A Proof of Lemma 1

Lemma 1 *For $\forall j \in S^*$, let $\mathbb{E}(R_j) > \frac{1}{2}$. With probability at least $1 - 2\exp(-2g(-\frac{1}{2} + \mathbb{E}(R_j))^2 + \log s)$, simultaneously $\forall j \in S^*$,we have $R_j \geqslant \frac{1}{2}$.*

Proof. Note that since w^* is fixed, for any $j \in [d]$, R_{ij} across $i \in [g]$ are independent of each other.

$$\begin{aligned} \mathbb{P}(R_j < \frac{1}{2}) &= \mathbb{P}(R_j - \mathbb{E}(R_j) < \frac{1}{2} - \mathbb{E}(R_j)) \\ &\leqslant \mathbb{P}(|R_j - \mathbb{E}(R_j)| > -\frac{1}{2} + \mathbb{E}(R_j)) \end{aligned} \tag{12}$$

The last inequality holds as long as $\mathbb{E}(R_j) > \frac{1}{2}$. We also note that for a fixed w^*, $R_j = f(R_{1j}, R_{2j}, \cdots, R_{gj})$ such that

$$(\forall k \in [g]), \quad \sup_{R_{1j}, \cdots, R_{kj}, R'_{kj}, \cdots, R_{gj}} |f(R_{1j}, \cdots, R_{kj}, \cdots, R_{gj}) - f(R_{1j}, \cdots, R'_{kj}, \cdots, R_{gj})| \leqslant \frac{1}{g} \tag{13}$$

Thus using McDiarmid's inequality McDiarmid (1989), we can write the following:

$$\mathbb{P}(|R_j - \mathbb{E}(R_j)| > -\frac{1}{2} + \mathbb{E}(R_j)) \leqslant 2\exp(-2g(-\frac{1}{2} + \mathbb{E}(R_j))^2) \tag{14}$$

Taking the union bound across $j \in S^*$, we get the desired result. ☐

B Proof of Lemma 2

We start by deriving some technical lemmas that will help us obtain our desired result.

Lemma 7. *For $0 \leqslant \delta_j \leqslant 8|w_j^*|\rho_i^2, \forall i \in [g], \forall j \in S^*$,*

$$\mathbb{P}(|w_j^* \frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| \geqslant \delta_j) \leqslant 2\exp(-\frac{n_i(\frac{\delta_j}{|w_j^*|\rho_i^2})^2}{64}) \tag{15}$$

Proof. Observe that,

$$\begin{aligned} \mathbb{P}(|w_j^* \frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| \geqslant \delta_j) &= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| \geqslant \frac{\delta_j}{|w_j^*|}) \\ &= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} ((\frac{X_{ij}^t}{\rho_i})^2 - (\frac{\sigma_{jj}^i}{\rho_i})^2)| \geqslant \frac{\delta_j}{|w_j^*|\rho_i^2}) \end{aligned} \tag{16}$$

Note that $\frac{X_{ij}}{\rho_i}$ is a zero mean sub-Gaussian random variable with a variance proxy 1. This implies that $(\frac{X_{ij}}{\rho_i})^2$ is a sub-exponential random variable with parameters $(4\sqrt{2}, 4)$. Thus, using concentration bounds for sub-exponential random variablesWainwright (2015), we can write:

$$\mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} ((\frac{X_{ij}^t}{\rho_i})^2 - (\frac{\sigma_{jj}^i}{\rho_i})^2)| \geqslant \frac{\delta_j}{|w_j^*|\rho_i^2}) \leqslant 2\exp(-\frac{n_i(\frac{\delta_j}{|w_j^*|\rho_i^2})^2}{64}), \quad \forall 0 \leqslant \frac{\delta_j}{|w_j^*|\rho_i^2} \leqslant 8 \tag{17}$$

