Supplementary Material for "An Asymptotic Rate for the LASSO Loss"

Cynthia Rush
Columbia University

1 SUPPLEMENTARY MATERIAL

In the supplementary material we include a number of Lemmas that will be used in the technical proofs of our main results. When the proof is straightforward, it is omitted.

Lemma 1 (Concentration of Square Roots). *Let* $c \neq 0$ *and* X_n *be a random variable. If* $P\left(X_n^2 - c^2 \geq \epsilon\right) \leq Ke^{-\kappa n \epsilon^2}$, *then* $P\left(|X_n| - |c| \geq \epsilon\right) \leq Ke^{-\kappa n |c|^2 \epsilon^2}$.

Lemma 2 (Gaussian Concentration). *For i.i.d. standard Gaussian random variables* $Z_1, Z_2, \ldots, Z_N$ *and* $0 \leq \epsilon \leq 1$,

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - 1\Big| \geq \epsilon\Big) \leq 2\exp\Big\{\frac{-n\epsilon^2}{8}\Big\}.$$

Lemma 3. *Let* $\ker(\mathbf{X})$ *denote the kernel of a matrix* $\mathbf{X} \in \mathbb{R}^{n\times p}$ *and* $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$ *vectors with* $\mathbf{v}_1 \in ker(\mathbf{X})^{\perp}$. *Then* $\sigma^2_{min}\|\mathbf{v}_2\|^2 \leq \|\mathbf{X}\mathbf{v}_2\|^2 \leq \sigma^2_{max}\|\mathbf{v}_2\|^2$ *where* $\sigma^2_{min}(\mathbf{X})$ *and* $\sigma^2_{max}(\mathbf{X})$ *are the minimum and maximum singular values of* $\mathbf{X}$ *and* $\hat{\sigma}^2_{min}\|\mathbf{v}_1\|^2 \leq \|\mathbf{X}\mathbf{v}_1\|^2 \leq \sigma^2_{max}\|\mathbf{v}_1\|^2$, *where* $\hat{\sigma}_{min}(\mathbf{X})$ *is the minimum,* non-zero *singular value of* $\mathbf{X}$.

Lemma 4 ([Kashin, 1977]). *For any dimension* $p > 0$, *with probability at least* $1 - 2^{-p}$, *for any uniformly random subspace* $\mathcal{V}_{(p,r)}$ *with dimension* $\lfloor p(1-r)\rfloor$ *where* $0 \leq r \leq 1$ *is a positive constant, for all vectors* $\mathbf{v} \in \mathcal{V}_{(p,r)}$,

$$c_r\|\mathbf{v}\| \leq \|\mathbf{v}\|_1/\sqrt{p},$$

where c_r *is a universal constant (not depending on* p*).*

Lemma 5. *[Rudelson and Vershynin, 2010] Let* $\mathbf{X}$ *be an* $n \times p$ *random matrix with i.i.d.* $\mathcal{N}(0, 1/n)$ *entries and* $n \leq p$. *If* $\sigma_{\max}(\mathbf{X})$ *and* $\hat{\sigma}_{\min}(\mathbf{X})$ *are the minimum and maximum* non-zero *singular values of* $\mathbf{X}$ *then for* $\delta = n/p$, *the probabilities* $P(\sigma_{\max}(\mathbf{X}) \geq 1 + \sqrt{\delta} + \epsilon)$ *and* $P(\hat{\sigma}_{\min}(\mathbf{X}) \leq 1 - \sqrt{\delta} - \epsilon)$ *are upper bounded by* $e^{-n\epsilon^2/2}$. *Let* $\widetilde{\mathbf{X}}$ *be an* $n \times p$ *random matrix with i.i.d.* $\mathcal{N}(0, 1/n)$ *entries and* $n \geq p$. *If* $\sigma_{\max}(\widetilde{\mathbf{X}})$ *and* $\sigma_{\min}(\widetilde{\mathbf{X}})$ *are the largest and smallest singular values of* $\widetilde{\mathbf{X}}$ *then, the probabilities* $P(\sigma_{\max}(\widetilde{\mathbf{X}}) \geq 1 + \frac{1}{\sqrt{\delta}} + \epsilon)$ *and* $P(\sigma_{\min}(\widetilde{\mathbf{X}}) \leq 1 - \frac{1}{\sqrt{\delta}} - \epsilon)$ *are upper bounded by* $e^{-n\epsilon^2/2}$.

