Supplemental Materials

Proof to Lemma 1

Proof. Denote $\text{supp}(\text{P}_{\Omega(\infty,\mathbf{t})}(\mathbf{w}))$ and $\text{supp}(\text{P}_{\Omega(s,\mathbf{t})}(\mathbf{w}))$ by A and B respectively for short. We first prove $B \subseteq A$.

Suppose $B \nsubseteq A$, then we can find an element $b \in B$ but $b \notin A$. Without the loss of generality, we assume that b is in a certain group g. Since $A \cap g$ contains the indices of the $\mathbf{t}_g$ largest (magnitude) elements of group g, there exists at least one element $a \in A \cap g$ and $a \notin B \cap g$ (otherwise $|B \cap g| \geq \mathbf{t}_g + 1$). Replacing b by a in B, the constraints are still satisfied, but we can get a better solution since $|\mathbf{w}_a| > |\mathbf{w}_b|$. This contradicts $B = \text{supp}(\text{P}_{\Omega(s,\mathbf{t})}(\mathbf{w}))$.

Because we already know $B \subseteq A$, we can construct B by selecting the A's elements corresponding to the largest s (magnitude) elements. Therefore, $\text{supp}(\text{P}_{\Omega(s,\mathbf{t})}(\mathbf{w})) = \text{supp}(\text{P}_{\Omega(s,\infty)}(\text{P}_{\Omega(\infty,\mathbf{t})}(\mathbf{w})))$, which proves Lemma 1. □

Lemma 5. *$\forall$supp$(\mathbf{w} - \bar{\mathbf{w}}) \subseteq S, S \in \Omega(s,\mathbf{t})$, if $2\eta - \eta^2\rho_+(s,\mathbf{t})) > 0$, then*

$$\|\mathbf{w} - \bar{\mathbf{w}} - \eta[\nabla f(\mathbf{w}) - \nabla f(\bar{\mathbf{w}})]_S\|^2 \leq (1 - 2\eta\rho_-(s,\mathbf{t}) + \eta^2\rho_-(s,\mathbf{t})\rho_+(s,\mathbf{t}))\|\mathbf{w} - \bar{\mathbf{w}}\|^2. \tag{6}$$

Proof.

$$\begin{aligned}
&\|\mathbf{w} - \bar{\mathbf{w}} - \eta[\nabla f(\mathbf{w}) - \nabla f(\bar{\mathbf{w}})]_S\|^2 \\
=&\|\mathbf{w} - \bar{\mathbf{w}}\|^2 + \eta^2\|[\nabla f(\mathbf{w}) - \nabla f(\bar{\mathbf{w}})]_S\|^2 - 2\eta\langle \mathbf{w} - \bar{\mathbf{w}}, [\nabla f(\mathbf{w}) - \nabla f(\bar{\mathbf{w}})]_S\rangle \\
\leq&\|\mathbf{w} - \bar{\mathbf{w}}\|^2 + (\eta^2\rho_+(s,\mathbf{t}) - 2\eta)\langle \mathbf{w} - \bar{\mathbf{w}}, [\nabla f(\mathbf{w}) - \nabla f(\bar{\mathbf{w}})]_S\rangle \\
\leq&\|\mathbf{w} - \bar{\mathbf{w}}\|^2 - (2\eta - \eta^2\rho_+(s,\mathbf{t}))\rho_-(s,\mathbf{t})\|\mathbf{w} - \bar{\mathbf{w}}\|^2 \\
=&(1 - 2\eta\rho_-(s,\mathbf{t}) + \eta^2\rho_+(s,\mathbf{t})\rho_-(s,\mathbf{t}))\|\mathbf{w} - \bar{\mathbf{w}}\|^2.
\end{aligned}$$

It completes the proof. □

Proof to Theorem 2

Proof. Let us prove the first claim.

$$\begin{aligned}
&\|\mathbf{w}^{k+1} - (\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k))\|^2 \\
=&\|\mathbf{w}^{k+1} - \bar{\mathbf{w}}\|^2 + \|\bar{\mathbf{w}} - (\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k))\|^2 + 2\langle \mathbf{w}^{k+1} - \bar{\mathbf{w}}, \bar{\mathbf{w}} - (\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k))\rangle ?
\end{aligned}$$