☐

Lemma 8. *For $0 \leqslant \delta_1 \leqslant 8\rho_i^2 \sqrt{\sum_{k \in S^*, k \neq j} {w_k^*}^2}, \forall i \in [g], \forall j \in [d]$, if predictors are mutually independent of each other then*

$$\mathbb{P}(|\sum_{k \in S^*, k \neq j} w_k^* \frac{1}{n_i} \sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \geqslant \delta_1) \leqslant 2\exp(-\frac{n_i(\frac{\delta_1}{\rho_i^2\sqrt{\sum_{k \in S^*, k \neq j} {w_k^*}^2}})^2}{64}) \tag{18}$$

Proof. Note that,

$$\sum_{k\in S*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t = \frac{1}{n_i}\sum_{t=1}^{n_i}(\sum_{k\in S*,k\neq j} w_k^* X_{ik}^t) X_{ij}^t \tag{19}$$

Here X_{ik}^t is a zero mean sub-Gaussian random variable with variance proxy ρ_i and $(\sum_{k\in S*,k\neq j} w_k^* X_{ik}^t)$ is a zero mean sub-Gaussian random variable with variance proxy $\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}\rho_i$. Thus,

$$\begin{aligned}\mathbb{P}(|\sum_{k\in S*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \geqslant \delta_1) &= \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(\sum_{k\in S*,k\neq j} w_k^* X_{ik}^t) X_{ij}^t| \geqslant \delta_1) \\ &= \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} \frac{(\sum_{k\in S*,k\neq j} w_k^* X_{ik}^t)}{\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}\rho_i}\frac{X_{ij}^t}{\rho_i}| \geqslant \frac{\delta_1}{\rho_i^2\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}})\end{aligned} \tag{20}$$

where $\frac{(\sum_{k\in S*,k\neq j} w_k^* X_{ik}^t)}{\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}\rho_i}$ and $\frac{X_{ij}^t}{\rho_i}$ are zero mean sub-Gaussian random variables with unit variance proxy. Thus their product is a sub-exponential random variable with parameters $(4\sqrt{2}, 4)$. Using concentration bound for sub-exponential random variables Wainwright (2015), we can write:

$$\begin{aligned}&\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} \frac{(\sum_{k\in S*,k\neq j} w_k^* X_{ik}^t)}{\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}\rho_i}\frac{X_{ij}^t}{\rho_i}| \geqslant \frac{\delta_1}{\rho_i^2\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}}) \leqslant 2\exp(-\frac{n_i(\frac{\delta_1}{\rho_i^2\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}})^2}{64}) \\ &\forall 0 \leqslant \frac{\delta_1}{\rho_i^2\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2}} \leqslant 8\end{aligned} \tag{21}$$

□

Lemma 9. *For* $0 \leqslant \delta_e \leqslant 8|\eta_i\rho_i|, \forall i \in [g], \forall j \in [d]$,

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e) \leqslant 2\exp(-\frac{n_i(\frac{\delta_e}{|\eta_i\rho_i|})^2}{64}) \tag{22}$$

Proof. Note that,

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e) = \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} \frac{e_i^t}{\eta_i}\frac{X_{ij}^t}{\rho_i}| \geqslant \frac{\delta_e}{|\eta_i\rho_i|}) \tag{23}$$

Again $\frac{X_{ij}}{\rho_i}$ and $\frac{e_i}{\eta_i}$ are mutually independent zero mean sub-Gaussian random variables with a variance proxy 1. Thus, $\frac{e_i^t}{\eta_i}\frac{X_{ij}^t}{\rho_i}$ is a sub-exponential random variable with parameters $(4\sqrt{2}, 4)$. Thus, using concentration bounds for sub-exponential random variablesWainwright (2015), we can write:

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} \frac{e_i^t}{\eta_i}\frac{X_{ij}^t}{\rho_i}| \geqslant \frac{\delta_e}{|\eta_i\rho_i|}) \leqslant 2\exp(-\frac{n_i(\frac{\delta_e}{|\eta_i\rho_i|})^2}{64}), \quad \forall 0 \leqslant \frac{\delta_e}{|\eta_i\rho_i|} \leqslant 8 \tag{24}$$

□

Armed with the previous technical lemmas, we provide our desired result.

