SUPPLEMENTARY MATERIALS: High Dimensional Data Enrichment: Interpretable, Fast, and Data-Efficient*

Amir Asiaee[†], Samet Oymak[‡], Kevin R. Coombes[§], and Arindam Banerjee[¶]

SM1. Proofs of Theorems.

Here, we present proof of Proposition 3.6 and full proof for Theorem 5.4 and each lemma used during the proofs of theorems.

SM1.1. Proof of Proposition 3.6.

Proof. Consider only one group for regression in isolation. Note that $\mathbf{y}_g = \mathbf{X}_g(\boldsymbol{\beta}_g^* + \boldsymbol{\beta}_0^*) + \mathbf{w}_g$ is a superposition model and as shown in [?] the sample complexity required for the RE condition and subsequently recovering $\boldsymbol{\beta}_0^*$ and $\boldsymbol{\beta}_g^*$ is $n_g \geq c(\max(\omega(\mathcal{A}_0), \omega(\mathcal{A}_g)) + \sqrt{\log 2})^2$. ■

SM1.2. Proof of Theorem 5.4.

Now we rewrite the same analysis using the tail bounds for the coefficients to clarify the probabilities. To simplify the notation, we define the following functions of τ:

$$
\begin{aligned}
r_{g1}(\tau) &\triangleq \frac{1}{2}\left[\left(1 - \frac{1}{a_g}\right) + \sqrt{2}c_g \frac{2\omega(\mathcal{A}_g) + \tau}{a_g\sqrt{n_g}}\right], \ \forall g \in [G_+] \\
r_{g2}(\tau) &\triangleq \frac{1}{a_g}\left(1 + c_{0g}\frac{\omega(\mathcal{A}_0) + \omega(\mathcal{A}_g) + \tau}{\sqrt{n_g}}\right), \ \forall g \in [G] \\
r_0(\tau) &\triangleq r_{01}(\tau) + \sum_{g=1}^{G}\sqrt{\frac{n_g}{n}}r_{g2}(\tau) \\
r_g(\tau) &\triangleq r_{g1}(\tau) + \sqrt{\frac{n_g}{n}}\frac{a_g}{a_0}r_{g2}(\tau), \ \forall g \in [G] \\
r(\tau) &\triangleq \max_{g\in[G_+]} r_g(\tau)
\end{aligned}
$$

All of which are computed using a_gs specified in the proof sketch of section 6. Basically $r(\tau)$ is an instantiation of an high probability upper bound of the ρ defined in Theorem 5.2. We are interested in

---

*Submitted to the editors DATE.

Funding: This work was funded by the Fog Research Institute under contract no. FRI-454.

[†]Mathematical Biosciences Institute, The Ohio State University, Columbus, OH (asiaeetaheri.1@osu.edu).
[‡]Department of Electrical and Computer Engineering, UC Riverside, Riverside, CA (oymak@ece.ucr.edu).
[§]Departments of Biomedical Informatics, The Ohio State University, Columbus, OH (coombes.3@osu.edu).
[¶]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN (banerjee@cs.umn.edu).

upper bounding the following probability:

$$
\begin{aligned}
&\mathbb{P}\left(\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\|\boldsymbol{\delta}_g^{(t+1)}\|_2 \geq r(\tau)^t \sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\|\boldsymbol{\beta}_g^*\|_2 + \frac{C(G+1)\sqrt{(2k_w^2+1)k_x^2}}{(1-r(\tau))\sqrt{n}}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) \qquad \text{(SM1.1)}\\
&\leq \mathbb{P}\left(\rho^t\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\|\boldsymbol{\beta}_g^*\|_2 + \frac{1-\rho^t}{1-\rho}\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\eta_g\left(\mu_g\right)\|\mathbf{w}_g\|_2 \geq \right.\\
&\qquad \left. r(\tau)^t\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\|\boldsymbol{\beta}_g^*\|_2 + \frac{C(G+1)\sqrt{(2k_w^2+1)k_x^2}}{(1-r(\tau))\sqrt{n}}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right)\\
&\leq \mathbb{P}\left(\rho\geq r(\tau)\right) + \mathbb{P}\left(\frac{1}{1-\rho}\sum_{g=0}^{G}\sqrt{n_g}\eta_g\left(\mu_g\right)\|\mathbf{w}_g\|_2 \geq \frac{C(G+1)\sqrt{(2k_w^2+1)k_x^2}}{(1-r(\tau))}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right)
\end{aligned}
$$

where the first inequality comes from the deterministic bound of Theorem 5.2 and the second one is based on the law of total probability.

We first focus on bounding the first term $\mathbb{P}\left(\rho\geq r(\tau)\right)$:

$$
\begin{aligned}
\mathbb{P}\left(\rho\geq r(\tau)\right) =& \mathbb{P}\left(\max\left(\rho_0\left(\mu_0\right) + \sum_{g=1}^{G}\sqrt{\frac{n_g}{n}}\phi_g\left(\mu_g\right), \max_{g\in[G]}\rho_g\left(\mu_g\right) + \sqrt{\frac{n}{n_g}}\frac{\mu_0}{\mu_g}\phi_g\left(\mu_g\right)\right) \geq \max_{g\in[G_+]} r(\tau)\right)\\
\text{(Union Bound)} \leq& \mathbb{P}\left(\rho_0\left(\mu_0\right) + \sum_{g=1}^{G}\sqrt{\frac{n_g}{n}}\phi_g\left(\mu_g\right) \geq r_0\right) + \sum_{g=1}^{G}\mathbb{P}\left(\rho_g\left(\mu_g\right) + \sqrt{\frac{n}{n_g}}\frac{\mu_0}{\mu_g}\phi_g\left(\mu_g\right) \geq r_g\right)\\
\leq& \mathbb{P}\left(\rho_0\left(\mu_0\right)\geq r_{01}\right) + \sum_{g=1}^{G}\mathbb{P}\left(\phi_g\left(\mu_g\right)\geq r_{g2}\right) + \sum_{g=1}^{G}\left[\mathbb{P}\left(\rho_g\left(\mu_g\right)\geq r_{g1}\right) + \mathbb{P}\left(\phi_g\left(\mu_g\right)\geq r_{g2}\right)\right]\\
\leq& \sum_{g=0}^{G}\mathbb{P}\left(\rho_g\left(\mu_g\right)\geq r_{g1}\right) + 2\sum_{g=1}^{G}\mathbb{P}\left(\phi_g\left(\mu_g\right)\geq r_{g2}\right)\\
\leq& \sum_{g=0}^{G}2\exp\left(-\gamma_g(\omega(\mathcal{A}_g)+\tau)^2\right) + 2\sum_{g=1}^{G}2\exp\left(-\gamma_g(\omega(\mathcal{A}_g)+\tau)^2\right)\\
(\gamma = \min_{g\in[G_+]}\gamma_g) \leq& 2(G+1)\exp\left(-\gamma\min_{g\in[G_+]}(\omega(\mathcal{A}_g)+\tau)^2\right) + 4G\exp\left(-\gamma\min_{g\in[G]}(\omega(\mathcal{A}_g)+\tau)^2\right)\\
\text{(SM1.2)} \qquad \leq& 6(G+1)\exp\left(-\gamma\min_{g\in[G_+]}(\omega(\mathcal{A}_g)+\tau)^2\right)
\end{aligned}
$$

