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 \( P(X_n^2 - c^2 \geq \epsilon) \leq Ke^{-\epsilon n c^2} \), then \( P(|X_n| - |c| \geq \epsilon \leq Ke^{-\epsilon n |c|^2} \).

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

\[
P\left( \frac{1}{n} \sum_{i=1}^{n} Z_i^2 - 1 \geq \epsilon \right) \leq 2 \exp \left( -\frac{n \epsilon^2}{8} \right).
\]

Lemma 3. Let \( \ker(X) \) denote the kernel of a matrix \( X \in \mathbb{R}^{n \times p} \) and \( v_1, v_2 \in \mathbb{R}^p \) vectors with \( v_i \in \ker(X)^\perp \). Then \( \sigma_{\min}^2 \|v_2\|^2 \leq \|Xv_2\|^2 \leq \sigma_{\max}^2 \|v_2\|^2 \) where \( \sigma_{\min}(X) \) and \( \sigma_{\max}(X) \) are the minimum and maximum singular values of \( X \) and \( \sigma_{\min}^2 \|v_1\|^2 \leq \|Xv_1\|^2 \leq \sigma_{\max}^2 \|v_1\|^2 \), where \( \sigma_{\min}(X) \) is the minimum, non-zero singular value of \( X \).

Lemma 4 (Kashin, 1977). For any dimension \( p > 0 \), with probability at least \( 1 - 2^{-p} \), for any uniformly random subspace \( V_{(p,r)} \) with dimension \( |p(1-r)| \) where \( 0 \leq r \leq 1 \) is a positive constant, for all vectors \( v \in V_{(p,r)} \),

\[
c_r \|v\| \leq \|v\|/\sqrt{p},
\]

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

Proof. The result for \( \tilde{X} \) follows directly from the cited result. The result for \( X \) uses the following fact: let \( M_1 \) be an \( m \times n \) matrix and let \( M_2 \) be an \( n \times m \) matrix, with \( n \geq m \), then the \( n \) eigenvalues of \( M_2 M_1 \) are the \( m \) eigenvalues of \( M_1 M_2 \) with the extra eigenvalues being 0. So we note that if \( X \) is an \( n \times p \) matrix with \( p \geq n \) then the \( n \) eigenvalues of \( X^* X \) are the \( n \) eigenvalues of \( XX^* \) with the extra eigenvalues being 0. The cited result tells us \( P(\sigma_{\max}(X^*) - \|B\| - n\epsilon) \geq 1 + \sqrt{\delta} - \epsilon \) and \( P(\sigma_{\min}(X^*) - 1 - \sqrt{\delta} - \epsilon) \) are upper bounded by \( e^{-n\epsilon^2/2} \).

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

\[
P(\|B - \tau_{t-1} Z_i\| > \theta_t) - P(\|B - \tau_{s} Z\| \geq \theta_s) \leq \frac{2\mathbb{E}[B]}{\min\{\tau_0^2, \tau_s^2\}} |\tau_t - \tau_s|.
\]

Proof. We first note that

\[
P(\|B - \tau_s Z\| \geq \theta_s | B = b) = P(\|b - \tau_s Z\| \geq \theta_s).
\]

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

\[
P(\|b - \tau_s Z\| \geq \theta_s) = P(b - \tau_s Z \geq \theta_s) + P(b - \tau_s Z \leq -\theta_s) = P(Z \leq (b/\tau_s) - \alpha) + P(Z \geq (b/\tau_s) + \alpha) = \Phi((-b/\tau_s) - \alpha) + \Phi((b/\tau_s) - \alpha).
\]
Similarly, 
\[ P(|x - \tau Z| \geq \theta) = \Phi(\frac{x}{\tau} - \alpha) + \Phi(-\frac{x}{\tau} - \alpha), \]
and by Jensen’s inequality,
\[
\begin{align*}
\left| P(|B - \tau Z| > \theta) - P(|B - \tau Z| \geq \theta) \right| \\
= \mathbb{E}\left\{ P(|B - \tau Z| > \theta | B) - P(|B - \tau Z| \geq \theta | B) \right\} \\
\leq \mathbb{E}\left\{ P(|B - \tau Z| > \theta | B) - P(|B - \tau Z| \geq \theta) \right\}.
\end{align*}
\]
(1)

Then, letting \( h(b) := \left| P(|B - \tau Z| > \theta | B = b) - P(|B - \tau Z| \geq \theta | B = b) \right| \), we have
\[
h(b) = \mathbb{E}\left\{ \mathbb{I}\{|b - \tau Z| > \theta\} - \mathbb{I}\{|b - \tau Z| \geq \theta\} \right\} \\
= \mathbb{E}\left\{ \Phi(b/\tau) - \Phi((b/\tau) - \alpha) \right\} + \mathbb{E}\left\{ \Phi(-b/\tau) - \Phi(-b/\tau - \alpha) \right\}.
\]
(2)