Define $\bar{\Omega} = \text{supp}(\bar{\mathbf{w}}), \Omega_{k+1} = \text{supp}(\mathbf{w}^{k+1})$, and $\bar{\Omega}_{k+1} = \bar{\Omega} \cup \Omega_{k+1}$. From $\|\mathbf{w}^{k+1} - (\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k))\|^2 \leq \|\bar{\mathbf{w}} - (\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k))\|^2$, we have

$$\begin{aligned}
\|\mathbf{w}^{k+1} - \bar{\mathbf{w}}\|^2 &\leq 2\langle \mathbf{w}^{k+1} - \bar{\mathbf{w}}, \mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}}\rangle \\
&= 2\langle \mathbf{w}^{k+1} - \bar{\mathbf{w}}, [\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}}]_{\bar{\Omega}_{k+1}}\rangle \\
&\leq 2\|\mathbf{w}^{k+1} - \bar{\mathbf{w}}\|\|[\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}}]_{\bar{\Omega}_{k+1}}\|.
\end{aligned}$$

It follows

$$\begin{aligned}
\|\mathbf{w}^{k+1} - \bar{\mathbf{w}}\| &\leq 2\|[\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}}]_{\bar{\Omega}_{k+1}}\| \\
&= 2\|[\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}} + \eta\nabla f(\bar{\mathbf{w}}) - \eta\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}}\| \\
&\leq 2\|[\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}} + \eta\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}}\| + 2\eta\|[\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}}\| \\
&\leq 2\|[\mathbf{w}^k - \eta\nabla f(\mathbf{w}^k) - \bar{\mathbf{w}} + \eta\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}\cup\Omega_k}\| + 2\eta\|[\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}}\| \\
&= 2\|\mathbf{w}^k - \bar{\mathbf{w}} - \eta[\nabla f(\mathbf{w}^k) - \nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}\cup\Omega_k}\| + 2\eta\|[\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}}\|.
\end{aligned}$$

From the inequality of Lemma 5, we have

$$\begin{aligned}\|\mathbf{w}^{k+1}-\bar{\mathbf{w}}\| &\leq \alpha\|\mathbf{w}^k-\bar{\mathbf{w}}\|+2\eta\|[\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{k+1}}\| \\ &\leq \alpha\|\mathbf{w}^k-\bar{\mathbf{w}}\|+2\eta\max_j\|[\nabla f(\bar{\mathbf{w}})]_{\bar{\Omega}_{j+1}}\| \\ &\leq \alpha\|\mathbf{w}^k-\bar{\mathbf{w}}\|+2\eta\Delta. \end{aligned} \tag{7}$$

Since Δ is constant, using the recursive relation of (7), we have

$$\begin{aligned}\|\mathbf{w}^k-\bar{\mathbf{w}}\| &\leq \alpha^k\|\mathbf{w}^0-\bar{\mathbf{w}}\|+2\eta\Delta\sum_{i=0}^{k}\alpha^i \\ &= \alpha^k\|\mathbf{w}^0-\bar{\mathbf{w}}\|+2\eta\Delta\frac{1-\alpha^k}{1-\alpha} \\ &\leq \alpha^k\|\mathbf{w}^0-\bar{\mathbf{w}}\|+2\eta\Delta\frac{1}{1-\alpha}. \end{aligned} \tag{8}$$

Then we move to (2), when $k \geq \lceil \log \frac{2\Delta}{(1-\alpha)\rho_+(3s,3t)\|\mathbf{w}^0-\bar{\mathbf{w}}\|}/\log\alpha\rceil$, from the conclusion of (1), we have

$$\|\mathbf{w}^k-\bar{\mathbf{w}}\|_\infty \leq \|\mathbf{w}^k-\bar{\mathbf{w}}\| \leq \frac{4\Delta}{(1-\alpha)\rho_+(3s,3t)}. \tag{9}$$

For any $j\in\bar{\Omega}$,

$$\begin{aligned}\|\mathbf{w}^k-\bar{\mathbf{w}}\|_\infty &\geq |[\mathbf{w}^k-\bar{\mathbf{w}}]_j| \\ &\geq -|[\mathbf{w}^k]_j|+|[\bar{\mathbf{w}}]_j|.\end{aligned}$$