Now we focus on bounding the second term:

$$
\begin{aligned}
&\mathbb{P}\left(\frac{1}{1-\rho}\sum_{g=0}^{G}\sqrt{n_g}\eta_g\left(\mu_g\right)\|\mathbf{w}_g\|_2 \geq \frac{C(G+1)\sqrt{(2k_w^2+1)k_x^2}}{(1-r(\tau))}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) \\
\leq&\mathbb{P}\left(\frac{1}{1-\rho}\sum_{g=0}^{G}\sqrt{n_g}\eta_g\left(\mu_g\right)\|\mathbf{w}_g\|_2 \geq \frac{C}{(1-r(\tau))}\sum_{g=0}^{G}\sqrt{(2k_w^2+1)k_x^2}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) \\
\leq&\mathbb{P}\left(\sum_{g=0}^{G}\sqrt{n_g}\eta_g\left(\mu_g\right)\|\mathbf{w}_g\|_2 \geq \sum_{g=0}^{G}c_g\sqrt{(2k_w^2+1)k_x^2}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) + \mathbb{P}\left(\rho \geq r(\tau)\right) \\
\text{(SM1.3)}\qquad \leq&\sum_{g=0}^{G}\mathbb{P}\left(\sqrt{n_g}\eta_g\left(\mu_g\right)\|\mathbf{w}_g\|_2 \geq c_g\sqrt{(2k_w^2+1)k_x^2}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) + \mathbb{P}\left(\rho \geq r(\tau)\right)
\end{aligned}
$$

Focusing on the first term, since $\eta_g(\frac{1}{a_g n_g}) = \frac{1}{a_g n_g}\sup_{\mathbf{v}\in\mathcal{B}_g}\mathbf{v}^T\mathbf{X}_g^T\frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2}, \forall g\in[G]$:

$$
\begin{aligned}
\text{(SM1.4)}\qquad &\mathbb{P}\left(\|\mathbf{w}_g\|_2\sup_{\mathbf{v}\in\mathcal{B}_g}\mathbf{v}^T\mathbf{X}_g^T\frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2} \geq a_g c_g\sqrt{(2k_w^2+1)k_x^2 n_g}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) \\
(a_g\geq 1)\ \leq&\mathbb{P}\left(\|\mathbf{w}_g\|_2\sup_{\mathbf{v}\in\mathcal{B}_g}\mathbf{v}^T\mathbf{X}_g^T\frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2} \geq c_g\sqrt{(2k_w^2+1)k_x^2 n_g}(\max_{g\in[G_+]}\omega(\mathcal{A}_g)+\tau)\right) \\
\text{((SM1.8) and (SM1.9))}\ \leq&2\exp(-\nu n_g) + \pi_g\exp(-\tau^2) \\
(\sigma_g = \max(\pi_g, 2))\ \leq&\sigma_g\exp(-\min(\nu n_g, \tau^2))
\end{aligned}
$$

where we used the intermediate form of Lemma 4.4 for $\tau > 0$. Putting all of the bounds (SM1.2), (SM1.3), and (SM1.4) back as the upper bound of subsection SM1.2:

$$
\begin{aligned}
(\pi = \max_{g\in[G]}\pi_g) \leq& 2(G+1)\exp(-\nu\min_{g\in[G]}n_g) + \pi(G+1)\exp(-\tau^2) + \\
&6(G+1)\exp\left(-\gamma\min_{g\in[G_+]}(\omega(\mathcal{A}_g)+\tau)^2\right) \\
(\upsilon = \max(6,\pi), \zeta = \min(1,\gamma)) \leq& 2\exp(-\nu\min_{g\in[G]}n_g + \log(G+1)) + \upsilon(G+1)\exp(-\zeta\tau^2) \\
(\tau = \theta + \sqrt{\log(G+1)/\zeta}) \leq& 2\exp(-\nu\min_{g\in[G]}n_g + \log(G+1)) + \upsilon\exp(-\zeta\theta^2) \\
(\sigma = \max(2,\upsilon)) \leq& \sigma\exp(-\min(\nu\min_{g\in[G]}n_g - \log(G+1), \zeta\theta^2))
\end{aligned}
$$

Note that setting $\tau = \theta + \sqrt{\log(G+1)}/\zeta$ increases the sample complexities to the followings:

$$
n > 2c_0^2\left(2\omega(\mathcal{A}_0) + \sqrt{\log(G+1)/\zeta} + \theta\right)^2, \forall g\in[G]: n_g \geq 2c_g^2(2\omega(\mathcal{A}_g) + C\sqrt{\log(G+1)/\zeta} + \theta)^2
$$

And it also affects step sizes as follows:

$$
\mu_0 = \frac{1}{4n}\times\min_{g\in[G]}\left(1 + c_{0g}\frac{\omega_{0g} + \sqrt{\log(G+1)/\zeta} + \theta}{\sqrt{n_g}}\right)^{-2}, \mu_g = \frac{1}{2\sqrt{n n_g}}\left(1 + c_{0g}\frac{\omega_{0g} + \sqrt{\log(G+1)/\zeta} + \theta}{\sqrt{n_g}}\right)^{-1}
$$

SM1.3. Proof of Lemma 3.10.

Proof. LHS of (3.2) is the weighted summation of $\xi_g Q_{2\xi_g}(\boldsymbol{\delta}_{0g}) = \|\boldsymbol{\delta}_{0g}\|_2 \xi \mathbb{P}(|\langle \mathbf{x}, , \boldsymbol{\delta}_{0g}/\|\boldsymbol{\delta}_{0g}\|_2 \rangle| > 2\xi) = \|\boldsymbol{\delta}_{0g}\|_2 \xi Q_{2\xi}(\mathbf{u})$ where $\xi > 0$ and $\mathbf{u} = \boldsymbol{\delta}_{0g}/\|\boldsymbol{\delta}_{0g}\|_2$ is a unit length vector. So we can rewrite the LHS of (3.2) as:

$$\sum_{g=1}^{G} \frac{n_g}{n} \xi_g Q_{2\xi_g}(\boldsymbol{\delta}_{0g}) = \sum_{g=1}^{G} \frac{n_g}{n} \|\boldsymbol{\delta}_0 + \boldsymbol{\delta}_g\|_2 \xi Q_{2\xi}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{S}^{p-1}$$

With this observation, the lower bound of the Lemma 3.10 is a direct consequence of the following two results:

Lemma SM1.1. *Let* $\mathbf{u}$ *be any unit length vector and suppose* $\mathbf{x}$ *obeys Definition 3.1. Then for any* $\mathbf{u}$*, we have*

$$Q_{2\xi}(\mathbf{u}) \geq \frac{(\alpha - 2\xi)^2}{4ck_x^2}. \tag{SM1.5}$$

Lemma SM1.2. *Suppose Definition 3.4 holds. Then, we have:*

$$\sum_{g=1}^{G} \frac{n_g}{n} \|\boldsymbol{\delta}_0 + \boldsymbol{\delta}_g\|_2 \geq \frac{\bar{\rho}\lambda_{\min}}{3} \left(G\|\boldsymbol{\delta}_0\|_2 + \sum_{g=1}^{G} \frac{n_g}{n} \|\boldsymbol{\delta}_g\|_2 \right), \quad \forall g \in [G_+] : \boldsymbol{\delta}_g \in \mathcal{C}_g. \tag{SM1.6}$$

SM1.4. Proof of Lemma 3.11.