Proof. The result for $\widetilde{\mathbf{X}}$ follows directly from the cited result. The result for $\mathbf{X}$ uses the following fact: let $\mathbf{M}_1$ be an $m \times n$ matrix and let $\mathbf{M}_2$ be an $n \times m$ matrix, with $n \geq m$, then the n eigenvalues of $\mathbf{M}_2\mathbf{M}_1$ are the m eigenvalues of $\mathbf{M}_1\mathbf{M}_2$ with the extra eigenvalues being 0. So we note that if $\mathbf{X}$ is an $n \times p$ matrix with $p \geq n$ then the p eigenvalues of $\mathbf{X}^*\mathbf{X}$ are the n eigenvalues of $\mathbf{X}\mathbf{X}^*$ with the extra eigenvalues being 0. The cited result tells us $P(\sigma_{\max}(\mathbf{X}^*) \geq 1 + \sqrt{\delta} + \epsilon)$ and $P(\sigma_{\min}(\mathbf{X}^*) \leq 1 - \sqrt{\delta} - \epsilon)$ are upper bounded by $e^{-n\epsilon^2/2}$. We note that $\sigma_{\min}(\mathbf{X}^*) = \lambda_{\min}(\mathbf{X}\mathbf{X}^*)$ and $\sigma_{\max}(\mathbf{X}^*) = \lambda_{\max}(\mathbf{X}\mathbf{X}^*)$. The lemma result follows from the stated fact.

□

Lemma 6. *Let* $\{\tau_t^2\}_{t\geq 0}$ *be the state evolution sequence and let* $\theta_t = \alpha\tau_t$ *for all* $t \geq 0$. *Then,*

$$\Big|P(|B - \tau_{t-1}Z_t| > \theta_t) - P(|B - \tau_* Z| \geq \theta_*)\Big| \leq \frac{2\mathbb{E}|B|}{\min\{\tau_0^2, \tau_*^2\}}|\tau_t - \tau_*|.$$

In the above $B \sim p_{\boldsymbol{\beta}}$ *is independent of* Z_t, Z *both standard Gaussian.*

Proof. We first note that

$$P(|B - \tau_* Z| \geq \theta_*) = \mathbb{E}\{P(|B - \tau_* Z| \geq \theta_* \,|\, B)\},$$

and so we first study the conditional probability

$$P(|B - \tau_* Z| \geq \theta_* \,|\, B = b) = P(|b - \tau_* Z| \geq \theta_*).$$

First, using $\Phi(\cdot)$ to denote the Gaussian cdf, namely $\Phi(x) = P(Z \leq x)$ for standard Gaussian Z, and the fact that $\theta_* = \alpha\tau_*$,

$$\begin{aligned} &P(|b - \tau_* Z| \geq \theta_*) \\ &= P(b - \tau_* Z \geq \theta_*) + P(b - \tau_* Z \leq -\theta_*) \\ &= P(Z \leq (b/\tau_*) - \alpha) + P(Z \geq (b/\tau_*) + \alpha) \\ &= \Phi((b/\tau_*) - \alpha) + \Phi((-b/\tau_*) - \alpha). \end{aligned}$$

Similarly,

$$P(|b-\tau_t Z_t|\geq \theta_t) = \Phi((b/\tau_t)-\alpha) + \Phi((-b/\tau_t)-\alpha),$$

and by Jensen's inequality,

$$\begin{aligned}&\Big|P(|B-\tau_t Z_t|>\theta_t) - P(|B-\tau_* Z|\geq \theta_*)\Big| \\ &= \Big|\mathbb{E}\Big\{P(|B-\tau_t Z_t|>\theta_t|B) - P(|B-\tau_* Z|\geq \theta_*|B)\Big\}\Big| \\ &\leq \mathbb{E}\Big\{\Big|P(|B-\tau_t Z_t|>\theta_t|B) - P(|B-\tau_* Z|\geq \theta_*|B)\Big|\Big\}.\end{aligned} \tag{1}$$