Lemma 2 *For $i \in [g]$, $j \in S^*$ and some $0 < \delta \leqslant 1$, if predictors are mutually independent of each other and $0 < \lambda_{ij} < |w_j^* (\sigma_{jj}^i)^2| - 8|w_j^*|\rho_i^2 \delta - 8\rho_i^2 \sqrt{\sum_{k \in S^*, k \neq j} {w_k^*}^2} \delta - 8|\eta_i \rho_i| \delta$ then we have $\mathbb{E}(R_j) \geqslant 1 - \frac{6}{g} \sum_{i=1}^{g} \exp(-n_i \delta^2)$.*

Proof. By using sum of expectations, we can write $\mathbb{E}(R_j)$ in the following form:

$$
\begin{aligned}
\mathbb{E}(R_j) &= \frac{1}{g} \sum_{i=1}^{g} \mathbb{E}(R_{ij}) \\
&= \frac{1}{g} \sum_{i=1}^{g} \mathbb{P}(\hat{w}_{ij} \neq 0) \\
&= \frac{1}{g} \sum_{i=1}^{g} \mathbb{P}(\frac{1}{\hat{\sigma}_{ij}} \operatorname{sign}(\hat{\alpha}_{ij}) \max(0, |\hat{\alpha}_{ij}| - \lambda_{ij}) \neq 0) \\
&= \frac{1}{g} \sum_{i=1}^{g} \mathbb{P}(\operatorname{sign}(\hat{\alpha}_{ij}) \neq 0 \wedge |\hat{\alpha}_{ij}| > \lambda_{ij}) \\
&= \frac{1}{g} \sum_{i=1}^{g} \mathbb{P}(\hat{\alpha}_{ij} \neq 0 \mid |\hat{\alpha}_{ij}| > \lambda_{ij}) \mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) \\
&= \frac{1}{g} \sum_{i=1}^{g} \mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij})
\end{aligned} \tag{25}
$$

The last equality follows since for $\lambda_{ij} > 0$, we have $\mathbb{P}(\hat{\alpha}_{ij} \neq 0 \mid |\hat{\alpha}_{ij}| > \lambda_{ij}) = 1$. Now, we put a bound on the term $\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij})$.

$$
\begin{aligned}
\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) &= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} y_i^t X_{ij}^t| > \lambda_{ij}) \\
&\text{Expanding } y_i^t \text{ using equation (1)} \\
&= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} ((X_i^t)^\intercal w^* + e_i^t) X_{ij}^t| > \lambda_{ij}) \\
&= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} (\sum_{k=1}^{d} X_{ik}^t w_k^* + e_i^t) X_{ij}^t| > \lambda_{ij}) \\
&= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 w_j^* + \sum_{k \in S^*, k \neq j} X_{ik}^t X_{ij}^t w_k^* + e_i^t X_{ij}^t)| > \lambda_{ij}) \\
&= \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} (X_{ij}^t)^2 w_j^* + \sum_{k \in S^*, k \neq j} \frac{1}{n_i} \sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t w_k^* + \frac{1}{n_i} \sum_{t=1}^{n_i} e_i^t X_{ij}^t| > \lambda_{ij})
\end{aligned} \tag{26}
$$

Recall that in this setting the covariance matrix for X_i is Σ^i with diagonal entries $\Sigma_{jj}^i \equiv (\sigma_{jj}^i)^2, \forall j \in [d]$ and non-diagonal entries $\Sigma_{jk}^i \equiv 0, \forall j, k \in [d], j \neq k$. Let $D_{ij} \triangleq w_j^* (\sigma_{jj}^i)^2$. Then,

$$
\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) = \mathbb{P}(|w_j^* \frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - (\sigma_{jj}^i)^2) + \sum_{k \in S^*, k \neq j} w_k^* \frac{1}{n_i} \sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t + \frac{1}{n_i} \sum_{t=1}^{n_i} e_i^t X_{ij}^t + D_{ij}| > \lambda_{ij}) \tag{27}
$$