Proof. Consider the following soft indicator function which we use in our derivation:

$$\psi_a(s) = \begin{cases} 0, & |s| \leq a \\ (|s| - a)/a, & a \leq |s| \leq 2a \\ 1, & 2a < |s| \end{cases}$$

Now using the definition of the marginal tail function we have:

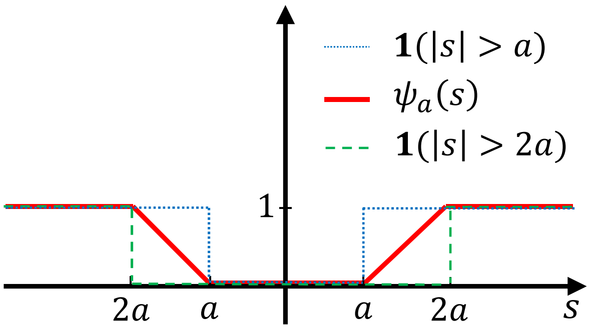

Figure SM1: $\mathbb{1}(|s| > 2a) \leq \psi_a(s) \leq \mathbb{1}(|s| > a)$

$$
\begin{aligned}
&\mathbb{E}\sup_{\boldsymbol{\delta}\in\mathcal{H}}\sum_{g=1}^{G}\xi_g\sum_{i=1}^{n_g}\left[Q_{2\xi_g}(\boldsymbol{\delta}_{0g})-\mathbb{1}(|\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle|\geq\xi_g)\right]\\
&=\mathbb{E}\sup_{\boldsymbol{\delta}\in\mathcal{H}}\sum_{g=1}^{G}\xi_g\sum_{i=1}^{n_g}\left[\mathbb{E}\mathbb{1}(|\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle|\geq 2\xi_g)-\mathbb{1}(|\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle|\geq\xi_g)\right]\\
(\text{Figure Figure SM1})\ &\leq\mathbb{E}\sup_{\boldsymbol{\delta}\in\mathcal{H}}\sum_{g=1}^{G}\xi_g\sum_{i=1}^{n_g}\left[\mathbb{E}\psi_{\xi_g}(\langle\mathbf{x},\boldsymbol{\delta}_{0g}\rangle)-\psi_{\xi_g}(\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle)\right]\\
(\text{Symmetrization [?]})\ &\leq 2\mathbb{E}\sup_{\boldsymbol{\delta}\in\mathcal{H}}\sum_{g=1}^{G}\xi_g\sum_{i=1}^{n_g}\epsilon_{gi}\psi_{\xi_g}(\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle)\\
(\text{Rademacher comparison [?]})\ &\leq 2\mathbb{E}\sup_{\boldsymbol{\delta}\in\mathcal{H}}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\epsilon_{gi}\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle
\end{aligned}
$$

where ϵ_{gi} are iid copies of Rademacher random variable which are independent of every other random variables and themselves.

Now we add back $\frac{1}{n}$ and expand $\boldsymbol{\delta}_{0g}=\boldsymbol{\delta}_0+\boldsymbol{\delta}_g$. Also, we substitute $\boldsymbol{\delta}\in\mathcal{H}$ constraint with $\boldsymbol{\delta}\in\mathcal{C}$ because $\mathcal{H}\subseteq\mathcal{C}$ where $\mathcal{C}=\{\boldsymbol{\delta}=(\boldsymbol{\delta}_0^T,\dots,\boldsymbol{\delta}_G^T)^T\big|\boldsymbol{\delta}_g\in\mathcal{C}_g\}$:

$$
\begin{aligned}
\frac{2}{n}\mathbb{E}\sup_{\boldsymbol{\delta}\in\mathcal{C}}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\epsilon_{gi}\langle\mathbf{x}_{gi},\boldsymbol{\delta}_{0g}\rangle &=\frac{2}{n}\mathbb{E}\sup_{\boldsymbol{\delta}_0\in\mathcal{C}_0}\sum_{i=1}^{n}\epsilon_i\langle\mathbf{x}_i,\boldsymbol{\delta}_0\rangle+\frac{2}{n}\mathbb{E}\sup_{\forall g\in[G]:\boldsymbol{\delta}_g\in\mathcal{C}_g}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\epsilon_{gi}\langle\mathbf{x}_{gi},\boldsymbol{\delta}_g\rangle\\
&=\frac{2}{\sqrt{n}}\mathbb{E}\sup_{\boldsymbol{\delta}_0\in\mathcal{C}_0}\sum_{i=1}^{n}\langle\frac{1}{\sqrt{n}}\epsilon_i\mathbf{x}_i,\boldsymbol{\delta}_0\rangle+\frac{2}{\sqrt{n}}\mathbb{E}\sup_{\forall g\in[G]:\boldsymbol{\delta}_g\in\mathcal{C}_g}\sum_{g=1}^{G}\sqrt{\frac{n_g}{n}}\sum_{i=1}^{n_g}\langle\frac{1}{\sqrt{n_g}}\epsilon_{gi}\mathbf{x}_{gi},\boldsymbol{\delta}_g\rangle\\
(n_0:=n,\epsilon_{0i}:=\epsilon_0,\mathbf{x}_{0i}:=\mathbf{x}_i) &=\frac{2}{\sqrt{n}}\mathbb{E}\sup_{\forall g\in[G_+]:\boldsymbol{\delta}_g\in\mathcal{C}_g}\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\sum_{i=1}^{n_g}\langle\frac{1}{\sqrt{n_g}}\epsilon_{gi}\mathbf{x}_{gi},\boldsymbol{\delta}_g\rangle\\
(\mathbf{h}_g:=\frac{1}{\sqrt{n_g}}\sum_{i=1}^{n_g}\epsilon_{gi}\mathbf{x}_{gi}) &=\frac{2}{\sqrt{n}}\mathbb{E}\sup_{\forall g\in[G_+]:\boldsymbol{\delta}_g\in\mathcal{C}_g}\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\langle\mathbf{h}_g,\boldsymbol{\delta}_g\rangle\\
(\mathbf{u}_g\in\frac{\boldsymbol{\delta}_g}{\|\boldsymbol{\delta}_g\|_2},\mathcal{A}_g\in\mathcal{C}_g\cap\mathbb{S}^{p-1}) &\leq\frac{2}{\sqrt{n}}\mathbb{E}\sup_{\forall g\in[G_+]:\mathbf{u}_g\in\mathcal{A}_g}\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\langle\mathbf{h}_g,\mathbf{u}_g\rangle\|\boldsymbol{\delta}_g\|_2\\
(\sup\sum<\sum\sup) &\leq\frac{2}{\sqrt{n}}\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}\mathbb{E}_{\mathbf{h}_g}\sup_{\mathbf{u}_g\in\mathcal{A}_g}\langle\mathbf{h}_g,\mathbf{u}_g\rangle\|\boldsymbol{\delta}_g\|_2\\
&\leq\frac{2}{\sqrt{n}}\sum_{g=0}^{G}\sqrt{\frac{n_g}{n}}c_gk_x\omega(\mathcal{A}_g)\|\boldsymbol{\delta}_g\|_2
\end{aligned}
$$

Note that the $\mathbf{h}_{gi}$ is a sub-Gaussian random vector which let us bound the $\mathbb{E}\sup$ using the Gaussian width [?] in the last step. ■

SM1.5. Proof of Lemma 4.4.