So

$$\begin{aligned}|[\mathbf{w}^k]_j| &\geq |[\bar{\mathbf{w}}]_j|-\|\mathbf{w}^k-\bar{\mathbf{w}}\|_\infty \\ &\geq |[\bar{\mathbf{w}}]_j|-\frac{4\Delta}{(1-\alpha)\rho_+(3s,3t)}.\end{aligned}$$

Therefore, $[\mathbf{w}^k]_j$ is non-zero if $|[\bar{\mathbf{w}}]_j| > \frac{4\Delta}{(1-\alpha)\rho_+(3s,3t)}$, and (2) is proved. □

Lemma 6. *The value of Δ is bounded by*

$$\Delta \leq \min\left(O\left(\sqrt{\frac{s\log p+\log 1/\eta'}{n}}\right),\ O\left(\sqrt{\frac{\max_{g\in\mathcal{G}}\log|g|\sum_{g\in\mathcal{G}}\mathbf{t}_g+\log 1/\eta'}{n}}\right)\right), \tag{10}$$

with high probability $1-\eta'$.

Proof. We introduce the following notation for matrix and it is different from the vector notation. For a matrix X in $\mathbb{R}^{n\times p}$, X_h will be a $\mathbb{R}^{n\times|h|}$ matrix that only keep the columns corresponding to the index set h. Here we restrict h by $\mathbf{w}_h \in \Omega(s,\mathbf{t})$ for any $\mathbf{w}\in\mathbb{R}^p$. We denote $\Sigma_h = X_h^\top X_g$, For the theorem, we can first show that $\|X_h^\top\epsilon\| \leqslant \sqrt{n}\left(\sqrt{|h|}+\sqrt{2\rho_+(2s,2\mathbf{t})\log(\frac{1}{\eta})}\right)$ with probability $1-\eta$. To this end, we have to point out that our columns of X are normalized to $\sqrt{n}$ and hence $X_h^\top\epsilon$ will be a $\frac{p}{m}$-variate Gaussian random variable with n on the diagonal of covariance matrix. We further use λ_i as the eigenvalues of Σ_h with decreasing order, i.e., λ_1 being the largest, or equivalently, $\lambda_1 = \|\Sigma_h\|_{spec}$.

Also, using the trick that $tr(\Sigma_h^2) = \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{|h|}^2$ and Proposition 1.1 from [16], we have

$$\begin{aligned}
e^{-t} \geqslant & \mathbf{Pr}\left(||X_h^\top \epsilon||^2 > \sum_{i=1}^{|h|} \lambda_i + 2\sqrt{\sum_{i=1}^{|h|} \lambda_i^2 t} + 2\lambda_1 t \right) \\
\geqslant & \mathbf{Pr}\left(||X_h^\top \epsilon||^2 > \sum_{i=1}^{|h|} \lambda_i + 2\sqrt{2\sum_{i=1}^{|h|} \lambda_i \lambda_1 t + 2\lambda_1 t} \right) \\
\geqslant & \mathbf{Pr}\left(||X_h^\top \epsilon|| \geqslant \sqrt{\sum_{i=1}^{|h|} \lambda_i} + \sqrt{2\lambda_1 t} \right).
\end{aligned}$$

Substitute t with $\log(\frac{1}{\eta})$ and the facts that $\sum_{i=1}^{|h|} \lambda_i = |h|n$ and $\lambda_1 = \|\Sigma\|_{spec} \leqslant n\rho_+(2s, 2\mathbf{t})$, we have

$$\|X_h^\top \epsilon\| \leqslant \sqrt{n}\left(\sqrt{|h|} + \sqrt{2\rho_+(2s, 2\mathbf{t})\log(1/\eta)}\right)$$

with probability $1 - \eta$.

For the least square loss, we have $\nabla f(\bar{\mathbf{w}}) = \frac{1}{n} X^\top (X\bar{\mathbf{w}} - y) = \frac{1}{n} X^\top \epsilon$. To estimate the upper bound of $\|\mathrm{P}_{\Omega(2s,2\mathbf{t})}(\nabla f(\mathbf{w}))\|$, we use the following fact

$$\|\mathrm{P}_{\Omega(2s,2\mathbf{t})}(\nabla f(\bar{\mathbf{w}}))\| = \|\mathrm{P}_{\Omega(2s,2\mathbf{t})}(X^\top \epsilon)\| \leq \min\left(\|\mathrm{P}_{\Omega(2s,\infty)}(X^\top \epsilon)\|,\ \|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top \epsilon)\|\right).$$