Using the reverse triangle inequality $|a+b| \geqslant |a| - |b|$ recursively,

$$\geqslant \mathbb{P}(|D_{ij}| - |w_j^* \frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| - \sum_{k \in S*, k \neq j} |w_k^* \frac{1}{n_i} \sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| - |\frac{1}{n_i} \sum_{t=1}^{n_i} e_i^t X_{ij}^t| > \lambda_{ij})$$

$$\geqslant \mathbb{P}(|D_{ij}| > \lambda_{ij} + \delta_j + \delta_1 + \delta_e \wedge |w_j^* \frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| < \delta_j \wedge |\sum_{k \in S*, k \neq j} w_k^* \frac{1}{n_i} \sum_{t=1}^{n_i} (X_{ik}^t X_{ij}^t)|$$

$$< \delta_1 \wedge |\frac{1}{n_i} \sum_{t=1}^{n_i} e_i^t X_{ij}^t| < \delta_e) \tag{28}$$

$$\geqslant 1 - \mathbb{P}(|D_{ij}| \leqslant \lambda_{ij} + \delta_j + \delta_1 + \delta_e) - \mathbb{P}(|w_j^* \frac{1}{n_i} \sum_{t=1}^{n_i} ((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| \geqslant \delta_j) -$$

$$\mathbb{P}(|\sum_{k \in S*, k \neq j} w_k^* \frac{1}{n_i} \sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \geqslant \delta_1) - \mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e)$$

We take $\delta_j = 8|w_j^*|\rho_i^2 \delta$ in Lemma 7, then

$$\mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} ((\frac{X_{ij}^t}{\rho_i})^2 - (\frac{\sigma_{jj}^i}{\rho_i})^2)| \geqslant \frac{\delta_j}{|w_j^*|\rho_i^2}) \leqslant 2\exp(-n_i \delta^2), \quad \forall 0 \leqslant \delta \leqslant 1 \tag{29}$$

We take $\delta_1 = 8\delta\rho_i^2 \sqrt{\sum_{k \in S*, k \neq j} {w_k^*}^2}$ in Lemma 8, which gives us

$$\mathbb{P}(|\sum_{k \in S*, k \neq j} w_k^* \frac{1}{n_i} \sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \geqslant \delta_1) \leqslant 2\exp(-n_i \delta^2), \quad \forall 0 \leqslant \delta \leqslant 1 \tag{30}$$

We take $\delta_e = 8|\eta_i \rho_i|\delta$ in Lemma 9, then

$$\mathbb{P}(|\frac{1}{n_i} \sum_{t=1}^{n_i} \frac{e_i^t}{\eta_i} \frac{X_{ij}^t}{\rho_i}| \geqslant \frac{\delta_e}{|\eta_i \rho_i|}) \leqslant 2\exp(-n_i \delta^2), \quad \forall 0 \leqslant \delta \leqslant 1 \tag{31}$$

By making the appropriate substitutions for δ_j, δ_1 and δ_e and noting that $|w_j^* {(\sigma_{jj}^i)}^2| > \lambda_{ij} + 8|w_j^*|\rho_i^2 \delta + 8\rho_i^2 \sqrt{\sum_{k \in S*, k \neq j} {w_k^*}^2} \delta + 8|\eta_i \rho_i|\delta$, we can write:

$$\mathbb{E}(R_j) \geqslant 1 - \frac{6}{g} \sum_{i=1}^{g} \exp(-n_i \delta^2) \tag{32}$$

□

C Proof of Lemma 3

Lemma 3 *For $\forall j \in S_c^*$, let $\mathbb{E}(R_j) < \frac{1}{2}$. With probability at least $1 - 2\exp(-2g(\frac{1}{2} - \mathbb{E}(R_j))^2 + \log(d-s))$, simultaneously $\forall j \in S_c^*$, we have $R_j \leqslant \frac{1}{2}$.*

Proof. We again note that,

$$\begin{aligned} \mathbb{P}(R_j \geqslant \frac{1}{2}) &= \mathbb{P}(R_j - \mathbb{E}(R_j) \geqslant \frac{1}{2} - \mathbb{E}(R_j)) \\ &\leqslant \mathbb{P}(|R_j - \mathbb{E}(R_j)| \geqslant \frac{1}{2} - \mathbb{E}(R_j)) \end{aligned} \tag{33}$$