Proof. To avoid cluttering let $h_g(\mathbf{w}_g, \mathbf{X}_g) \triangleq \|\mathbf{w}_g\|_2 \sup_{\mathbf{u}_g \in \mathcal{A}_g} \langle \mathbf{X}_g^T \frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2}, \mathbf{u}_g \rangle$ be a random quantity and $e_g(\tau) \triangleq c_g k_x(\omega(\mathcal{A}_g) + \sqrt{\log(G+1)} + \tau)$, and $s_g \triangleq \sqrt{(2k_w^2+1)n_g}$ constants. From the law of total probability, we have:

$$
\begin{aligned}
\mathbb{P}\left(h_g(\mathbf{w}_g, \mathbf{X}_g) > s_g e_g(\tau)\right) =& \mathbb{P}\left(h_g(\mathbf{w}_g, \mathbf{X}_g) > s_g e_g(\tau) \Big| \|\mathbf{w}_g\|_2 > s_g\right) \mathbb{P}\left(\|\mathbf{w}_g\|_2 > s_g\right) \\
&+ \mathbb{P}\left(h_g(\mathbf{w}_g, \mathbf{X}_g) > s_g e_g(\tau) \Big| \|\mathbf{w}_g\|_2 < s_g\right) \mathbb{P}\left(\|\mathbf{w}_g\|_2 < s_g\right) \\
\leq& \mathbb{P}\left(\|\mathbf{w}_g\|_2 > s_g\right) + \mathbb{P}\left(\|\mathbf{w}_g\|_2 \sup_{\mathbf{u}_g \in \mathcal{A}_g} \langle \mathbf{X}_g^T \frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2}, \mathbf{u}_g \rangle > s_g e_g(\tau) \Big| \|\mathbf{w}_g\|_2 < s_g\right) \\
\leq& \mathbb{P}\left(\|\mathbf{w}_g\|_2 > \sqrt{(2k_w^2+1)n_g}\right) + \mathbb{P}\left(\sup_{\mathbf{u}_g \in \mathcal{C}_g \cap \mathbb{S}^{p-1}} \langle \mathbf{X}_g^T \frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2}, \mathbf{u}_g \rangle > e_g(\tau)\right) \\
\leq& \mathbb{P}\left(\|\mathbf{w}_g\|_2 > \sqrt{(2k_w^2+1)n_g}\right) + \sup_{\mathbf{v} \in \mathbb{S}^{p-1}} \mathbb{P}\left(\sup_{\mathbf{u}_g \in \mathcal{C}_g \cap \mathbb{S}^{p-1}} \langle \mathbf{X}_g^T \mathbf{v}, \mathbf{u}_g \rangle > e_g(\tau)\right) \qquad \text{(SM1.7)}
\end{aligned}
$$

Let's focus on the first term. Since $\mathbf{w}_g$ consists of i.i.d. centered unit-variance sub-Gaussian elements with $|||w_{gi}|||_{\psi_2} < k_w$, w_{gi}^2 is sub-exponential with $|||\mathbf{w}_{gi}|||_{\psi_1} < 2k_w^2$. Let's apply the Bernstein's inequality [?] to $\|\mathbf{w}_g\|_2^2 = \sum_{i=1}^{n_g} \mathbf{w}_{gi}^2$:

$$
\mathbb{P}\left(\left|\|\mathbf{w}_g\|_2^2 - \mathbb{E}\|\mathbf{w}_g\|_2^2\right| > \tau\right) \leq 2\exp\left(-\nu \min\left[\frac{\tau^2}{4k_w^4 n_g}, \frac{\tau}{2k_w^2}\right]\right)
$$

We also know that $\mathbb{E}\|\mathbf{w}_g\|_2^2 \leq n_g$ [?] which gives us:

$$
\mathbb{P}\left(\|\mathbf{w}_g\|_2 > \sqrt{n_g + \tau}\right) \leq 2\exp\left(-\nu \min\left[\frac{\tau^2}{4k_w^4 n_g}, \frac{\tau}{2k_w^2}\right]\right)
$$

Finally, we set $\tau = 2k_w^2 n_g$:

$$
\mathbb{P}\left(\|\mathbf{w}_g\|_2 > \sqrt{(2k_w^2+1)n_g}\right) \leq 2\exp\left(-\nu n_g\right) = \frac{2}{(G+1)} \exp\left(-\nu n_g + \log(G+1)\right) \qquad \text{(SM1.8)}
$$

Now we upper bound the second term of (SM1.7). Given any fixed $\mathbf{v} \in \mathbb{S}^{p-1}$, $\mathbf{X}_g \mathbf{v}$ is a sub-Gaussian random vector with $\left|\left|\left|\mathbf{X}_g^T \mathbf{v}\right|\right|\right|_{\psi_2} \leq C_g k_x$ [?]. From Theorem 9 of [?] for any $\mathbf{v} \in \mathbb{S}^{p-1}$ we have:

$$
\mathbb{P}\left(\sup_{\mathbf{u}_g \in \mathcal{A}_g} \langle \mathbf{X}_g^T \mathbf{v}, \mathbf{u}_g \rangle > \upsilon_g C_g k_x \omega(\mathcal{A}_g) + t\right) \leq \pi_g \exp\left(-\left(\frac{t}{\theta_g C_g k_x \phi_g}\right)^2\right) \qquad \text{(SM1.9)}
$$

where $\phi_g = \sup_{\mathbf{u}_g \in \mathcal{A}_g} \|\mathbf{u}_g\|_2$ and in our problem $\phi_g = 1$. To simplify we take $c_g = C_g \max(\upsilon_g, \theta_g)$ and then substitute $t = c_g k_x(\tau + \sqrt{\log(G+1)})$:

$$
\begin{aligned}
\mathbb{P}\left(\sup_{\mathbf{u}_g \in \mathcal{A}_g} \langle \mathbf{X}_g^T \mathbf{v}, \mathbf{u}_g \rangle > c_g k_x \left(\omega(\mathcal{A}_g) + \sqrt{\log(G+1)} + \tau\right)\right) &\leq \pi_g \exp\left(-\left(\tau + \sqrt{\log(G+1)}\right)^2\right) \\
&\leq \pi_g \exp\left(-\log(G+1) - \tau^2\right) \\
&\leq \frac{\pi_g}{(G+1)} \exp(-\tau^2)
\end{aligned}
$$

Now we put back results to the original inequality (SM1.7):

$$\mathbb{P}\left(h_g(\mathbf{w}_g, \mathbf{X}_g) > \sqrt{(2k_w^2+1)n_g} \times c_g k_x \left(\omega(\mathcal{A}_g) + \sqrt{\log(G+1)} + \tau\right)\right)$$
$$\leq \frac{2}{(G+1)} \exp\left(-\nu n_g + \log(G+1)\right) + \frac{\pi_g}{(G+1)} \exp\left(-\tau^2\right)$$
$$\leq \frac{\sigma_g}{(G+1)} \exp\left(-\min\left[\nu n_g - \log(G+1), \tau^2\right]\right)$$

where $\sigma_g = \pi_g + 2$. ∎

SM1.6. Proof of Lemma 5.3.

