of an absolute value of the elements in the vector b and $\hat{b}$ will be the same. Furthermore these vectors differ only in the i-th position and from the form of a vector b we know that the absolute value of an i-th element in vector b is smaller then in vector $\hat{b}$. In consequence we obtain for certain index k that:

$$\sum_{j=1}^{p} \lambda_j |\hat{b}|_{(j)} - \sum_{j=1}^{p} \lambda_j |b|_{(j)} = h\lambda_k \geqslant h\lambda_r$$

The last inequality is a consequence of assumptions $\#\{i : \hat{b}_i \neq 0\} = r$ and $\hat{b}_i \neq 0$ and a fact that λ_r is the smallest element of vector λ associated with non-zero elements of the vector $\hat{b}$.

Let's consider now more general scenario where two or more elements of the vector $\hat{b}$ have the same absolute value as $|\hat{b}_i|$. The reasoning is analogous, however in this situation value $|\hat{b}_i|$ will be associated with a set of elements of vector λ. Of course the power of associated set will be the same as the number of elements of vector $\hat{b}$ with absolute value the same as $|\hat{b}_i|$. Moreover the associated set of lambdas will contain consecutive elements of whole sequence. It is also straightforward that the indexing in a group of elements of vector $\hat{b}$ with the same module as $|\hat{b}_i|$ is ambiguous. For such situation if we take

$$h < \min_{j,\ |\hat{b}_i| \neq |\hat{b}_j|} \left(\left| |\hat{b}_i| - |\hat{b}_j| \right| \right)$$

we obtain for certain index k that:

$$\sum_{j=1}^{p} \lambda_j |\hat{b}|_{(j)} - \sum_{j=1}^{p} \lambda_j |b|_{(j)} = h\lambda_k \geqslant h\lambda_r$$

however this time λ_k is the smallest element associated with $|\hat{b}_i|$. The justification for the last inequality is analogous to the previously discussed situation.

Using proven relation (S.1.2) in the inequality (S.1.1) we obtain:

$$\lambda_r \leqslant \frac{l(b) - l(\hat{b})}{h}$$

Via transformation of an expression on the right side we obtain:

$$\lambda_r \leqslant \frac{l(b) - l(\hat{b})}{h} = \frac{l(\hat{b} + hL_i) - l(\hat{b})}{h} \to sgn(\hat{b}_i)U_i(\hat{b}) \tag{S.1.3}$$

for $h \to 0$. By adding to both sides factor $a|\hat{b}_i|$ we obtain:

$$\lambda_r < \lambda_r + a|\hat{b}_i| \leqslant sgn(\hat{b}_i)(U_i(\hat{b}) + a\hat{b}_i) \leqslant |U_i(\hat{b}) + a\hat{b}_i|$$

Naturally if $a = 0$ then:

$$\lambda_r \leqslant |U_i(\hat{b})|$$

This ends the proof in one direction.
It remains to show that

$$\left(\hat{b}_i \neq 0\right) \Leftarrow (|T_i(a)| > \lambda_r)$$

and that when $\lambda_1 > ... > \lambda_p \geqslant 0$ then

$$\left(\hat{b}_i \neq 0\right) \Leftarrow \left(|U_i(\hat{b})| \geqslant \lambda_r\right)$$

The proof of implications will be shown via contradiction. Let's assume that $\hat{b}_i = 0$ and $\lambda_r < |U_i(\hat{b}) + a\hat{b}_i|$ and consider vector $b = \hat{b} + h(0, ..., 0, \underbrace{1}_{i-th\ \ pos.}, 0, ..., 0)^T = \hat{b} + hL_i$. Let's recall that λ_{r+1} is the largest element of the sequence λ associated with the elements of the vector $\hat{b}$ with value 0. It is easy to notice that for small enough, positive h ($h < \min_i\{|\hat{b}_i|,\ \hat{b}_i \neq 0\}$) we obtain:

$$\sum_{j=1}^{p} \lambda_j |\hat{b}|_{(j)} - \sum_{j=1}^{p} \lambda_j |b|_{(j)} = -h\lambda_{r+1}$$

and in consequence following inequality holds:

$$0 \geqslant f(\hat{b}) - f(b) = l(\hat{b}) - l(b) - h\lambda_{r+1}$$

Transforming above inequality we obtain:

$$\lambda_{r+1} \geqslant -\frac{l(\hat{b} + hL_i) - l(\hat{b})}{h} \to U_i(\hat{b})$$

