A Proofs

Lemma A.1. For any $\mathbf{v} \in K$ we have the following inequality

$$\frac{1}{\sqrt{n}}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2 \leq \frac{4}{\sqrt{n}}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2 + \sqrt{\left(\frac{4}{n}\langle \mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v}), \boldsymbol\varepsilon\rangle - \frac{2}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2\right) \vee 0}$$

Proof of Lemma A.1. Since we have that the set $\mathbf{X}K$ is convex conditionally on the matrix $\mathbf{X}$, for any $\mathbf{v} \in K$, we have the inequality

$$\frac{1}{n}\|\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol\beta}\|_2^2 \leq \frac{1}{n}\|\boldsymbol{Y} - \mathbf{X}\mathbf{v}\|_2^2 - \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2.$$

Expanding the square leads to

$$\frac{1}{n}\|\mathbf{X}(\boldsymbol\beta^* - \widehat{\boldsymbol\beta})\|_2^2 \leq \frac{1}{n}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2^2 + \frac{2}{n}\langle \mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v}), \boldsymbol\varepsilon\rangle - \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2.$$

Next note that by the triangle inequality:

$$\frac{1}{n}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2^2 + \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2 - \frac{2}{n}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2 \leq \frac{1}{n}\|\mathbf{X}(\boldsymbol\beta^* - \widehat{\boldsymbol\beta})\|_2^2.$$

We obtain:

$$\frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2 \leq \frac{2}{n}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2 + \frac{2}{n}\langle \mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v}), \boldsymbol\varepsilon\rangle - \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2,$$

Using $c \leq a + b$ implies $c \leq 2\max(a \vee 0, b \vee 0)$ we get

$$\frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2 \leq \max\left(\frac{4}{n}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2, (\frac{4}{n}\langle \mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v}), \boldsymbol\varepsilon\rangle - \frac{2}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2) \vee 0\right)$$

Equivalently, provided that $\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2 \neq 0$, we have

$$\frac{1}{\sqrt{n}}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2 \leq \max\left(\frac{4}{\sqrt{n}}\|\mathbf{X}(\boldsymbol\beta^* - \mathbf{v})\|_2, \sqrt{(\frac{4}{n}\langle \mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v}), \boldsymbol\varepsilon\rangle - \frac{2}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2) \vee 0}\right)$$

The above also obviously holds when $\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2 = 0$. Using $c \leq \max(a, b)$ for $a, b \geq 0$ implies $c \leq a + b$ completes the proof. □

Proof of Theorem 2.3. We start by controlling the empirical process term appearing in Lemma A.1:

$$I := \frac{2}{n}\langle \mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v}), \boldsymbol\varepsilon\rangle - \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol\beta} - \mathbf{v})\|_2^2.$$

Below we state a useful lemma which we take from the classical work Gordon [1988].

Lemma A.2 (Gordon's escape through mesh)**.** Let $D \subset \mathbb{R}^n$ be a cone and $\mathbf{X}$ be an $m \times n$ standard Gaussian matrix. Then

$$\inf_{\mathbf{w} \in D \cap S^{n-1}} \|\mathbf{X}\mathbf{w}\|_2 \geq \sqrt{m-1} - w_1(D) - t,$$
$$\sup_{\mathbf{w} \in D \cap S^{n-1}} \|\mathbf{X}\mathbf{w}\|_2 \leq \sqrt{m} + w_1(D) + t,$$

with probability at least $1 - e^{-t^2/2}$ each.

Note that the unit vector $\frac{\widehat{\boldsymbol\beta} - \mathbf{v}}{\|\widehat{\boldsymbol\beta} - \mathbf{v}\|_2} \in \mathcal{T}_{K,\mathbf{v}} = \{t\mathbf{u} : t \geq 0, \mathbf{u} \in K - \mathbf{v}\}$. By Gordon's escape through a mesh result we have

$$\left\|\mathbf{X}\frac{\widehat{\boldsymbol\beta} - \mathbf{v}}{\|\widehat{\boldsymbol\beta} - \mathbf{v}\|_2}\right\|_2 \geq \inf_{\mathbf{w} \in \mathcal{T}_{K,\mathbf{v}} \cap S^{p-1}} \|\mathbf{X}\mathbf{w}\|_2 \geq \sqrt{n-1} - w_1(\mathcal{T}_{K,\mathbf{v}}) - t,$$