Proof. We upper bound the individual error $\|\boldsymbol{\delta}_g^{(t+1)}\|_2$ and the common one $\|\boldsymbol{\delta}_0^{(t+1)}\|_2$ in the followings:

$$\begin{aligned}
\|\boldsymbol{\delta}_g^{(t+1)}\|_2 &= \|\boldsymbol{\beta}_g^{(t+1)} - \boldsymbol{\beta}_g^*\|_2 \\
&= \left\|\Pi_{\Omega_{f_g}}\left(\boldsymbol{\beta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\mathbf{y}_g - \mathbf{X}_g(\boldsymbol{\beta}_0^{(t)} + \boldsymbol{\beta}_g^{(t)})\right)\right) - \boldsymbol{\beta}_g^*\right\|_2 \\
\text{(Lemma 6.3 of [?])} &= \left\|\Pi_{\Omega_{f_g} - \{\boldsymbol{\beta}_g^*\}}\left(\boldsymbol{\beta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\mathbf{y}_g - \mathbf{X}_g(\boldsymbol{\beta}_0^{(t)} + \boldsymbol{\beta}_g^{(t)})\right) - \boldsymbol{\beta}_g^*\right)\right\|_2 \\
&= \left\|\Pi_{\mathcal{E}_g}\left(\boldsymbol{\delta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\mathbf{y}_g - \mathbf{X}_g(\boldsymbol{\beta}_0^{(t)} + \boldsymbol{\beta}_g^{(t)}) - \mathbf{X}_g(\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_g^*) + \mathbf{X}_g(\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_g^*)\right)\right)\right\|_2 \\
&= \left\|\Pi_{\mathcal{E}_g}\left(\boldsymbol{\delta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\boldsymbol{\omega}_g - \mathbf{X}_g(\boldsymbol{\delta}_0^{(t)} + \boldsymbol{\delta}_g^{(t)})\right)\right)\right\|_2 \\
\text{(Lemma 6.4 of [?])} &\leq \left\|\Pi_{\mathcal{C}_g}\left(\boldsymbol{\delta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\boldsymbol{\omega}_g - \mathbf{X}_g(\boldsymbol{\delta}_0^{(t)} + \boldsymbol{\delta}_g^{(t)})\right)\right)\right\|_2 \\
\text{(Lemma 6.2 of [?])} &\leq \sup_{\mathbf{v} \in \mathcal{C}_g \cap \mathbb{B}^p} \mathbf{v}^T\left(\boldsymbol{\delta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\boldsymbol{\omega}_g - \mathbf{X}_g(\boldsymbol{\delta}_0^{(t)} + \boldsymbol{\delta}_g^{(t)})\right)\right) \\
(\mathcal{B}_g = \mathcal{C}_g \cap \mathbb{B}^p) &= \sup_{\mathbf{v} \in \mathcal{B}_g} \mathbf{v}^T\left(\boldsymbol{\delta}_g^{(t)} + \mu_g \mathbf{X}_g^T\left(\boldsymbol{\omega}_g - \mathbf{X}_g(\boldsymbol{\delta}_0^{(t)} + \boldsymbol{\delta}_g^{(t)})\right)\right) \\
&\leq \sup_{\mathbf{v} \in \mathcal{B}_g} \mathbf{v}^T(\mathbf{I}_g - \mu_g \mathbf{X}_g^T \mathbf{X}_g)\boldsymbol{\delta}_g^{(t)} + \mu_g \sup_{\mathbf{v} \in \mathcal{B}_g} \mathbf{v}^T \mathbf{X}_g^T \boldsymbol{\omega}_g + \mu_g \sup_{\mathbf{v} \in \mathcal{B}_g} -\mathbf{v}^T \mathbf{X}_g^T \mathbf{X}_g \boldsymbol{\delta}_0^{(t)} \\
&\leq \left\|\boldsymbol{\delta}_g^{(t)}\right\|_2 \sup_{\mathbf{u}, \mathbf{v} \in \mathcal{B}_g} \mathbf{v}^T(\mathbf{I}_g - \mu_g \mathbf{X}_g^T \mathbf{X}_g)\mathbf{u} + \mu_g \|\boldsymbol{\omega}_g\|_2 \sup_{\mathbf{v} \in \mathcal{B}_g} \mathbf{v}^T \mathbf{X}_g^T \frac{\boldsymbol{\omega}_g}{\|\boldsymbol{\omega}_g\|_2} \\
&+ \mu_g \|\boldsymbol{\delta}_0^{(t)}\|_2 \sup_{\mathbf{v} \in \mathcal{B}_g, \mathbf{u} \in \mathcal{B}_0} -\mathbf{v}^T \mathbf{X}_g^T \mathbf{X}_g \mathbf{u} \\
&= \rho_g(\mu_g)\|\boldsymbol{\delta}_g^{(t)}\|_2 + \xi_g(\mu_g)\|\boldsymbol{\omega}_g\|_2 + \phi_g(\mu_g)\|\boldsymbol{\delta}_0^{(t)}\|_2
\end{aligned}$$

So the final bound becomes:

$$\|\boldsymbol{\delta}_g^{(t+1)}\|_2 \leq \rho_g(\mu_g)\|\boldsymbol{\delta}_g^{(t)}\|_2 + \xi_g(\mu_g)\|\boldsymbol{\omega}_g\|_2 + \phi_g(\mu_g)\|\boldsymbol{\delta}_0^{(t)}\|_2 \tag{SM1.10}$$

Now we upper bound the error of common parameter. Remember common parameter's update: $\beta_0^{(t+1)} = \Pi_{\Omega_{f_0}}\left(\beta_0^{(t)} + \mu_0 \mathbf{X}_0^T \begin{pmatrix} (\mathbf{y}_1 - \mathbf{X}_1(\beta_0^{(t)} + \beta_1^{(t)})) \\ \vdots \\ (\mathbf{y}_G - \mathbf{X}_G(\beta_0^{(t)} + \beta_G^{(t)})) \end{pmatrix}\right)$.