Analogical reasoning for vector $b = \hat{b} + h(0, ..., 0, \underbrace{-1}_{i-th\ \ pos.}, 0, ..., 0)^T$ provides us:

$$\lambda_{r+1} \geqslant -U_i(\hat{b})$$

In consequence we obtain that $|U_i(\hat{b})| \leqslant \lambda_{r+1}$. On the other hand we assumed that $\lambda_r < |U_i(\hat{b}) + a\hat{b}_i| = |U_i(\hat{b})|$ (assumpt. $\hat{b}_i = 0$) which leads to contradiction and ends the proof of first equivalence.

It is easy to notice that we used the assumption $\lambda_r < |U_i(\hat{b}) + a\hat{b}_i| = |U_i(\hat{b})|$ at the end of the above reasoning. Therefore in the case of the second equivalence the calculation are the same and in consequence we can again use condition $|U_i(\hat{b})| \leqslant \lambda_{r+1}$. On the other hand we assume that $\lambda_{r+1} < \lambda_r \leqslant |U_i(\hat{b})|$ which again leads to contradiction and ends the proof. ■

Proof. The proof of Proposition S.1.2.
The proof is analogical to a proof of the Theorem 3.2.
Let's consider a vector $b = \hat{b} + h(0, ..., 0, \underbrace{(-1)sgn(\hat{b}_j)}_{j-th\ \ pos.}, 0, ..., 0, \underbrace{sgn(\hat{b}_k)}_{k-th\ \ pos.}, 0, ..., 0)^T = \hat{b} + hL$. For small enough positive h:

$$0 \geqslant f(\hat{b}) - f(b) = l(\hat{b}) - l(b) + h\lambda_l - h\lambda_m$$

where λ_l and λ_m are the smallest and the largest element of the sequence λ associated with the elements $|\hat{b}_j|$ and $|\hat{b}_k|$, respectively. This is a consequence of the vector b form. By its definition the $j-th$ element is pulled towards 0 and the $k-th$ element is pushed away from zero. Now, from the assumption $|\hat{b}_j| > |\hat{b}_k|$ we know that $\lambda_l \geqslant \lambda_m$. Therefore, after transformation we obtain:

$$0 \leqslant \lambda_l - \lambda_m \leqslant \frac{l(b) - l(\hat{b})}{h} \to \nabla_L l(\hat{b}) = sgn(\hat{b}_j)U_j(\hat{b}) - sgn(\hat{b}_k)U_k(\hat{b})$$

for $h \to 0$. The last equation is a consequence of differentiability of function $l(b)$. Using relation $sgn(\hat{b}_i)U_i(\hat{b}) = |U_i(\hat{b})|$ (see corollary S.1.1) we obtain:

$$0 \leqslant |U_j(\hat{b})| - |U_k(\hat{b})|$$

This ends the proof. ■

Proof. The proof of the Theorem 3.1.
Recall that $T(a) = U(\hat{b}) + a\hat{b}$ and that from the definition of the set H_r we know that

$\left(U(\hat{b})+a\hat{b}\right)\in H_r$ is fulfilled if and only if both following condition are satisfied:
$i)$

$$\forall_{j\leqslant r}\sum_{i=j}^{r}\lambda_i<\sum_{i=j}^{r}|U(\hat{b})+a\hat{b}|_{(i)}$$

$ii)$

$$\forall_{j\geqslant r+1}\sum_{i=r+1}^{j}\lambda_i\geqslant\sum_{i=r+1}^{j}|U(\hat{b})+a\hat{b}|_{(i)}$$

Let's assume $\#\{i:\hat{b}_i\neq 0\}=r$ and determine $j\leqslant r$.
Furthermore, let I be a set of indexes for which the rank of absolute value of the element of a vector $\hat{b}$ is between j and r. Similarly to the proof of Theorem 3.2 we consider vector b defined in a following way:

$$b=\begin{cases}\hat{b}_i-h\ sgn(\hat{b}_i) & i\in I\\ \hat{b}_i\quad otherwise\end{cases}$$