with probability at least $1 - e^{-t^2/2}$. Hence on this event we have

$$\begin{aligned} I &\leq \frac{2}{n}(\sqrt{n-1} - w_1(\mathcal{T}_{K,\mathbf{v}}) - t)\|\widehat{\boldsymbol\beta} - \mathbf{v}\|_2 \times \\ &\times \left\langle \frac{\widehat{\boldsymbol\beta} - \mathbf{v}}{\|\widehat{\boldsymbol\beta} - \mathbf{v}\|_2}, \frac{\mathbf{X}^\top\boldsymbol\varepsilon}{\sqrt{n-1} - w_1(\mathcal{T}_{K,\mathbf{v}}) - t}\right\rangle \\ &- \frac{(\sqrt{n-1} - w_1(\mathcal{T}_{K,\mathbf{v}}) - t)^2\|\widehat{\boldsymbol\beta} - \mathbf{v}\|_2^2}{n} \\ &\leq \frac{\left(\sup_{\mathbf{u} \in \mathcal{T}_{K,\mathbf{v}}, \|\mathbf{u}\|_2 \leq 1}\langle \mathbf{u}, \frac{1}{\sqrt{n}}\mathbf{X}^\top\boldsymbol\varepsilon\rangle\right)^2}{n(\sqrt{\frac{n-1}{n}} - \frac{w_1(\mathcal{T}_{K,\mathbf{v}}) + t}{\sqrt{n}})^2}, \end{aligned}$$

where we used $-a^2 + 2ab \leq b^2$. Note that conditionally on the error term the vector $\frac{1}{\sqrt{n}}\mathbf{X}^\top\boldsymbol\varepsilon \sim \mathcal{N}(0, \mathbf{I}\frac{\|\boldsymbol\varepsilon\|_2^2}{n})$. Set

$$Z = \sup_{\mathbf{u} \in \mathcal{T}_{K,\mathbf{v}} \cap S^{n-1}} \langle \mathbf{u}, \frac{1}{\sqrt{n}}\mathbf{X}^\top\boldsymbol\varepsilon\rangle.$$

By Theorem 5.8 of Boucheron et al. [2013], we have that

$$\mathbb{P}(Z - \mathbb{E}Z \geq \sqrt{2t}\|\boldsymbol\varepsilon\|_2/\sqrt{n}) \leq e^{-t}.$$

Note that conditionally on the error term $\mathbb{E}Z = w_1(\mathcal{T}_{K,\mathbf{v}})\frac{\|\boldsymbol\varepsilon\|_2}{\sqrt{n}}$. It follows that

$$Z \leq (w_1(\mathcal{T}_{K,\mathbf{v}}) + \sqrt{2t})\frac{\|\boldsymbol\varepsilon\|_2}{\sqrt{n}}, \quad \text{(A.1)}$$

with probability at least $1-e^{-t}$, and therefore

$$I \le \frac{\left((w_1(\mathcal{T}_{K,\mathbf{v}}) + \sqrt{2t})\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}}\right)^2}{n(\sqrt{\frac{n-1}{n}} - \frac{w_1(\mathcal{T}_{K,\mathbf{v}})+t}{\sqrt{n}})^2},$$

with probability at least $1-e^{-t}-e^{-t^2/2}$. Therefore by Lemma A.1 conclude that the following holds:

$$\frac{1}{\sqrt{n}}\|\mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v})\|_2 \le \frac{4}{\sqrt{n}}\|\mathbf{X}(\boldsymbol{\beta}^*-\mathbf{v})\|_2 + \frac{\sqrt{2}(w_1(\mathcal{T}_{K,\mathbf{v}}) + \sqrt{2t})\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}}}{\sqrt{n}(\sqrt{\frac{n-1}{n}} - \frac{w_1(\mathcal{T}_{K,\mathbf{v}})+t}{\sqrt{n}})}. \quad \text{(A.2)}$$