Then, letting $h(b) := \Big|P(|B-\tau_t Z_t|>\theta_t|B=b) - P(|B-\tau_* Z|\geq \theta_*|B=b)\Big|$, we have

$$\begin{aligned}h(b) &= \Big|\mathbb{E}\Big\{\mathbb{I}\{|b-\tau_t Z_t|>\theta_t\} - \mathbb{I}\{|b-\tau_* Z|\geq \theta_*\}\Big\}\Big| \\ &= \Big|P(|b-\tau_t Z_t|>\theta_t) - P(|b-\tau_* Z|\geq \theta_*)\Big| \\ &\leq \mathbb{E}\Big|\Phi((b/\tau_t)-\alpha) - \Phi((b/\tau_*)-\alpha)\Big| \\ &\quad + \mathbb{E}\Big|\Phi((-b/\tau_t)-\alpha) - \Phi((-b/\tau_*)-\alpha)\Big|\end{aligned} \tag{2}$$

Now we note that the Gaussian cdf is Lipschitz. Indeed, for $x>y$, and $\phi(\cdot)$ the Gaussian pdf,

$$|\Phi(x)-\Phi(y)| = \int_y^x \phi(z)dz \leq |x-y|,$$

and so (2) gives the following upper bound

$$h(b) \leq |b|\Big|\frac{1}{\tau_t}-\frac{1}{\tau_*}\Big| + \mathbb{E}|B|\Big|\frac{1}{\tau_*}-\frac{1}{\tau_t}\Big| = 2|b|\frac{|\tau_*-\tau_t|}{\tau_t\tau_*}. \tag{3}$$

Finally, note from (1) and (3) that $T \leq \mathbb{E}[h(B)] \leq \frac{2\mathbb{E}|B|}{\min\{\tau_0^2,\tau_*^2\}}|\tau_*-\tau_{t-1}|$. ☐

Lemma 7. *Let $\lambda>0$ and $\alpha=\alpha(\lambda)$, then for $0\leq t\leq T^*$,*

$$P\Big(\Big|\frac{1}{p}\|\boldsymbol{\beta}^t\|^2 - \mathbb{E}\Big[(\eta(B+\tau_t Z;\theta_t))^2\Big]\Big| \geq \epsilon\Big) \leq K_t e^{-\kappa_t n\epsilon^2}. \tag{4}$$

Moreover, with $\hat{\mathsf{B}} := \tilde{\mathsf{C}}\mathsf{C}^2 + 4\tilde{\mathsf{C}}\mathsf{C}/c_{min}$ and $\tilde{\mathsf{B}} = 1 + \max(6\sigma_0^2 + 6\max\{\tau_0^2,\tau_*^2\}(1+\alpha^2), \tilde{\mathsf{C}})$, where $\tilde{\mathsf{C}}$ and C are independent of t and are defined in the proof below, c_{min} is the concentrating values for the minimum (non-zero) singular values of the matrix $\mathbf{X}$ as defined in Lemma 4, Condition (4) in the main document, σ_0^2 is defined in Eq. (8) in the main document, and σ^2 is the noise variance in the problem, we have

$$P(\|\widehat{\boldsymbol{\beta}}(\lambda)\|^2/p \geq \hat{\mathsf{B}}) \leq Ke^{-\kappa n},$$

and

$$P(\|\boldsymbol{\beta}^t\|^2/p \geq \tilde{\mathsf{B}}) \leq K_t e^{-\kappa_t n}.$$