$$
\begin{aligned}
\|\delta_0^{(t+1)}\|_2 =& \|\beta_0^{(t+1)} - \beta_0^*\|_2 \\
=& \left\| \Pi_{\Omega_{f_0}}\left(\beta_0^{(t)} + \mu_0 \sum_{g=1}^G \mathbf{X}_g^T \left(\mathbf{y}_g - \mathbf{X}_g(\beta_0^{(t)} + \beta_g^{(t)})\right)\right) - \beta_0^* \right\|_2 \\
\text{(Lemma 6.3 of [?])} =& \left\| \Pi_{\Omega_{f_0} - \{\beta_0^*\}}\left(\beta_0^{(t)} + \mu_0 \sum_{g=1}^G \mathbf{X}_g^T \left(\mathbf{y}_g - \mathbf{X}_g(\beta_0^{(t)} + \beta_g^{(t)})\right) - \beta_0^*\right) \right\|_2 \\
=& \left\| \Pi_{\mathcal{E}_0}\left(\delta_0^{(t)} + \mu_0 \sum_{g=1}^G \mathbf{X}_g^T \left(\mathbf{y}_g - \mathbf{X}_g(\beta_0^{(t)} + \beta_g^{(t)})\right)\right) \right\|_2 \\
\text{(Lemma 6.4 of [?])} \le& \left\| \Pi_{\mathcal{C}_0}\left(\delta_0^{(t)} + \mu_0 \sum_{g=1}^G \mathbf{X}_g^T \left(\omega_g - \mathbf{X}_g(\delta_0^{(t)} + \delta_g^{(t)})\right)\right) \right\|_2 \\
\text{(Lemma 6.2 of [?])} \le& \sup_{\mathbf{v}\in\mathcal{B}_0} \mathbf{v}^T \left(\delta_0^{(t)} + \mu_0 \sum_{g=1}^G \mathbf{X}_g^T \left(\omega_g - \mathbf{X}_g(\delta_0^{(t)} + \delta_g^{(t)})\right)\right) \\
\le& \sup_{\mathbf{v}\in\mathcal{B}_0} \mathbf{v}^T (\mathbf{I} - \mu_0 \sum_{g=1}^G \mathbf{X}_g^T \mathbf{X}_g) \delta_0^{(t)} + \mu_0 \sup_{\mathbf{v}\in\mathcal{B}_0} \mathbf{v}^T \sum_{g=1}^G \mathbf{X}_g^T \omega_g \\
&+ \mu_0 \sup_{\mathbf{v}\in\mathcal{B}_0} -\mathbf{v}^T \sum_{g=1}^G \mathbf{X}_g^T \mathbf{X}_g \delta_g^{(t)} \\
\le& \|\delta_0^{(t)}\|_2 \sup_{\mathbf{u},\mathbf{v}\in\mathcal{B}_0} \mathbf{v}^T \left(\mathbf{I} - \mu_0 \mathbf{X}_0^T \mathbf{X}_0\right) \mathbf{u} + \mu_0 \sup_{\mathbf{v}\in\mathcal{B}_0} \mathbf{v}^T \mathbf{X}_0^T \frac{\omega_0}{\|\omega_0\|_2} \|\omega_0\|_2 \\
&+ \mu_0 \sum_{g=1}^G \sup_{\mathbf{v}_g\in\mathcal{B}_0, \mathbf{u}_g\in\mathcal{B}_g} -\mathbf{v}_g^T \mathbf{X}_g^T \mathbf{X}_g \mathbf{u}_g \|\delta_g^{(t)}\|_2 \\
\le& \rho_0(\mu_0)\|\delta_0^{(t)}\|_2 + \xi_0(\mu_0)\|\omega_0\|_2 + \mu_0 \sum_{g=1}^G \frac{\phi_g(\mu_g)}{\mu_g} \|\delta_g^{(t)}\|_2 \qquad \text{(SM1.11)}
\end{aligned}
$$

■

To avoid cluttering we drop μ_g as the arguments. Putting together (SM1.10) and (SM1.11) inequalities we reach to the followings:

$$
\begin{aligned}
\|\boldsymbol{\delta}_g^{(t+1)}\|_2 &\leq \rho_g \|\boldsymbol{\delta}_g^{(t)}\|_2 + \xi_g \|\boldsymbol{\omega}_g\|_2 + \phi_g \|\boldsymbol{\delta}_0^{(t)}\|_2 \\
\|\boldsymbol{\delta}_0^{(t+1)}\|_2 &\leq \rho_0 \|\boldsymbol{\delta}_0^{(t)}\|_2 + \xi_0 \|\boldsymbol{\omega}_0\|_2 + \mu_0 \sum_{g=1}^{G} \frac{\phi_g}{\mu_g} \|\boldsymbol{\delta}_g^{(t)}\|_2
\end{aligned}
$$

SM1.7. Proof of Lemma 6.1.

We will need the following lemma in our proof. It establishes the RE condition for individual isotropic sub-Gaussian designs and provides us with the essential tool for proving high probability bounds.

Lemma SM1.3 (Theorem 11 of [?]). *For all $g \in [G]$, for the matrix $\mathbf{X}_g \in \mathbb{R}^{n_g \times p}$ with independent isotropic sub-Gaussian rows, i.e., $|||\mathbf{x}_{gi}|||_{\psi_2} \leq k_x$ and $\mathbb{E}[\mathbf{x}_{gi}\mathbf{x}_{gi}^T] = \mathbf{I}$, the following result holds with probability at least $1 - 2\exp\left(-\gamma_g(\omega(\mathcal{A}_g) + \tau)^2\right)$ for $\tau > 0$:*

$$
\forall \mathbf{u}_g \in \mathcal{C}_g : n_g \left(1 - c_g \frac{\omega(\mathcal{A}_g) + \tau}{\sqrt{n_g}}\right) \|\mathbf{u}_g\|_2^2 \leq \|\mathbf{X}_g \mathbf{u}_g\|_2^2 \leq n_g \left(1 + c_g \frac{\omega(\mathcal{A}_g) + \tau}{\sqrt{n_g}}\right) \|\mathbf{u}_g\|_2^2
$$

where $c_g > 0$ is constant.

The statement of Lemma SM1.3 characterizes the distortion in the Euclidean distance between points $\mathbf{u}_g \in \mathcal{C}_g$ when the matrix $\mathbf{X}_g / n_g$ is applied to them and states that any sub-Gaussian design matrix is approximately isometry, with high probability:

$$
(1 - \alpha)\|\mathbf{u}_g\|_2^2 \leq \frac{1}{n_g} \|\mathbf{X}_g \mathbf{u}_g\|_2^2 \leq (1 + \alpha)\|\mathbf{u}_g\|_2^2
$$

where $\alpha = c_g \frac{\omega(\mathcal{A}_g)}{\sqrt{n_g}}$. Now the proof for Lemma 6.1:

SM1.7.1. Bounding $\rho_g(\mu_g)$.