Observe that both $\widehat{\boldsymbol{\beta}}-\mathbf{v}$ and $\boldsymbol{\beta}^*-\mathbf{v}$ belong to the cone $\mathcal{T}_{K,\mathbf{v}}$. Hence using Gordon's escape through mesh once again we obtain:

$$\|\widehat{\boldsymbol{\beta}}-\mathbf{v}\|_2 \le \frac{4(\sqrt{n}+w_1(\mathcal{T}_{K,\mathbf{v}})+t)}{(\sqrt{n-1}-w_1(\mathcal{T}_{K,\mathbf{v}})-t)}\|\boldsymbol{\beta}^*-\mathbf{v}\|_2 + \frac{\sqrt{2}(w_1(\mathcal{T}_{K,\mathbf{v}}) + \sqrt{2t})\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}}}{\sqrt{n}(\sqrt{\frac{n-1}{n}} - \frac{w_1(\mathcal{T}_{K,\mathbf{v}})+t}{\sqrt{n}})^2},$$

on an event of probability at least $1-e^{-t}-3e^{-t^2/2}$. Finally, by Chebyshev's inequality we have that

$$\mathbb{P}\left(\left|\frac{\|\boldsymbol{\varepsilon}\|_2^2}{n}-\sigma^2\right| \ge t\right) \le \frac{\operatorname{Var}\varepsilon^2}{nt^2}. \quad \text{(A.3)}$$

Setting $t=\sigma^2$ completes the proof. □

Proof of Theorem 2.5. We need the following lemma whose proof can be found in Plan and Vershynin [2016].

Lemma A.3. Let $S \subset \mathbb{R}^n$ be a star shaped set (i.e., $\lambda S \subset S$ for all $0 \le \lambda \le 1$). Let $x>0$ and suppose that $n \gtrsim w_x^2(S)/x^2$. Then with probability at least $1-\exp(-n/8)$

$$\inf_{\mathbf{w}\in S, \|\mathbf{w}\|_2 \ge x} \|\mathbf{X}\mathbf{w}\|_2 \ge \sqrt{cn}\|\mathbf{w}\|_2,$$

for some absolute constant $c>0$.

By Lemma A.1 we have

$$\frac{1}{\sqrt{n}}\|\mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v})\|_2 \le \frac{4}{\sqrt{n}}\|\mathbf{X}(\boldsymbol{\beta}^*-\mathbf{v})\|_2 + \sqrt{2\left\{\frac{2}{n}\langle \mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v}), \boldsymbol{\varepsilon}\rangle - \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v})\|_2^2\right\} \vee 0},$$

Suppose that $\|\widehat{\boldsymbol{\beta}}-\mathbf{v}\|_2 \ge x$ otherwise the proposition is clearly true. Let $a = \frac{x}{\|\widehat{\boldsymbol{\beta}}-\mathbf{v}\|_2} \le 1$. Since $0 \le a \le 1$ we have that $\mathbf{u} = a\widehat{\boldsymbol{\beta}} + (1-a)\mathbf{v} \in K$. This implies the simple identity $a(\widehat{\boldsymbol{\beta}}-\mathbf{v}) = (\mathbf{u}-\mathbf{v})$. Therefore

$$\begin{aligned}
&\frac{2}{n}\langle \mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v}), \boldsymbol{\varepsilon}\rangle - \frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v})\|_2^2 \\
&\quad \le \frac{2}{\sqrt{n}}\langle \widehat{\boldsymbol{\beta}}-\mathbf{v}, \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{\sqrt{n}}\rangle - c\|\widehat{\boldsymbol{\beta}}-\mathbf{v}\|_2^2 \\
&\quad = \frac{2}{a\sqrt{n}}\langle \mathbf{u}-\mathbf{v}, \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{\sqrt{n}}\rangle - c\frac{\|\mathbf{u}-\mathbf{v}\|_2^2}{a^2} \\
&\quad \le \frac{2x\sqrt{c}}{a\sqrt{n}x\sqrt{c}} \sup_{\mathbf{w}\in K-\mathbf{v}, \|\mathbf{w}\|_2 \le x}\langle \mathbf{w}, \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{\sqrt{n}}\rangle - c\frac{x^2}{a^2} \\
&\quad \le \left(\frac{\sup_{\mathbf{w}\in K-\mathbf{v}, \|\mathbf{w}\|_2 \le x}\langle \mathbf{w}, \frac{\mathbf{X}^\top\boldsymbol{\varepsilon}}{\sqrt{n}}\rangle}{\sqrt{n}x\sqrt{c}}\right)^2 \\
&\quad \le \left(\frac{w_x(K-\mathbf{v})\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}} + x\sqrt{2t}\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}}}{\sqrt{n}x\sqrt{c}}\right)^2.
\end{aligned}$$