Proof. Note that the concentration for the norm of $\boldsymbol{\beta}^t$ follows from Theorem 1 in the main document using the pseudo-Lipschitz function $\phi(a,b)=a^2$. Namely, we know that

$$P\Big(\Big|\frac{1}{p}\|\boldsymbol{\beta}^t\|^2 - \mathbb{E}\Big[(\eta(B+\tau_t Z;\theta_t))^2\Big]\Big| \geq \epsilon\Big) \leq K_t e^{-\kappa_t n\epsilon^2},$$

where, for each $0\leq t\leq T^*$,

$$\begin{aligned}&\mathbb{E}\Big[(\eta(B+\tau_t Z;\theta_t))^2\Big] \\ &\overset{(a)}{=} \mathbb{E}\Big[(B+\tau_t Z-\theta_t)^2\mathbb{I}\{B+\tau_t Z>\theta_t\}\Big] \\ &\qquad + \mathbb{E}\Big[(B+\tau_t Z+\theta_t)^2\mathbb{I}\{B+\tau_t Z<-\theta_t\}\Big] \\ &\overset{(b)}{\leq} \mathbb{E}[(B+\tau_t Z-\theta_t)^2] + \mathbb{E}[(B+\tau_t Z+\theta_t)^2] \\ &\overset{(c)}{\leq} 6\mathbb{E}\{B^2\} + 6\tau_t^2\mathbb{E}\{Z^2\} + 6\theta_t^2 \\ &\leq 6\sigma_0^2 + 6\tau_t^2 + 6\alpha^2\tau_t^2 \overset{(d)}{\leq} 6\sigma_0^2 + 6\max\{\tau_0^2,\tau_*^2\}(1+\alpha^2).\end{aligned}$$

In the chain above, step (a) follows from the definition of the soft-thresholding function in Eq. (4) of the main document, step (b) since the indicator function is upper bounded by 1 and we're considering the expectation of a positive term, step (c) from [Rush and Venkataramanan, 2015, Lemma C.3], and step (d) since $\tau_t^2 \leq \max\{\tau_0^2,\tau_*^2\}$. The result in the lemma statement follows since

$$\begin{aligned}&P(\|\boldsymbol{\beta}^t\|^2/p \geq \tilde{\mathsf{B}}) \\ &\leq P(\|\boldsymbol{\beta}^t\|^2/p \geq \mathbb{E}[(\eta(B+\tau_t Z;\theta_t))^2]+\epsilon) \\ &\leq P(|\|\boldsymbol{\beta}^t\|^2/p - \mathbb{E}[(\eta(B+\tau_t Z;\theta_t))^2]|\geq \epsilon).\end{aligned}$$

The first inequality above follows since $\tilde{\mathsf{B}} \geq 1 + 6\sigma_0^2 + 6\max\{\tau_0^2,\tau_*^2\}(1+\alpha^2) \geq \epsilon + \mathbb{E}[(\eta(B+\tau_t Z;\theta_t))^2]$.

Now consider concentration for $\widehat{\boldsymbol{\beta}}$. We will first show that $\mathcal{C}(\widehat{\boldsymbol{\beta}})$ is lower bounded by a constant with high probability. We have the following upper bound on the LASSO cost function:

$$\mathcal{C}(\widehat{\boldsymbol{\beta}}) \leq \mathcal{C}(0) = \frac{1}{2}\|\mathbf{y}\|^2 = \frac{1}{2}\|\mathbf{X}\boldsymbol{\beta}+\mathbf{w}\|^2 \tag{5}$$

$$\overset{(a)}{\leq} \|\mathbf{X}\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2 \overset{(b)}{\leq} \sigma_{max}^2(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2, \tag{6}$$

where step (a) follows by Cauchy-Schwarz and step (b) from Lemma 3. Now for $\epsilon' \in (0,1)$, note that $(c_{max}+\epsilon')(\sigma_0^2+\epsilon') + \sigma^2 + \epsilon' \leq (c_{max}+1)(\sigma_0^2+1) + \sigma^2 + 1$ and we label $\mathsf{C} := (c_{max}+1)(\sigma_0^2+1)+\sigma^2+1$. Considering the above, then,

$$\begin{aligned}&P((\sigma_{max}^2(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2)/p \geq \mathsf{C}) \\ &\leq P(\sigma_{max}^2(\mathbf{X}) \geq c_{max}+\epsilon') + P(\|\boldsymbol{\beta}\|^2/p \geq \sigma_0^2+\epsilon') \\ &\qquad + P(\|\mathbf{w}\|^2/p \geq \sigma^2+\epsilon') \\ &\overset{(c)}{\leq} Ke^{-\kappa n\epsilon'^2}.\end{aligned} \tag{7}$$