Proof. First we upper bound each of the coefficients $\forall g \in [G]$:

$$
\rho_g(\mu_g) = \sup_{\mathbf{u}, \mathbf{v} \in \mathcal{B}_g} \mathbf{v}^T \left(\mathbf{I}_g - \mu_g \mathbf{X}_g^T \mathbf{X}_g\right) \mathbf{u}
$$

We upper bound the argument of the sup as follows:

$$\begin{aligned}
\mathbf{v}^T(\mathbf{I}_g - \mu_g \mathbf{X}_g^T \mathbf{X}_g)\mathbf{u} &= \frac{1}{4}\left[(\mathbf{u}+\mathbf{v})^T(\mathbf{I} - \mu_g \mathbf{X}_g^T \mathbf{X}_g)(\mathbf{u}+\mathbf{v}) - (\mathbf{u}-\mathbf{v})^T(\mathbf{I} - \mu_g \mathbf{X}_g^T \mathbf{X}_g)(\mathbf{u}-\mathbf{v})\right] \\
&= \frac{1}{4}\left[\|\mathbf{u}+\mathbf{v}\|_2^2 - \mu_g\|\mathbf{X}_g(\mathbf{u}+\mathbf{v})\|_2^2 - \|\mathbf{u}-\mathbf{v}\|_2^2 + \mu_g\|\mathbf{X}_g(\mathbf{u}-\mathbf{v})\|_2^2\right] \\
(\text{Lemma SM1.3}) &\le \frac{1}{4}\Bigg[\left(1 - \mu_g n_g\left(1 - c_g\frac{2\omega(\mathcal{A}_g)+\tau}{\sqrt{n_g}}\right)\right)\|\mathbf{u}+\mathbf{v}\|_2 \\
&\quad - \left(1 - \mu_g n_g\left(1 + c_g\frac{2\omega(\mathcal{A}_g)+\tau}{\sqrt{n_g}}\right)\right)\|\mathbf{u}-\mathbf{v}\|_2\Bigg] \\
\left(\mu_g = \frac{1}{a_g n_g}\right) &\le \frac{1}{4}\left[\left(1 - \frac{1}{a_g}\right)(\|\mathbf{u}+\mathbf{v}\|_2 - \|\mathbf{u}-\mathbf{v}\|_2) + c_g\frac{2\omega(\mathcal{A}_g)+\tau}{a_g\sqrt{n_g}}(\|\mathbf{u}+\mathbf{v}\|_2 + \|\mathbf{u}-\mathbf{v}\|_2)\right] \\
&\le \frac{1}{4}\left[\left(1 - \frac{1}{a_g}\right)2\|\mathbf{v}\|_2 + c_g\frac{2\omega(\mathcal{A}_g)+\tau}{a_g\sqrt{n_g}}2\sqrt{2}\right]
\end{aligned}$$

where the last line follows from the triangle inequality and the fact that $\|\mathbf{u}+\mathbf{v}\|_2 + \|\mathbf{u}-\mathbf{v}\|_2 \le 2\sqrt{2}$ which itself follows from $\|\mathbf{u}+\mathbf{v}\|_2^2 + \|\mathbf{u}-\mathbf{v}\|_2^2 \le 4$. Note that we applied the Lemma SM1.3 for bigger sets of $\mathcal{A}_g + \mathcal{A}_g$ and $\mathcal{A}_g - \mathcal{A}_g$ where Gaussian width of both of them are upper bounded by $2\omega(\mathcal{A}_g)$.

The above holds with high probability which is computed as follows. Let's set $\mu_g = \frac{1}{a_g n_g}$, $d_g := \frac{1}{2}\left(1 - \frac{1}{a_g}\right) + \sqrt{2}c_g\frac{\omega(\mathcal{A}_g)+\tau/2}{a_g\sqrt{n_g}}$ and name the bad events of $\|\mathbf{X}_g(\mathbf{u}+\mathbf{v})\|_2^2 < n_g\left(1 - c_g\frac{2\omega(\mathcal{A}_g)+\tau}{\sqrt{n_g}}\right)$ and $\|\mathbf{X}_g(\mathbf{u}-\mathbf{v})\|_2^2 > n_g\left(1 + c_g\frac{2\omega(\mathcal{A}_g)+\tau}{\sqrt{n_g}}\right)$ as $\mathcal{E}_1$ and $\mathcal{E}_2$ respectively. Then from the law of total probability we have:

$$\begin{aligned}
\mathbb{P}(\rho_g(\mu_g) \ge d_g) &\le \mathbb{P}(\rho_g(\mu_g) \ge d_g | \neg\mathcal{E}_1, \neg\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_1, \mathcal{E}_2) \\
&\le 0 + \mathbb{P}(\mathcal{E}_1|\mathcal{E}_2)\mathbb{P}(\mathcal{E}_2) \le \mathbb{P}(\mathcal{E}_2) \\
(\text{Lemma SM1.3}) &\le 2\exp\left(-\gamma_g(\omega(\mathcal{A}_g)+\tau)^2\right)
\end{aligned}$$

which concludes the proof. ■

SM1.7.2. Bounding $\eta_g(\mu_g)$.

Proof. The proof of this bound has been worked out during the proof of Lemma 4.4 where we show the following in equations (SM1.7) and (SM1.9)

$$\eta_g(\mu_g) = \mu_g \sup_{\mathbf{v}\in\mathcal{B}_g} \mathbf{v}^T\mathbf{X}_g^T\frac{\mathbf{w}_g}{\|\mathbf{w}_g\|_2} = \mu_g c_g k_x(\omega(\mathcal{A}_g)+\tau), \quad \text{w.p. at least} \quad 1 - \pi_g\exp\left(-\tau^2\right)$$ ■

SM1.7.3. Bounding $\phi_g(\mu_g)$.

Proof. The following holds for any $\mathbf{u}$ and $\mathbf{v}$ because of $\|\mathbf{X}_g(\mathbf{u}+\mathbf{v})\|_2^2 \ge 0$:

$$-\mathbf{v}^T\mathbf{X}_g^T\mathbf{X}_g\mathbf{u} \le \frac{1}{2}\left(\|\mathbf{X}_g\mathbf{u}\|_2^2 + \|\mathbf{X}_g\mathbf{v}\|_2^2\right)$$

Now we can bound ϕ_g as follows:

$$\phi_g(\mu_g) = \mu_g \sup_{\mathbf{v}\in\mathcal{B}_g, \mathbf{u}\in\mathcal{B}_0} -\mathbf{v}^T\mathbf{X}_g^T\mathbf{X}_g\mathbf{u} \leq \frac{\mu_g}{2}\left(\sup_{\mathbf{u}\in\mathcal{B}_0}\|\mathbf{X}_g\mathbf{u}\|_2^2 + \sup_{\mathbf{v}\in\mathcal{B}_g}\|\mathbf{X}_g\mathbf{v}\|_2^2\right)$$

So we have:

$$\begin{aligned}
\phi_g\left(\frac{1}{a_g n_g}\right) &\leq \frac{1}{2a_g}\left(\frac{1}{n_g}\sup_{\mathbf{u}\in\mathcal{B}_0}\|\mathbf{X}_g\mathbf{u}\|_2^2 + \frac{1}{n_g}\sup_{\mathbf{v}\in\mathcal{B}_g}\|\mathbf{X}_g\mathbf{v}\|_2^2\right) \\
(\text{Lemma SM1.3}) &\leq \frac{1}{a_g}\left(1 + c_{0g}\frac{\omega(\mathcal{A}_g)+\omega(\mathcal{A}_0)+2\tau}{2\sqrt{n_g}}\right) \\
(\omega_{0g} = \max(\omega(\mathcal{A}_0),\omega(\mathcal{A}_g))) &\leq \frac{1}{a_g}\left(1 + c_{0g}\frac{\omega_{0g}+2\tau}{\sqrt{n_g}}\right)
\end{aligned}$$

where $c_{0g} = \max(c_0, c_g)$.