where we used $-a^2+2ab \le b^2$ in the next to last bound, and the last bound holds on an event of probability at least $1-e^{-t}$ by the same argument as in (A.1) in the proof of Theorem 2.3. Next by Lemma A.3

$$\frac{1}{n}\|\mathbf{X}(\widehat{\boldsymbol{\beta}}-\mathbf{v})\|_2^2 \ge c\|\widehat{\boldsymbol{\beta}}-\mathbf{v}\|_2^2,$$

and therefore we conclude

$$\|\mathbf{v}-\widehat{\boldsymbol{\beta}}\|_2 \le \frac{4}{\sqrt{c}\sqrt{n}}\|\mathbf{X}(\boldsymbol{\beta}^*-\mathbf{v})\|_2 + \sqrt{2}\frac{w_x(K-\mathbf{v})\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}} + x\sqrt{2t}\frac{\|\boldsymbol{\varepsilon}\|_2}{\sqrt{n}}}{\sqrt{n}xc}.$$

Using Chebyshev's inequality as in (A.3) completes the proof. □

Remark A.4. Observe that by Corollary 5.35 of Vershynin [2012], we have

$$\frac{\|\mathbf{X}(\boldsymbol{\beta}^*-\mathbf{v})\|_2}{\sqrt{n}} \le \left(1+\sqrt{\frac{p}{n}}+\sqrt{\frac{2t}{n}}\right)\|\boldsymbol{\beta}^*-\mathbf{v}\|_2,$$

with probability at least $1-e^{-t}$.

Proof of Corollary 2.6. The proof is very similar to that of Theorem 2.3. We omit the details. □

Proof of Corollary 2.7. The proof is very similar to that of Theorem 2.5. We omit the details. □

Proof of Lemma 3.1. Let $k=\lfloor p/\ell \rfloor$ and $r=p-k\ell$. Fix $v_1=\beta_1$. For $i \in \{2,\dots,k\}$ take $v_i = v_1 + (i-1)\frac{L}{p}$. Next for $i \in \{k+1,\dots,2k\}$ take $v_i = v_k + (i-k)\frac{L}{p}\operatorname{sign}(\beta_k - v_k)$. For $i \in \{2k+1,\dots,3k\}$ take $v_i =$

$v_{2k} + (i-2k)\frac{L}{p}\operatorname{sign}(\beta_{2k} - v_{2k})$, and so on. We will now show by induction that $|v_i - \beta_i| \leq 2k\frac{L}{p}$ for all $i \in [p]$. Note that the final result follows from this observation. This is clearly true over the first k numbers, so suppose it is true for the first $m-1 \geq k$ numbers. Let q be the largest integer such that $qk < m$. We have

$$\begin{aligned}|v_m - \beta_m| & \\ &\leq |\beta_m - \beta_{qk}| + |v_m - \beta_{qk}| \\ &= |\beta_m - \beta_{qk}| \\ &+ |\beta_{qk} - v_{qk} - (m-qk)\frac{L}{p}\operatorname{sign}(\beta_{qk} - v_{qk})| \\ &\leq \frac{(m-qk)L}{p} + ||\beta_{qk} - v_{qk}| - (m-qk)\frac{L}{p}| \\ &\leq |\beta_{qk} - v_{qk}| \vee 2\frac{(m-qk)L}{p} \leq \frac{2kL}{p},\end{aligned}$$

which completes the proof. □

Proof of Lemma 3.2. Take any vector $\mathbf{u} \in K$. Observe that $-\frac{L}{p} - (v_i - v_{i-1}) \leq u_i - u_{i-1} - (v_i - v_{i-1}) \leq \frac{L}{p} - (v_i - v_{i-1})$. It follows that when $v_i - v_{i-1} = \frac{L}{p}$ we have that $u_i - v_i$ form a decreasing sequence. On the other hand on stretches where $v_i - v_{i-1} = -\frac{L}{p}$ we have an increasing sequence. Clearly, this conclusion remains true even when we scale $\mathbf{u} - \mathbf{v}$ by an arbitrary positive constant. This completes the proof. □

Proof of Corollary 3.4. We start by showing the first part. Take a $\mathbf{v}$ with $\ell + 1 \leq 2\ell$ affine pieces which approximates $\boldsymbol{\beta}^*$ as in Lemma 3.1. Using Corollary 2.4 we have

$$\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2 \lesssim \frac{L\sqrt{p}}{\ell} + \sqrt{\frac{\ell\log(ep/(\ell+1))}{n}}\sigma + \frac{\sigma}{\sqrt{n}}.$$