In the above, step (c) follows from Lemma 5, the assumption given in Eq. (8) in the main document, and Lemma 2. Then using (6) and (7),

$$P(\mathcal{C}(\widehat{\boldsymbol{\beta}})/p \geq \mathsf{C}) \leq P((\sigma^2_{max}(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2)/p \geq \mathsf{C}) \leq Ke^{-\kappa n\epsilon'^2}. \tag{8}$$

Now we will relate $\|\widehat{\boldsymbol{\beta}}\|^2/p$ to $\mathcal{C}(\widehat{\boldsymbol{\beta}})/p$ and other terms lower-bounded by a constant with high probability. We write $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\perp} + \widehat{\boldsymbol{\beta}}^{\parallel}$ where $\widehat{\boldsymbol{\beta}}^{\perp} \in ker(\mathbf{X})^{\perp}$ and $\widehat{\boldsymbol{\beta}}^{\parallel} \in ker(\mathbf{X})$. Since $\widehat{\boldsymbol{\beta}}^{\parallel} \in \ker(\mathbf{X})$ and $\ker(\mathbf{X})$ is a random subspace of size $p - n = p(1-\delta)$, by Kashin Theorem (Lemma 4), we have that for some constant $\nu_1 = \nu_1(\delta)$,

$$P(\|\widehat{\boldsymbol{\beta}}^{\parallel}\|_2^2 \geq \nu_1 \|\widehat{\boldsymbol{\beta}}^{\parallel}\|_1^2/N) \leq Ke^{-p}. \tag{9}$$

Denote the event $\{\|\widehat{\boldsymbol{\beta}}^{\parallel}\|_2^2 \geq \nu_1\|\widehat{\boldsymbol{\beta}}\|_1^2/p\}$ by $\mathcal{E}$, and by the above $P(\mathcal{E}) \leq Ke^{-p}$. Conditioned on $\mathcal{E}^c$, we have the following bound

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}}\|^2 &= \|\widehat{\boldsymbol{\beta}}^{\parallel}\|^2 + \|\widehat{\boldsymbol{\beta}}^{\perp}\|^2 \overset{(a)}{\leq} \frac{\nu_1}{p}\|\widehat{\boldsymbol{\beta}}^{\parallel}\|_1^2 + \|\widehat{\boldsymbol{\beta}}^{\perp}\|^2 \\ &\overset{(b)}{\leq} \frac{2\nu_1}{p}\|\widehat{\boldsymbol{\beta}}\|_1^2 + \frac{2\nu_1}{p}\|\widehat{\boldsymbol{\beta}}^{\perp}\|_1^2 + \|\widehat{\boldsymbol{\beta}}^{\perp}\|^2 \\ &\overset{(c)}{\leq} \frac{2\nu_1}{p}\Big(\frac{1}{\lambda}\mathcal{C}(\widehat{\boldsymbol{\beta}})\Big)^2 + (2\nu_1+1)\|\widehat{\boldsymbol{\beta}}^{\perp}\|^2,\end{aligned} \tag{10}$$

where step (a) holds since we are conditioning on $\mathcal{E}^c$, step (b) from the triangle inequality and Cacuh-Schwarz, and step (c) since $\lambda\|\widehat{\boldsymbol{\beta}}\|_1 \leq \mathcal{C}(\widehat{\boldsymbol{\beta}})$ by the definition of the cost function and $\|\widehat{\boldsymbol{\beta}}^{\perp}\|_1 \leq \sqrt{p}\|\widehat{\boldsymbol{\beta}}^{\perp}\|$ by Cauchy-Schwarz. Now we bound the second term on the RHS of (10):