The last inequality holds as long as $\mathbb{E}(R_j) < \frac{1}{2}$. Again, by noting that for a fixed w^*, $R_j = f(R_{1j}, R_{2j}, \cdots, R_{gj})$ such that

$$(\forall k \in [g]), \quad \sup_{R_{1j},\cdots,R_{kj},R'_{kj},\cdots,R_{gj}} |f(R_{1j},\cdots,R_{kj},\cdots,R_{gj}) - f(R_{1j},\cdots,R'_{kj},\cdots,R_{gj})| \leqslant \frac{1}{g} \tag{34}$$

and using McDiarmid's inequality McDiarmid (1989), we can write the following:

$$\mathbb{P}(|R_j - \mathbb{E}(R_j)| \geqslant \frac{1}{2} - \mathbb{E}(R_j)) \leqslant 2\exp(-2g(\frac{1}{2} - \mathbb{E}(R_j))^2) \tag{35}$$

Taking a union bound across $j \in S_c^*$, we get the desired result. □

D Proof of Lemma 4

Lemma 4 *For $i \in [g]$, $j \in S_c^*$ and $0 < \delta \leqslant 1$, if predictors are mutually independent of each other and if $\lambda_{ij} > 8\delta\rho_i^2\sqrt{\sum_{k\in S^*} w_k^2} + 8|\eta_i\rho_i|\delta$ then we have $\mathbb{E}(R_j) \leqslant \frac{4}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2)$.*

Proof. Like the proof of Lemma 2, we have the same formula for $\mathbb{E}(R_j)$ with the difference that $j \in S_c^*$, i.e.,

$$\mathbb{E}(R_j) = \frac{1}{g}\sum_{i=1}^{g}\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) \tag{36}$$

This time, we will put an upper bound on $\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij})$.

$$\begin{aligned}
\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) &= \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} y_i^t X_{ij}^t| > \lambda_{ij}) \\
&\text{Expanding } y_i^t \text{ using equation (1)} \\
&= \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}((X_i^t)^\intercal w^* + e_i^t)X_{ij}^t| > \lambda_{ij}) \\
&= \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(\sum_{k=1}^{d} X_{ik}^t w_k^* + e_i^t)X_{ij}^t| > \lambda_{ij}) \\
&= \mathbb{P}(|\frac{1}{n_i}(\sum_{k\in S^*} X_{ik}^t X_{ij}^t w_k^* + e_i^t X_{ij}^t)| > \lambda_{ij}) \\
&= \mathbb{P}(|\sum_{k\in S^*}\frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t w_k^* + \frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| > \lambda_{ij})
\end{aligned} \tag{37}$$

Using the triangle inequality $|a+b| \leqslant |a| + |b|$, we can rewrite the above equation as:

$$\begin{aligned}
\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) &\leqslant \mathbb{P}(|\sum_{k\in S^*} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| + |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| > \lambda_{ij}) \\
&= 1 - \mathbb{P}(|\sum_{k\in S^*} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| + |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \leqslant \lambda_{ij}) \\
&\leqslant 1 - \mathbb{P}(|\sum_{k\in S^*} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \leqslant \delta_1 \wedge |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \leqslant \delta_e \wedge 0 \leqslant \lambda_{ij} - \delta_1 - \delta_e) \\
&\leqslant \mathbb{P}(|\sum_{k\in S^*} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \geqslant \delta_1) + \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e) + \mathbb{P}(0 \geqslant \lambda_{ij} - \delta_1 - \delta_e)
\end{aligned} \tag{38}$$

We take $\delta_1 = 8\rho_i^2\sqrt{\sum_{k\in S^*} {w_k^*}^2}\delta$ for $0 < \delta < 1$ in Lemma 8. Then,

$$\mathbb{P}(|\sum_{k\in S^*} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i} X_{ik}^t X_{ij}^t| \geqslant \delta_1) \leqslant 2\exp(-n_i\delta^2) \tag{39}$$