Setting $\ell = \sqrt[3]{\frac{L^2pn}{\log(ep)}}$ gives a rate

$$\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2 \lesssim \frac{[\log(ep)L]^{1/3}p^{1/6}}{n^{1/3}}(1+\sigma) + \frac{\sigma}{\sqrt{n}}.$$

The second part is a direct consequence to (3.1) and Corollary 2.4. To elaborate, select $\mathbf{v} = \boldsymbol{\beta}^*$ in Corollary 2.4. By (3.1), and the assumptions on $\boldsymbol{\beta}^*$ we have that

$$w_1(\mathcal{T}_{K,\boldsymbol{\beta}^*}) \leq \sqrt{(\ell+1)\log(ep/(\ell+1))}.$$

Since $\ell+1 \leq 2\ell$ and $\log ep/(\ell+1) \leq \log ep/\ell$ the second bound follows by Corollary 2.4. □

Proof of Corollary 3.5. Set $\mathbf{v} = \boldsymbol{\beta}^*$ in Corollary 2.4. To obtain an upper bound on $w_1(\mathcal{T}_{K,\boldsymbol{\beta}^*})$ one may use Proposition 3.1 of Bellec et al. [2018] and the same logic as the bound on the tangent cone of Lipschitz sequences. We omit the details. □

Proof of Corollary 3.6. Chatterjee et al. [2014] showed (proof of Theorem 2.2) that

$$w_x(K - \boldsymbol{\beta}^*) \leq \Omega_1\sqrt{(\beta_n - \beta_1)x}p^{1/4} + \Omega_2 x^2,$$

for some absolute constants Ω_1 and Ω_2. Plugging in the bound of Theorem 2.5 (with $\mathbf{v} = \boldsymbol{\beta}^*$) we obtain:

$$\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2 \lesssim \frac{\sqrt{(\beta_n - \beta_1)x}p^{1/4} + x^2 + x\sqrt{t}}{\sqrt{n}x}\sigma + x,$$

and therefore

$$\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2 \lesssim \frac{\sqrt{(\beta_n - \beta_1)}p^{1/4}}{\sqrt{n}\sqrt{x}}\sigma + \frac{\sigma}{\sqrt{n}} + x,$$

holds with probability at least .99. Setting x such that equates the two rates completes the proof.

□

Proof of Corollary 3.7. Set $\mathbf{v} = \boldsymbol{\beta}^*$ in Corollary 2.4. To obtain an upper bound on $w_1(\mathcal{T}_{K,\boldsymbol{\beta}^*})$ one may use Proposition 4.2 of Bellec et al. [2018] and the same logic as the bound on the tangent cone of Lipschitz sequences. To elaborate, we have that

$$w_1^2(\mathcal{T}_{K,\boldsymbol{\beta}^*}) \leq \delta(\mathcal{T}_{K,\boldsymbol{\beta}^*}) \leq 8q(\boldsymbol{\beta}^*)\log\left(\frac{ep}{q(\boldsymbol{\beta}^*)}\right),$$

where the final bound follows by Proposition 4.2 of Bellec et al. [2018], and $q(\boldsymbol{\beta}^*) = \ell$ denotes the number of affine pieces of $\boldsymbol{\beta}^*$. Therefore by Corollary 2.4 (with $\mathbf{v} = \boldsymbol{\beta}^*$), we have that

$$\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2 \lesssim \frac{w_1(\mathcal{T}_{K,\boldsymbol{\beta}^*})}{\sqrt{n}} + \frac{\sigma}{\sqrt{n}} \lesssim \sqrt{\frac{\ell\log ep/\ell}{n}} + \frac{\sigma}{\sqrt{n}},$$

which is what we wanted to show. □

Proof of Corollary 3.8. We start by applying Proposition 3.1 of Chatterjee et al. [2016] which shows that

$$\begin{aligned}w_x(K - \boldsymbol{\beta}^*) &\lesssim ((\max\beta_i - \min\beta_i) + 1)^{1/4}p^{1/8}x^{3/4} \\ &+ \sqrt{\log p}x + \frac{x^2}{4}.\end{aligned}$$