To compute the exact probabilities lets define $s_g := \frac{1}{a_g}\left(1 + c_{0g}\frac{\omega(\mathcal{A}_g)+\omega(\mathcal{A}_0)+2\tau}{2\sqrt{n_g}}\right)$ and name the bad events of $\frac{1}{n_g}\sup_{\mathbf{u}\in\mathcal{B}_0}\|\mathbf{X}_g\mathbf{u}\|_2^2 > 1 + c_0\frac{\omega(\mathcal{A}_0)+\tau}{\sqrt{n_g}}$ and $\frac{1}{n_g}\sup_{\mathbf{v}\in\mathcal{B}_g}\|\mathbf{X}_g\mathbf{v}\|_2^2 > 1 + c_g\frac{\omega(\mathcal{A}_g)+\tau}{\sqrt{n_g}}$ as $\mathcal{E}_1$ and $\mathcal{E}_2$ respectively. Then from the law of total probability we have:

$$\begin{aligned}
\mathbb{P}(\phi_g(\mu_g) > s_g) &\leq \mathbb{P}(\phi_g(\mu_g) > s_g|\neg\mathcal{E}_1, \neg\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_1, \mathcal{E}_2) \\
&\leq 0 + \mathbb{P}(\mathcal{E}_1|\mathcal{E}_2)\mathbb{P}(\mathcal{E}_2) \leq \mathbb{P}(\mathcal{E}_2) \\
(\text{Lemma SM1.3}) &\leq 2\exp\left(-\gamma_g(\omega(\mathcal{A}_g)+\tau)^2\right)
\end{aligned}$$

which concludes the proof. ∎

SM1.8. Proof of Lemma SM1.1.

Proof. To obtain lower bound, we use the Paley–Zygmund inequality for the zero-mean, non-degenerate ($0 < \alpha \leq \mathbb{E}|\langle \mathbf{x}, \mathbf{u}\rangle|, \mathbf{u} \in \mathbb{S}^{p-1}$) sub-Gaussian random vector $\mathbf{x}$ with $|||\mathbf{x}|||_{\psi_2} \leq k_x$ [**?**].

$$Q_{2\xi}(\mathbf{u}) \geq \frac{(\alpha - 2\xi)^2}{4ck_x^2}.$$

SM1.9. Proof of Lemma SM1.2.

Proof. We split $[G] - \mathcal{I}$ into two groups $\mathcal{J}, \mathcal{K}$. $\mathcal{J}$ consists of $\boldsymbol{\delta}_g$'s with $\|\boldsymbol{\delta}_g\|_2 \geq 2\|\boldsymbol{\delta}_0\|_2$ and $\mathcal{K} = [G] - \mathcal{I} - \mathcal{J}$. We use the bounds

$$\|\boldsymbol{\delta}_0 + \boldsymbol{\delta}_g\|_2 \geq \begin{cases} \lambda_{\min}(\|\boldsymbol{\delta}_g\|_2 + \|\boldsymbol{\delta}_0\|_2) & \text{if } g \in \mathcal{I} \\ \|\boldsymbol{\delta}_g\|_2/2 & \text{if } g \in \mathcal{J} \\ 0 & \text{if } g \in \mathcal{K} \end{cases}$$

This implies

$$\sum_{g=1}^{G} n_g\|\boldsymbol{\delta}_0 + \boldsymbol{\delta}_g\|_2 \geq \sum_{g\in\mathcal{J}} \frac{n_g}{2}\|\boldsymbol{\delta}_g\|_2 + \lambda_{\min}\sum_{g\in\mathcal{I}} n_g(\|\boldsymbol{\delta}_g\|_2 + \|\boldsymbol{\delta}_0\|_2).$$

∎

Let $S_{\mathcal{S}} = \sum_{g\in\mathcal{S}} n_g\|\boldsymbol{\delta}_g\|_2$ for $\mathcal{S} = \mathcal{I}, \mathcal{J}, \mathcal{K}$. We know that over $\mathcal{K}$, $\|\boldsymbol{\delta}_g\|_2 \leq 2\|\boldsymbol{\delta}_0\|_2$ which implies $S_{\mathcal{K}} = \sum_{g\in\mathcal{K}} n_g\|\boldsymbol{\delta}_g\|_2 \leq 2\sum_{g\in\mathcal{K}} n_g\|\boldsymbol{\delta}_0\|_2 \leq 2n\|\boldsymbol{\delta}_0\|_2$. Set $\psi_{\mathcal{I}} = \min\{1/2, \lambda_{\min}\bar{\rho}/3\}$. Using $1/2 \geq \psi_{\mathcal{I}}$, we write:

$$\sum_{g=1}^{G} n_g\|\boldsymbol{\delta}_0 + \boldsymbol{\delta}_g\|_2 \geq \psi_{\mathcal{I}} S_{\mathcal{J}} + \lambda_{\min}\sum_{g\in\mathcal{I}} n_g(\|\boldsymbol{\delta}_g\|_2 + \|\boldsymbol{\delta}_0\|_2)$$

$$(S_{\mathcal{K}} \leq 2n\|\boldsymbol{\delta}_0\|_2) \geq \psi_{\mathcal{I}} S_{\mathcal{J}} + \psi_{\mathcal{I}} S_{\mathcal{K}} - 2\psi_{\mathcal{I}} n\|\boldsymbol{\delta}_0\|_2 + \left(\sum_{g\in\mathcal{I}} n_g\right)\lambda_{\min}\|\boldsymbol{\delta}_0\|_2 + \lambda_{\min} S_{\mathcal{I}}$$

$$(\lambda_{\min} \geq \psi_{\mathcal{I}}) \geq \psi_{\mathcal{I}}(S_{\mathcal{I}} + S_{\mathcal{J}} + S_{\mathcal{K}}) + \left(\left(\sum_{g\in\mathcal{I}} n_g\right)\lambda_{\min} - 2\psi_{\mathcal{I}} n\right)\|\boldsymbol{\delta}_0\|_2.$$

Now, observe that, assumption of the Definition 3.4, $\sum_{g\in\mathcal{I}} n_g \geq \bar{\rho} n$ implies:

$$\left(\sum_{g\in\mathcal{I}} n_g\right)\lambda_{\min} - 2\psi_{\mathcal{I}} n \geq (\bar{\rho}\lambda_{\min} - 2\psi_{\mathcal{I}})n \geq \psi_{\mathcal{I}} n.$$

Combining all, we obtain:

$$\sum_{g=1}^{G} n_g\|\boldsymbol{\delta}_0 + \boldsymbol{\delta}_g\|_2 \geq \psi_{\mathcal{I}}(S_{\mathcal{I}} + S_{\mathcal{J}} + S_{\mathcal{K}} + \|\boldsymbol{\delta}_0\|_2) = \psi_{\mathcal{I}}(n\|\boldsymbol{\delta}_0\|_2 + \sum_{g=1}^{G} n_g\|\boldsymbol{\delta}_g\|_2).$$