Now we note that the Gaussian cdf is Lipschitz. Indeed, for \( x > y \), and \( \phi(\cdot) \) the Gaussian pdf,
\[
\mathbb{E}\left\{ \mathbb{I}\{|b - \tau Z| > \theta\} - \mathbb{I}\{|b - \tau Z| \geq \theta\} \right\} \\
= \mathbb{E}\left\{ \Phi(b/\tau) - \Phi((b/\tau) - \alpha) \right\} + \mathbb{E}\left\{ \Phi(-b/\tau) - \Phi(-b/\tau - \alpha) \right\}.
\]

Proof. Note that the concentration for the norm of \( \beta^t \) follows from Theorem 1 in the main document using the pseudo-Lipschitz function \( \phi(a, b) = a^2 \). Namely, we know that
\[
P\left( \frac{1}{p} \|eta^t\|^2 - \mathbb{E}\left[ (\eta(B + \tau Z; \theta_i))^2 \right] \geq \epsilon \right) \leq K e^{-\kappa \epsilon n c^2},
\]
where, for each \( 0 \leq t \leq T^* \),
\[
\mathbb{E}\left[ (\eta(B + \tau Z; \theta_i))^2 \right] \\
\leq \mathbb{E}\left[ (B + \tau Z - \theta_i)^2 \mathbb{I}\{B + \tau Z > \theta_i\} \right] \\
+ \mathbb{E}\left[ (B + \tau Z + \theta_i)^2 \mathbb{I}\{B + \tau Z < -\theta_i\} \right]
\]
\[
\leq 6\mathbb{E}\{B^2\} + 6\tau^2\eta^2\mathbb{E}\{Z^2\} + 6\theta^2
\]
\[
\leq 6\sigma_0^2 + 6\tau^2 + 6\epsilon^2 \leq 6\sigma_0^2 + 6\max\{\tau_0^2, \tau^2\}(1 + \alpha^2).
\]
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 \( \tau_0^2 \leq \max\{\tau_0^2, \tau^2\} \). The result in the lemma statement follows since
\[
P\left( \frac{1}{p} \|eta^t\|^2 / p \geq \hat{B} \right) \\
\leq P\left( \frac{1}{p} \|eta^t\|^2 / p \geq \mathbb{E}[\eta(B + \tau Z; \theta_i)^2] + \epsilon \right) \\
\leq P\left( \frac{1}{p} \|eta^t\|^2 / p - \mathbb{E}[\eta(B + \tau Z; \theta_i)^2] \geq \epsilon \right).
\]
The first inequality above follows since \( \hat{B} \geq 1 + 6\sigma_0^2 + 6\max\{\tau_0^2, \tau^2\}(1 + \alpha^2) \geq \epsilon + \mathbb{E}[\eta(B + \tau Z; \theta_i)^2] \).

Now consider concentration for \( \hat{\beta} \). We will first show that \( C(\hat{\beta}) \) is lower bounded by a constant with high probability. We have the following upper bound on the LASSO cost function:
\[
C(\hat{\beta}) \leq C(0) = \frac{1}{2} |y|^2 = \frac{1}{2} \|X\beta + w\|^2
\]
(5)
\[
\leq \frac{1}{2} \|X\beta\|^2 + \|w\|^2 \leq \sigma_{\max}^2(X)\|\beta\|^2 + \|w\|^2,
\]
(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 \( C := (c_{\max} + 1)(\sigma_0^2 + 1) + \sigma^2 + 1 \). Considering the above, then
\[
P(\sigma_{\max}^2(X)\|\beta\|^2 + \|w\|^2 / p \geq C) \\
\leq P(\sigma_{\max}^2(X) \geq c_{\max} + \epsilon') + P(\|\beta\|^2 / p \geq \sigma_0^2 + \epsilon') \\
+ P(\|w\|^2 / p \geq \sigma^2 + \epsilon') \\
\leq K e^{-\kappa \epsilon \sigma^2},
\]
(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),

\[
P(C(\hat{\beta})/p ≥ C) ≤ P((σ^2_{\text{max}}(X)||\beta||^2 + ||w||^2)/p ≥ C) ≤ Ke^{-κν^2}. (8)
\]

Now we will relate \(∥\hat{\beta}∥^2/p\) to \(C(\hat{\beta})/p\) and other terms lower-bounded by a constant with high probability. We write \(\hat{\beta} = \hat{β}^\perp + \hat{β}^∥\) where \(\hat{β}^∥ \in \ker(X)^⊥\) and \(\hat{β}^\perp \in \ker(X)\). Since \(\beta^∥ \in \ker(X)\) and \(\ker(X)\) is a random subspace of size \(p - n = p(1 - δ)\), by Kashin Theorem (Lemma 4), we have that for some constant \(ν_1 = ν_1(δ)\),

\[
P(∥\hat{β}∥^2 ≥ ν_1||β||^2_p/N) ≤ Ke^{-p}. (9)
\]