Again, for h positive and small enough we obtain:

$$0\geqslant f(\hat{b})-f(b)=l(\hat{b})-l(b)+h\sum_{i=j}^{r}\lambda_i$$

Above is a consequence of a construction of the vector b. It is obtained via pulling $r-(j-1)$ smallest nonzero elements in the vector $\hat{b}$ towards zero by the same factor h.
Transforming above inequality we obtain:

$$\sum_{i=j}^{r}\lambda_i\leqslant\frac{l(b)-l(\hat{b})}{h}\longrightarrow\sum_{i=j}^{r}sgn\left(\hat{b}_{(i)\#}\right)U_{(i)\#}(\hat{b})=\sum_{i=j}^{r}|U_{(i)\#}(\hat{b})|=\sum_{i=j}^{r}|U(\hat{b})|_{(i)}$$

for $h\to 0$. By adding to both sides of the inequality $\sum_{i=j}^{r}|a\hat{b}|_{(i)}$ we obtain:

$$\sum_{i=j}^{r}\lambda_i<\sum_{i=j}^{r}(\lambda_i+|a\hat{b}|_{(i)})\leqslant\sum_{i=j}^{r}\left(|U(\hat{b})|_{(i)}+|a\hat{b}|_{(i)}\right)=\sum_{i=j}^{r}|U(\hat{b})+a\hat{b}|_{(i)}$$

which ends the proof that $\#\{i:\hat{b}_i\neq 0\}=r$ ensure (i).
Now let's determine $j\geqslant r+1$ and introduce set J of indexes for which the rank of

an absolute value of the element of a vector $\hat{b}$ is between $r+1$ and j. Let's consider following set of vectors b:

$$b = \begin{cases} hE_i & i \in J \\ \hat{b}_i & otherwise \end{cases}$$

where $E_i = \pm 1$ and $h > 0$.
For small enough h and for any vector b we obtain:

$$0 \geqslant f(\hat{b}) - f(b) = l(\hat{b}) - l(b) - h \sum_{i=r+1}^{j} \lambda_i$$

Again it is associated with a form of vector b. However this time $j - r$ elements with value 0 in the vector $\hat{b}$ are pushed away from 0 by a factor h.
By transformation of the above inequality we obtain:

$$\sum_{i=r+1}^{j} \lambda_i \geqslant \frac{l(\hat{b}) - l(b)}{h} \longrightarrow \sum_{i=r+1}^{j} E_{(i)\#} U_{(i)\#}(\hat{b})$$

for $h \to 0$. This inequality is valid for any sequence of $\{E_i\}$. Therefore:

$$\sum_{i=r+1}^{j} \lambda_i \geqslant \sum_{i=r+1}^{j} |U_{(i)\#}(\hat{b})| = \sum_{i=r+1}^{j} |U(\hat{b})|_{(i)} = \sum_{i=r+1}^{j} |U(\hat{b}) + a\hat{b}|_{(i)}$$

The last equality is a consequence of a fact that $\hat{b}_{(i)} = 0$ for $i > r$.
This ends the proof that $\#\{i : \hat{b}_i \neq 0\} = r$ ensure (ii).
To prove the equivalence in other direction we will show that there are at least and at most r nonzero elements in $\hat{b}$.
Suppose that conditions (i) and (ii) are fulfilled and that $\#\{i : \hat{b}_i \neq 0\} = j < r$. From the first part of the proof we know that when $\#\{i : \hat{b}_i \neq 0\} = j$ then for $m > j$ we have $|\hat{b}|_{(m)} = 0$ and:

$$\sum_{i=j+1}^{r} \lambda_i \geqslant \sum_{i=j+1}^{r} |U(\hat{b})|_{(i)}$$

On the other hand from (i) we have:

$$\sum_{i=j+1}^{r} \lambda_i < \sum_{i=j+1}^{r} |U(\hat{b})|_{(i)}$$

which leads to a contradiction.
Now let's assume that $\#\{i : \hat{b}_i \neq 0\} = j > r$. This implies that:

$$\sum_{i=r}^{j} \lambda_i < \sum_{i=r}^{j} |U(\hat{b}) + a\hat{b}|_{(i)}$$

On the other hand from (ii) we obtain:

$$\sum_{i=r}^{j} \lambda_i \geqslant \sum_{i=r}^{j} |U(\hat{b}) + a\hat{b}|_{(i)}$$

which again leads to a contradiction and ends the proof of the Theorem. ■

2 Results used in the proof of the Theorem 2.1

In this section we present theoretical results that are used to prove the Theorem 2.1.

Definition S.2.1 (Resolvent set). *Fix $S = supp(b^0)$ of cardinality k, and an integer k^* obeying $k < k^* < p$. The set $S^* = S^*(S, k^*)$ is said to be a resolvent set if it is the union of S and the $k^* - k$ indexes with the largest values of $|X_i'\epsilon|$ among all $i \in \{1, ..., p\} \setminus S$.*

Lemma S.2.2 (Su and Candès (2016), Lemma 4.4). *Suppose we consider the sequence of linear models (1.1) where $p \to \infty, k/p \to 0, (k \log p)/n \to 0$. Moreover, let*

$$k^* \geqslant max\left(\frac{1+c}{1-q}k, k+d\right) \tag{S.2.1}$$

for an arbitrary small constant $c > 0$, and where d is a deterministic sequence diverging to infinity in such a way that $k^*/p \to 0$ and $(k^* \log p)/n \to 0$. Then

$$P(\{Supp(b^0) \cup Supp(\hat{b}) \subseteq S^*\}) \to 1. \tag{S.2.2}$$

Proof. Proof of Corollary 3.5.
Naturally, it is easy to notice that assumptions of the main Theorem in Article (2.1) are stronger than those of the Lemma S.2.2. In consequence under assumptions of the Theorem 2.1, we can choose a sequence d and k^* in such a way that S.2.2 holds, $k^*/p \to 0$ and $((k^*)^2 \log p)/n \to 0$. ■

Definition S.2.3. *Given two metric spaces (X, d_X) and (Y, d_Y), where d_X denotes the metric on the set X and d_Y is the metric on set Y, a function $f : X \to Y$ is called L - lipschitz continuous if there exists a real constant $L \geqslant 0$ such that, for all x_1 and x_2 in X,*

$$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2)$$

Theorem S.2.4. *(Vershynin, 2012) Let $\zeta \sim N(0, \mathbb{I}_n)$ and f be a L-lipschitz continous function in $\mathbb{R}^n$. Then*

$$P(f(\zeta) \leqslant Ef(\zeta) + t) \geqslant 1 - e^{-\frac{t^2}{2L^2}}$$

for all $t \geqslant 0$.

We shall denote by $X_I, \hat{b}_I, b_I^0$ a submatrix (subvector) of $X, \hat{b}, b^0$ consisted of columns (vector elements) with indexes in a set I.

Lemma S.2.5 (Su and Candès (2016), Lemma A.12)**.** *Under assumptions of the Theorem 2.1 (in original paper the assumptions where less stringent) there exist a constant C_q only depending on q such that for all j we have:*

$$\sum_{i=1}^{j} |X'_{(S^*)^c} X_{S^*}(\hat{b}_{S^*} - b^0_{S^*})|_{(i)} \leqslant C_q \sqrt{\frac{k^* \log p}{n}} \sum_{i=k^*+1}^{k^*+j} \lambda_i^{BH}$$

with probability tending to 1.