We consider the upper bounds of $\|\mathrm{P}_{\Omega(2s,\infty)}(X^\top \epsilon)\|$ and $\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top \epsilon)\|$ respectively:

$$\begin{aligned}
& \mathbf{Pr}\left(\|\mathrm{P}_{\Omega(2s,\infty)}(X^\top \epsilon)\| \geq n^{-1/2}\left(\sqrt{2s} + \sqrt{2\rho_+(2s, 2\mathbf{t})\log(1/\eta)}\right)\right) \\
= & \mathbf{Pr}\left(\max_{|h|=2s} \|X_h^\top \epsilon\| \geq n^{-1/2}\left(\sqrt{2s} + \sqrt{2\rho_+(2s, 2\mathbf{t})\log(1/\eta)}\right)\right) \\
\leq & \sum_{|h|=2s} \mathbf{Pr}\left(\|X_h^\top \epsilon\| \geq n^{-1/2}\left(\sqrt{2s} + \sqrt{2\rho_+(2s, 2\mathbf{t})\log(1/\eta)}\right)\right) \\
\leq & \binom{p}{2s}\eta.
\end{aligned}$$

By taking $\eta' = \eta \binom{p}{2s}$, we obtain

$$\begin{aligned}
\eta' \geq & \mathbf{Pr}\left(\|\mathrm{P}_{\Omega(2s,\infty)}(X^\top \epsilon)\| \geq n^{-1/2}\left(\sqrt{2s} + \sqrt{2\rho_+(2s, 2\mathbf{t})\log\left(\binom{p}{2s}/\eta'\right)}\right)\right) \\
\geq & \mathbf{Pr}\left(\|\mathrm{P}_{\Omega(2s,\infty)}(X^\top \epsilon)\| \geq O\left(\sqrt{\frac{s\log(p) + \log 1/\eta'}{n}}\right)\right),
\end{aligned}$$

where the last inequality uses the fact that $\rho_+(2s, 2\mathbf{t})$ is bounded by a constant with high probability.

Next we consider the upper bound of $\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top\epsilon)\|$. Similarly, we have

$$\begin{aligned}
&\mathbf{Pr}\left(\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top\epsilon)\|\geq n^{-1/2}\left(\sqrt{2\sum_{g\in\mathcal{G}} g}+\sqrt{2\rho_+(2s,2\mathbf{t})\log(1/\eta)}\right)\right)\\
=&\mathbf{Pr}\left(\max_{|h\cap g|=2\mathbf{t}_g\ \forall g\in\mathcal{G}}\|X_h^\top\epsilon\|\geq n^{-1/2}\left(\sqrt{2\sum_{g\in\mathcal{G}} g}+\sqrt{2\rho_+(2s,2\mathbf{t})\log(1/\eta)}\right)\right)\\
\leq&\sum_{|h\cap g|\leq 2\mathbf{t}_g\ \forall g\in\mathcal{G}}\mathbf{Pr}\left(\|X_h^\top\epsilon\|\geq n^{-1/2}\left(\sqrt{2\sum_{g\in\mathcal{G}} g}+\sqrt{2\rho_+(2s,2\mathbf{t})\log(1/\eta)}\right)\right)\\
\leq&\eta\prod_{g\in\mathcal{G}}\binom{|g|}{2\mathbf{t}_g}.
\end{aligned}$$

Thus, by taking $\eta'=\eta\prod_{g\in\mathcal{G}}\binom{|g|}{2\mathbf{t}_g}$, we have