$$\begin{aligned}\|\widehat{\boldsymbol{\beta}}^{\perp}\|^2 &\overset{(d)}{\leq} \frac{\|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\perp}\|^2}{\hat{\sigma}^2_{min}(\mathbf{X})} \leq \frac{\|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\perp} - \mathbf{y} + \mathbf{y}\|^2}{\hat{\sigma}^2_{min}(\mathbf{X})} \\ &\overset{(e)}{\leq} \frac{2\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{\perp}\|^2}{\hat{\sigma}^2_{min}(\mathbf{X})} + \frac{2\|\mathbf{y}\|^2}{\hat{\sigma}^2_{min}(\mathbf{X})} \\ &\leq \frac{2\mathcal{C}(\widehat{\boldsymbol{\beta}})}{\hat{\sigma}^2_{min}(\mathbf{X})} + \frac{2\|\mathbf{X}\boldsymbol{\beta} + \mathbf{w}\|^{\mathbf{2}}}{\hat{\sigma}^2_{min}(\mathbf{X})}.\end{aligned} \tag{11}$$

In the above, step (d) follows from the fact that $\hat{\sigma}^2_{min}(\mathbf{X})\|\widehat{\boldsymbol{\beta}}^{\perp}\|^2 \leq \|\mathbf{X}\widehat{\boldsymbol{\beta}}^{\perp}\|^2$ by Lemma 3 and step (e) by Cauchy-Schwarz. Next note

$$\|\mathbf{X}\boldsymbol{\beta}+\mathbf{w}\|^2 \leq 2(\|\mathbf{X}\boldsymbol{\beta}\|^2+\|\mathbf{w}\|^2) \leq 2(\sigma^2_{max}(\mathbf{X})\|\boldsymbol{\beta}\|^2+\|\mathbf{w}\|^2), \tag{12}$$

by Cauchy-Schwarz and Lemma 3. Now plugging (11) and (12) into (10) we have for some constant $\tilde{\mathsf{C}} = \max\{\frac{2\nu_1}{\lambda^2}, 4(2\nu_1+1)\} > 0$,

$$\frac{1}{p}\|\widehat{\boldsymbol{\beta}}\|^2 \leq \tilde{\mathsf{C}}\Big(\frac{\mathcal{C}(\widehat{\boldsymbol{\beta}})}{p}\Big)^2 + \frac{\tilde{\mathsf{C}}(\mathcal{C}(\widehat{\boldsymbol{\beta}}) + \sigma^2_{max}(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2)}{p\hat{\sigma}^2_{min}(\mathbf{X})}. \tag{13}$$

Now considering the above,

$$\begin{aligned}&P\Big(\frac{1}{p}\|\widehat{\boldsymbol{\beta}}\|^2 \geq \tilde{\mathsf{C}}\mathsf{C}^2 + \frac{2\tilde{\mathsf{C}}\mathsf{C}}{c_{min} - \epsilon'}\Big) \\ &\leq P(\mathcal{E}) + P\Big(\frac{1}{p}\|\widehat{\boldsymbol{\beta}}\|^2 \geq \tilde{\mathsf{C}}\mathsf{C}^2 + \frac{2\tilde{\mathsf{C}}\mathsf{C}}{c_{min} - \epsilon'}\Big|\mathcal{E}^c\Big).\end{aligned} \tag{14}$$