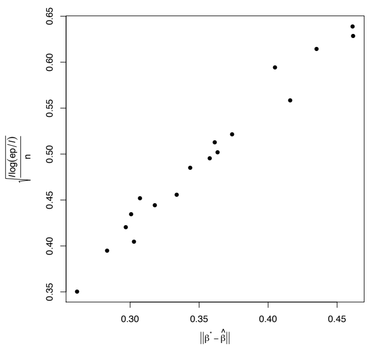
It is simple to see that setting

$$x = \frac{\sigma^{4/5}((\max\beta_i - \min\beta_i) + 1)^{1/5}p^{1/10}}{n^{2/5}}$$

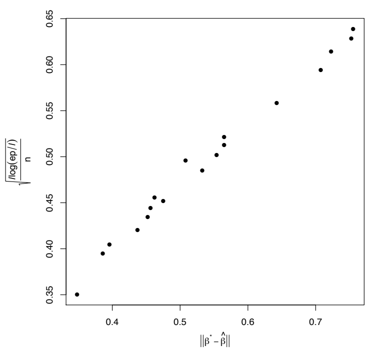
and applying Theorem 2.5 completes the proof. □

A.1 Additional Numerical Simulations

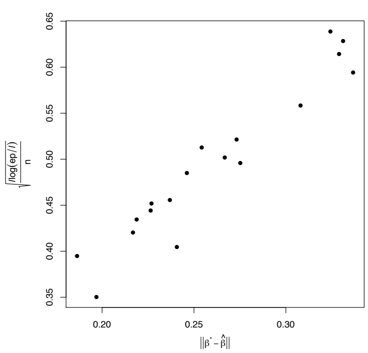
In this section we present additional simulation studies. To verify our findings we simulated 100 repetitions of each of the 6 settings from Section 4, for



(a) Lipschitz Regression

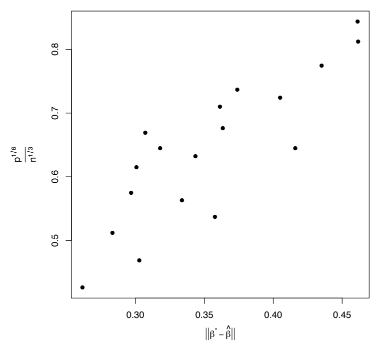


(b) Monotone Regression

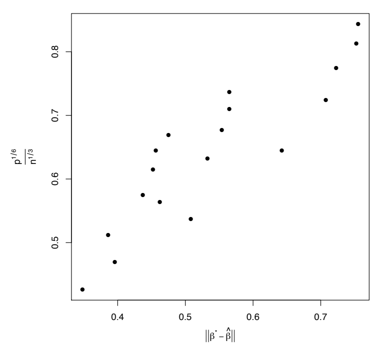


(c) Convex Regression

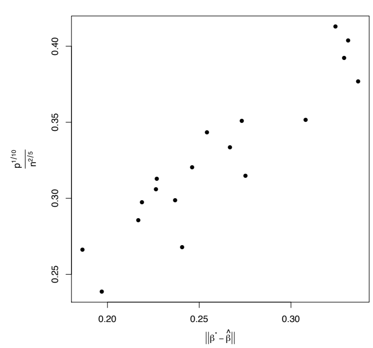
Figure 4: Example of the adaptive rates vs the averaged values of $\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2$ over 100 repetitions. In all three examples we observe a linear alignment confirming our theoretical findings.



(a) Lipschitz Regression



(b) Monotone Regression



(c) Convex Regression

Figure 5: Examples of the non-adaptive rates vs the averaged values of $\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2$ over 100 repetitions. We observe linear alignment confirming our theoretical findings. The points appear to be more variable as compared to those on Fig 4. We attribute that to the greater variability of the estimates in the non-adaptive setting (as is also evident on Fig 3).

each of the following choices of $n \in \{50, 75, 100\}$ and $p \in \{60, 180, 360, 540, 720, 900\}$. We then plot the averaged values of $\|\boldsymbol{\beta}^* - \widehat{\boldsymbol{\beta}}\|_2$ over the 100 simulations vs the corresponding adaptive or non-adaptive function of n and p. The results for the adaptive and non-adaptive rates can be found in Figs 4 and 5 respectively. We observe linear alignment of the points confirming our theoretical findings. In addition it appears that the non-adaptive rates appear to be more variable than the adaptive rates, which is possibly due to the greater variability of the estimates.