$$\begin{aligned}
\eta'\geq&\mathbf{Pr}\left(\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top\epsilon)\|\geq n^{-1/2}\left(\sqrt{2\sum_{g\in\mathcal{G}}\mathbf{t}_g}+\sqrt{2\rho_+(2s,2\mathbf{t})\log\left(\prod_{g\in\mathcal{G}}\binom{|g|}{2\mathbf{t}_g}/\eta'\right)}\right)\right)\\
\geq&\mathbf{Pr}\left(\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top\epsilon)\|\geq n^{-1/2}\left(\sqrt{2\sum_{g\in\mathcal{G}}\mathbf{t}_g}+\sqrt{4\rho_+(2s,2\mathbf{t})\sum_{g\in\mathcal{G}}\mathbf{t}_g\log|g|+2\varphi_+(1)\log 1/\eta'}\right)\right)\\
\geq&\mathbf{Pr}\left(\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top\epsilon)\|\geq n^{-1/2}\left(\sqrt{2\sum_{g\in\mathcal{G}}\mathbf{t}_g}+\sqrt{4\rho_+(2s,2\mathbf{t})\max_{g\in\mathcal{G}}\log|g|\sum_{g\in\mathcal{G}}\mathbf{t}_g+2\varphi_+(1)\log 1/\eta'}\right)\right)\\
\geq&\mathbf{Pr}\left(\|\mathrm{P}_{\Omega(\infty,2\mathbf{t})}(X^\top\epsilon)\|\geq O\left(\sqrt{\frac{\max_{g\in\mathcal{G}}\log|g|\sum_{g\in\mathcal{G}}\mathbf{t}_g+\log 1/\eta'}{n}}\right)\right).
\end{aligned}$$

Summarizing two upper bounds, we have with high probability $(1-2\eta')$

$$\|\mathrm{P}_{\Omega(2s,2\mathbf{t})}(\nabla f(\bar{\mathbf{w}}))\|\leq\min\left(O\left(\sqrt{\frac{s\log p+\log 1/\eta'}{n}}\right),\ O\left(\sqrt{\frac{\max_{g\in\mathcal{G}}\log|g|\sum_{g\in\mathcal{G}}\mathbf{t}_g+\log 1/\eta'}{n}}\right)\right).$$

□

Lemma 7. *For the least square loss, assume that matrix X to be sub-Gaussian with zero mean and has independent rows or columns. If the number of samples n is more than*

$$O\left(\min\left\{s\log p,\ \log(\max_{g\in\mathcal{G}}|g|)\sum_{g\in\mathcal{G}}\mathbf{t}_g\right\}\right),$$

then with high probability, we have with high probability

$$\rho_+(3s,3\mathbf{t})\leq\frac{3}{2} \tag{11}$$

$$\rho_-(3s,3\mathbf{t})\geq\frac{1}{2}. \tag{12}$$

Thus, α defined in (3) *is less than* 1 *by appropriately choosing η (for example, $\eta=1/\rho_+(3s,3\mathbf{t})$).*

Proof. For the linear regression loss, we have

$$\rho_+^{1/2}(3s, 3\mathbf{t}) \le \frac{1}{\sqrt{n}} \max_{\mathbf{w} \in \Omega(3s,3\mathbf{t})} \frac{\|X\mathbf{w}\|}{\|\mathbf{w}\|} = \max_{|h| \le 3s, |h \cap g| \le \mathbf{t}_g} \|X_h\|$$
$$\rho_-^{1/2}(3s, 3\mathbf{t}) \ge \frac{1}{\sqrt{n}} \min_{\mathbf{w} \in \Omega(3s,3\mathbf{t})} \frac{\|X\mathbf{w}\|}{\|\mathbf{w}\|} = \min_{1 \le |h| \le 3s, |h \cap g| \le \mathbf{t}_g} \|X_h\|$$

From the random matrix theory [35, Theorem 5.39], we have

$$\mathbf{Pr}\left(\|X_h\| \ge \sqrt{n} + O(\sqrt{3s}) + O(\sqrt{\log \frac{1}{\eta}}\right) \le O(\eta)$$

Then we have

$$\begin{aligned}
&\mathbf{Pr}\left(\sqrt{n}\rho_+^{1/2}(3s, 3\mathbf{t}) \ge \sqrt{n} + O(\sqrt{s}) + O(\sqrt{\log \frac{1}{\eta}})\right) \\
\le &\mathbf{Pr}\left(\max_{|h| \le 3s, |h \cap g| \le \mathbf{t}_g} \|X_h\| \ge \sqrt{n} + O(\sqrt{s}) + O(\sqrt{\log \frac{1}{\eta}})\right) \\
\le &|\{h \mid |h| = 3s\}| \mathbf{Pr}\left(\|X_h\| \ge \sqrt{n} + O(\sqrt{s}) + O(\sqrt{\log \frac{1}{\eta}})\right) \\
= &\binom{p}{3s} \mathbf{Pr}\left(\|X_h\| \ge \sqrt{n} + O(\sqrt{s}) + O(\sqrt{\log \frac{1}{\eta}})\right) \le O\left(\binom{p}{3s}\eta\right)
\end{aligned}$$

which implies (by taking $\eta' = \binom{p}{3s}\eta$):

$$\mathbf{Pr}\left(\sqrt{n}\rho_+^{1/2}(3s, 3\mathbf{t}) \ge \sqrt{n} + O\left(\sqrt{s \log p}\right) + O\left(\sqrt{\log \frac{1}{\eta'}}\right)\right) \le \eta'$$