First note $P(\mathcal{E}) \leq Ke^{-p}$ by (9), and

$$\begin{aligned}&P\Big(\frac{1}{p}\|\widehat{\boldsymbol{\beta}}\|^2 \geq \tilde{\mathsf{C}}\mathsf{C}^2 + \frac{2\tilde{\mathsf{C}}\mathsf{C}}{c_{min} - \epsilon'}\Big|\mathcal{E}^c\Big) \\ &\overset{(a)}{\leq} P\Big(\tilde{\mathsf{C}}\Big(\frac{\mathcal{C}(\widehat{\boldsymbol{\beta}})}{p}\Big)^2 + \frac{\tilde{\mathsf{C}}(\mathcal{C}(\widehat{\boldsymbol{\beta}}) + \sigma^2_{max}(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2)}{N\hat{\sigma}^2_{min}(\mathbf{X})} \\ &\qquad\qquad \geq \tilde{\mathsf{C}}\mathsf{C}^2 + \frac{2\tilde{\mathsf{C}}\mathsf{C}}{c_{min} - \epsilon'}\Big) \\ &\overset{(b)}{\leq} P\Big(\frac{\mathcal{C}(\widehat{\boldsymbol{\beta}})}{p} \geq \mathsf{C}\Big) + P\Big(\sigma^2_{max}(\mathbf{X})\frac{\|\boldsymbol{\beta}\|^2}{p} + \frac{\|\mathbf{w}\|^2}{p} \geq \mathsf{C}\Big) \\ &\qquad\qquad + P(\hat{\sigma}^2_{min}(\mathbf{X}) \leq c_{min} - \epsilon') \\ &\overset{(c)}{\leq} Ke^{-\kappa n\epsilon'^2}.\end{aligned} \tag{15}$$

In the above, step (a) follows by (13), step (b) since if

$$\Big\{\frac{\mathcal{C}(\widehat{\boldsymbol{\beta}})}{p} \leq \mathsf{C}\Big\} \cap \Big\{\frac{1}{p}(\sigma^2_{max}(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2) \leq \mathsf{C}\Big\} \cap \Big\{\hat{\sigma}^2_{min}(\mathbf{X}) \geq c_{min} - \epsilon'\Big\},$$

then

$$\begin{aligned}&\tilde{\mathsf{C}}\Big(\frac{\mathcal{C}(\widehat{\boldsymbol{\beta}})}{p}\Big)^2 + \frac{\tilde{\mathsf{C}}(\mathcal{C}(\widehat{\boldsymbol{\beta}}) + \sigma^2_{max}(\mathbf{X})\|\boldsymbol{\beta}\|^2 + \|\mathbf{w}\|^2)}{p\hat{\sigma}^2_{min}(\mathbf{X})} \\ &\qquad\qquad \leq \tilde{\mathsf{C}}\mathsf{C}^2 + \frac{2\tilde{\mathsf{C}}\mathsf{C}}{c_{min} - \epsilon'},\end{aligned} \tag{16}$$

and step (c) from (7), (8), and Lemma 5. Finally, labeling $\hat{\mathsf{B}} := \tilde{\mathsf{C}}\mathsf{C}^2 + 4\tilde{\mathsf{C}}\mathsf{C}/c_{min} \geq \tilde{\mathsf{C}}\mathsf{C}^2 + 2\tilde{\mathsf{C}}\mathsf{C}/(c_{min}/2)$, it follows from (14) -(15) that,

$$\begin{aligned}P\Big(\frac{\|\widehat{\boldsymbol{\beta}}\|^2}{p} \geq \hat{\mathsf{B}}\Big) &\leq P\Big(\frac{\|\widehat{\boldsymbol{\beta}}\|^2}{p} \geq \tilde{\mathsf{C}}\mathsf{C}^2 + \frac{2\tilde{\mathsf{C}}\mathsf{C}}{c_{min} - c_{min}/2}\Big) \\ &\leq Ke^{-\kappa n c^2_{min}/4}.\end{aligned}$$

□

References

[Kashin, 1977] Kashin, B. S. (1977). Diameters of some finite-dimensional sets and classes of smooth functions. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, 41(2):334–351.

[Rudelson and Vershynin, 2010] Rudelson, M. and Vershynin, R. (2010). Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 1576–1602. World Scientific.

[Rush and Venkataramanan, 2015] Rush, C. and Venkataramanan, R. (2015). Finite sample analysis of approximate message passing. *Proc. IEEE Int. Symp. Inf. Theory*. Full version: https://arxiv.org/abs/1606.01800.