Denote the event \(\{∥\hat{β}∥^2 ≥ ν_1||β||^2_p/p\}\) by \(E\), and by the above \(P(E) ≤ Ke^{-p}\). Conditioned on \(E^c\), we have the following bound

\[
\hat{β} = \hat{β}^\perp + \hat{β}^∥ \leq \frac{2ν_1}{p} ||\beta||_1^2 + \frac{2ν_1}{p} ||\beta||_1^2 + ||\hat{β}^∥||^2 \leq 2(2ν_1 + 1)||\hat{β}^∥||^2, (10)
\]

where step (a) holds since we are conditioning on \(E^c\), step (b) from the triangle inequality and Cauchy-Schwarz, and step (c) since \(λ∥\hat{β}∥_1 ≤ C(\hat{β})\) by the definition of the cost function and \(∥\hat{β}^∥_1 ≤ √p||\hat{β}^∥\) by Cauchy-Schwarz. Now we bound the second term on the RHS of (10):

\[
∥\hat{β}^∥^2 ≤ \frac{2∥X\hat{β}∥^2}{σ^2_{\text{min}}(X)} ≤ \frac{2∥X\hat{β}^∥^2 - y^2}{σ^2_{\text{min}}(X)} \leq \frac{\hat{β}^∥^2 - y^2 + y^2}{σ^2_{\text{min}}(X)} \leq \frac{\hat{β}^∥^2}{σ^2_{\text{min}}(X)} + \frac{2∥y∥^2}{σ^2_{\text{min}}(X)} \leq \frac{2\hat{β}^∥^2}{σ^2_{\text{min}}(X)} + \frac{2∥X\beta + w∥^2}{σ^2_{\text{min}}(X)}. (11)
\]

In the above, step (d) follows from the fact that \(σ^2_{\text{min}}(X)||\hat{β}^∥||^2 ≤ ||X\hat{β}^∥||^2\) by Lemma 3 and step (e) by Cauchy-Schwarz. Next note

\[
\|X\beta + w\|^2 ≤ 2(\|X\beta\|^2 + \|w\|^2) ≤ 2(σ^2_{\text{max}}(X)||\beta||^2 + \|w\|^2), (12)
\]

by Cauchy-Schwarz and Lemma 3. Now plugging (11) and (12) into (10) we have for some constant \(C = \max\{\frac{2σ^2_{\text{min}}(X)}{p}, 4(2ν_1 + 1)\} > 0\),

\[
\frac{1}{p} ||\hat{β}||^2 ≤ \hat{C}(\frac{1}{p} ||\hat{β}||^2 + \frac{\hat{C}(\hat{β}) + σ^2_{\text{max}}(X)||\beta||^2 + \|w\|^2}{pσ^2_{\text{min}}(X)}), (13)
\]

Now considering the above,

\[
P\left(\frac{1}{p} ||\hat{β}||^2 ≥ \hat{C}^2 + \frac{2\hat{C}C}{c_{\text{min}} - ε}\right) ≥ P(E) + P\left(\frac{1}{p} ||\hat{β}||^2 ≥ \hat{C}^2 + \frac{2\hat{C}C}{c_{\text{min}} - ε}\right).
\]

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

\[
P\left(\frac{1}{p} ||\hat{β}||^2 ≥ \hat{C}^2 + \frac{2\hat{C}C}{c_{\text{min}} - ε}\right) ≤ P\left(\hat{C}(\frac{1}{p} ||\hat{β}||^2 + \frac{σ^2_{\text{max}}(X)||\beta||^2 + \|w\|^2}{Nσ^2_{\text{min}}(X)} ≥ \hat{C}^2 + \frac{2\hat{C}C}{c_{\text{min}} - ε}\right)
\]

\[
+ P\left(σ^2_{\text{min}}(X) ≤ c_{\text{min}} - ε\right)
\]

\[
(15)
\]

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

\[
\text{\{C(\hat{β}) ≤ C\} \cap \left\{\frac{1}{p} \left(σ^2_{\text{max}}(X)||\beta||^2 + \|w\|^2\right) ≤ C\right\} \cap \left\{σ^2_{\text{min}}(X) ≥ c_{\text{min}} - ε\right\},
\]

then

\[
\hat{C}(\frac{1}{p} ||\hat{β}||^2 + \frac{\hat{C}(\hat{β}) + σ^2_{\text{max}}(X)||\beta||^2 + \|w\|^2}{pσ^2_{\text{min}}(X)}) ≤ \hat{C}^2 + \frac{2\hat{C}C}{c_{\text{min}} - ε}.
\]

and step (c) from (7), (8), and Lemma 3. Finally, labeling \(B := \hat{C}^2 + 4\hat{C}C/c_{\text{min}} ≥ \hat{C}^2 + 2\hat{C}/(c_{\text{min}}/2)\), it follows from (14) and (15) that,

\[
P\left(\frac{1}{p} ||\hat{β}||^2 ≥ B\right) ≤ P\left(\frac{1}{p} ||\hat{β}||^2 ≥ C\right) + \frac{2\hat{C}C}{c_{\text{min}} - c_{\text{min}}/2} \leq Ke^{-κν^2_{\text{min}}/4}.
\]

References