Taking $n = O(s \log p)$, we have $\rho_+^{1/2}(3s, 3\mathbf{t}) \le \sqrt{\frac{3}{2}}$ with high probability. Next, we consider it from a different perspective.

$$\begin{aligned}
&\mathbf{Pr}\left(\sqrt{n}\rho_+^{1/2}(3s, 3\mathbf{t}) \ge \sqrt{n} + O\left(\sqrt{\sum_{g \in \mathcal{G}} \mathbf{t}_g}\right) + O\left(\sqrt{\log \frac{1}{\eta}}\right)\right) \\
\le &\mathbf{Pr}\left(\sqrt{n}\rho_+^{1/2}(+\infty, 3\mathbf{t}) \ge \sqrt{n} + O\left(\sqrt{\sum_{g \in \mathcal{G}} \mathbf{t}_g}\right) + O\left(\sqrt{\log \frac{1}{\eta}}\right)\right) \\
= &\mathbf{Pr}\left(\max_{|h \cap g| \le \mathbf{t}_g, g \in \mathcal{G}} \|X_h\| \ge \sqrt{n} + O\left(\sqrt{\sum_{g \in \mathcal{G}} \mathbf{t}_g}\right) + O\left(\sqrt{\log \frac{1}{\eta}}\right)\right) \\
\le &\prod_{g \in \mathcal{G}} \binom{|g|}{\mathbf{t}_g} \mathbf{Pr}\left(\|X_h\| \ge \sqrt{n} + O\left(\sqrt{\sum_{g \in \mathcal{G}} \mathbf{t}_g}\right) + O\left(\sqrt{\log \frac{1}{\eta}}\right)\right) \\
\le &\eta \prod_{g \in \mathcal{G}} \binom{|g|}{\mathbf{t}_g} \le \eta \log \max_{g \in \mathcal{G}} |g| \sum_{g \in \mathcal{G}} \mathbf{t}_g \\
\Rightarrow & \\
&\mathbf{Pr}\left(\sqrt{n}\rho_+^{1/2}(3s, 3\mathbf{t}) \ge \sqrt{n} + O\left(\sqrt{\sum_{g \in \mathcal{G}} \mathbf{t}_g \log \max_{g \in \mathcal{G}} |g|}\right) + O\left(\log \frac{1}{\eta'}\right)\right) \le \eta'
\end{aligned}$$

It indicates that if $n \geq O(\sum_{g \in \mathcal{G}} \mathbf{t}_g \max_{g \in \mathcal{G}} |g|)$, then we have $\rho_+^{1/2}(3s, 3\mathbf{t}) \leq \sqrt{\frac{3}{2}}$ with high probability as well. Similarly, we can prove $\rho_-^{1/2}(3s, 3\mathbf{t}) \leq \sqrt{\frac{1}{2}}$ with high probability. □

Proof to Theorem 3

Proof. Since n is large enough as shown in (4), from Lemma 7, we have $\alpha < 1$ and are allowed to apply Theorem 2. Since $\Delta = 0$ for the noiseless case, we prove the theorem by letting $\bar{\mathbf{w}}$ be $\mathbf{w}^*$. □

Proof to Theorem 4

Proof. Since n is large enough as shown in (4), from Lemma 7, we have $\alpha < 1$ and are allowed to apply Theorem 2. From Lemma 6, we obtain the upper bound for Δ. When the number of iterations k is large enough such that $\alpha^k \|\mathbf{w}^0 - \bar{\mathbf{w}}\|$ reduces the magnitude of Δ, we can easily prove the error bound of $\mathbf{w}^k$ letting $\bar{\mathbf{w}}$ be $\mathbf{w}^*$. The second claim can be similarly proven by applying the second claim in Theorem 2. □