Proof. The proof of the Corollary 6.1.
Directly from the Lemma S.2.5 for $j = 1$ we obtain:

$$\max_{i \in (S^*)^c} |[X'_{(S^*)^c} X_{S^*}(\hat{b}_{S^*} - b^0_{S^*})]_i| = |X'_{(S^*)^c} X_{S^*}(\hat{b}_{S^*} - b^0_{S^*})|_{(1)} \leqslant C_q \sqrt{\frac{k^* \log p}{n}} \lambda^{BH}_{k^*+1} \tag{S.2.3}$$

with probability tending to 1. ■

Lemma S.2.6. *(Su and Candès (2016), Lemma A.11) Let $k < k^* < \min\{n, p\}$ be any (deterministic) integer. Denote by σ_{max} and σ_{min}, the largest and the smallest singular value of X_{S^*}. Then for any $t > 0$,*

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

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

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

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

Proof. The proof of relations 6.16 and 6.17.
Recall the form of 6.16:

$$\sum_{r=k+2}^{k^*} F_E(2p^{-\delta}, r-k-1, 1) \to 0 \tag{S.2.4}$$

Furthermore, recall that

$$F_E(x, m, 1) = 1 - \exp(-x)\sum_{l=0}^{m-1}\frac{1}{l!}(x)^l = \exp(-x)\sum_{l=m}^{\infty}\frac{1}{l!}(x)^l$$

Therefore

$$\sum_{r=k+2}^{k^*} F_E(2p^{-\delta}, r-k-1, 1) = \exp(-2p^{-\delta})\sum_{r=k+2}^{k^*}\sum_{l=r-k-1}^{\infty}\frac{1}{l!}(2p^{-\delta})^l =$$

$$\exp(-2p^{-\delta})\left(\sum_{l=1}^{\infty}\frac{1}{l!}(2p^{-\delta})^l + \sum_{l=2}^{\infty}\frac{1}{l!}(2p^{-\delta})^l + ... + \sum_{l=k^*-k-1}^{\infty}\frac{1}{l!}(2p^{-\delta})^l\right) =$$

$$= \exp(-2p^{-\delta})\left[\sum_{l=1}^{k^*-k-1} l\left(\frac{1}{l!}(2p^{-\delta})^l\right) + \sum_{l=k^*-k}^{\infty}\underbrace{(k^*-k-1)}_{\leqslant l}\left(\frac{1}{l!}(2p^{-\delta})^l\right)\right] \leqslant$$

$$\exp(-2p^{-\delta})(2p^{-\delta})\sum_{l=1}^{\infty}\frac{1}{(l-1)!}(2p^{-\delta})^{(l-1)} = (2p^{-\delta}) \to 0$$

which we wanted to show. Now, recall the form of the 6.17:

$$\sum_{r=k+2}^{k^*} P\left(\frac{1}{p-k}\sum_{j=1}^{p-k}E_j > 2\right) = (k^*-k-1)P\left(\frac{1}{p-k}\sum_{j=1}^{p-k}E_j > 2\right) \to 0 \tag{S.2.5}$$

To see that above is also valid observe that:

$$E\left(\frac{1}{p-k}\sum_{j=1}^{p-k}E_j\right) = 1$$

$$Var\left(\frac{1}{p-k}\sum_{j=1}^{p-k}E_j\right) = \left(\frac{1}{p-k}\right)^2\sum_{j=1}^{p-k}Var\left(E_j\right) = \frac{1}{p-k} = \sigma_1^2$$

Therefore do to the Chebyshev's inequality we obtain that

$$P\left(\frac{1}{p-k}\sum_{j=1}^{p-k} E_j > 2\right) = P\left(\frac{1}{p-k}\sum_{j=1}^{p-k} E_j - 1 > 1\right) \leqslant$$

$$\leqslant P\left(\left|\frac{1}{p-k}\sum_{j=1}^{p-k} E_j - 1\right| > \sigma_1\sqrt{(p-k)}\right) \leqslant \frac{1}{p-k}$$

and in consequence

$$(k^* - k - 1)P\left(\frac{1}{p-k}\sum_{j=1}^{p-k} E_j > 2\right) \leqslant \frac{k^* - k - 1}{p-k} \to 0$$

which ends the proof. ■

References

Weijie Su and Emmanuel Candès. Slope is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.*, 44(3):1038–1068, 06 2016. doi: 10.1214/15-AOS1397. URL `https://doi.org/10.1214/15-AOS1397`.

Roman Vershynin. *Introduction to the non-asymptotic analysis of random matrices*, page 210–268. Cambridge University Press, 2012. doi: 10.1017/CBO9780511794308.006.