We take $\delta_e = 8|\eta_i\rho_i|\delta$ for $0 < \delta < 1$ in Lemma 9, then

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e) \leqslant 2\exp(-n_i\delta^2) \tag{40}$$

Using results of Lemmas 8 and 9 and making the appropriate substitutions for δ_1 and δ_e and noticing that $\lambda_{ij} > 8\delta\rho_i^2\sqrt{\sum_{k\in S^*} w_k^2} + 8|\eta_i\rho_i|\delta$, we can write

$$\mathbb{E}(R_j) \leqslant \frac{4}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2) \tag{41}$$

□

E Proof of Lemma 5

We start by deriving a technical lemma that will help us obtain our desired result.

Lemma 10. *For* $0 \leqslant \delta_k \leqslant 8\sqrt{2}|w_k^*|(1+4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})\max_j (\sigma_{jj}^i)^2, \forall i \in [g], \forall j \in [d]$, *then*

$$\mathbb{P}(|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_k) \leqslant 4\exp(-\frac{n_i(\frac{\delta_k}{|w_k^*|})^2}{128(1+4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})^2 \max_j {\sigma_{jj}^i}^4}) \tag{42}$$

Proof. Note that,

$$\mathbb{P}(|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_k) = \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \frac{\delta_k}{|w_k^*|}) \tag{43}$$

Using Lemma 1 from Ravikumar et al. (2011), we can write

$$\begin{aligned}
&\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \frac{\delta_k}{|w_k^*|}) \leqslant 4\exp(-\frac{n_i(\frac{\delta_k}{|w_k^*|})^2}{128(1+4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})^2 \max_j {\sigma_{jj}^i}^4}),\\
&\forall 0 \leqslant \frac{\delta_k}{|w_k^*|} \leqslant 8\max_j (\sigma_{jj}^i)^2(1+4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})
\end{aligned} \tag{44}$$

□

Armed with the previous technical lemma, we provide our desired result.

Lemma 5 *For* $i \in [g]$, $j \in S^*$ *and some* $0 < \delta \leqslant \frac{1}{\sqrt{2}}$, *if* $0 < \lambda_{ij} < |(w_j^*(\sigma_{jj}^i)^2 + \sum_{k\in S^*, k\neq j} w_k^*\sigma_{jk}^i)| - 8|w_j^*|\rho_i^2\delta - \sum_{k\in S^*,k\neq j} 8\sqrt{2}|w_k^*|(1+4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})\max_j (\sigma_{jj}^i)^2\delta - 8|\eta_i\rho_i|\delta$ *then we have* $\mathbb{E}(R_j) \geqslant 1 - \frac{4s}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2)$.

Proof. We know from the proof of Lemma 2 that

$$\mathbb{E}(R_j) = \frac{1}{g}\sum_{i=1}^{g}\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) \tag{45}$$

and

$$\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda) = \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ij}^t)^2 w_j^* + \sum_{k\in S*,k\neq j}\frac{1}{n_i}\sum_{t=1}^{n_i}X_{ik}^t X_{ij}^t w_k^* + \frac{1}{n_i}\sum_{t=1}^{n_i}e_i^t X_{ij}^t| > \lambda_{ij}) \tag{46}$$

Let $D_{ij} = w_j^*\mathbb{E}(X_{ij}^2) + \sum_{k\in S*,k\neq j} w_k^*\mathbb{E}(X_{ik}X_{ij}) = w_j^*{(\sigma_{jj}^i)}^2 + \sum_{k\in S*,k\neq j} w_k^*\sigma_{jk}^i$. Then,

$$\begin{aligned}\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) = \mathbb{P}(|w_j^*\frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2) + \sum_{k\in S*,k\neq j} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)\\ + \frac{1}{n_i}\sum_{t=1}^{n_i}e_i^t X_{ij}^t + D_{ij}| > \lambda_{ij})\end{aligned} \tag{47}$$

Using the reverse triangle inequality $|a+b| \geqslant |a| - |b|$ recursively,

$$\begin{aligned}&\geqslant \mathbb{P}(|D_{ij}| - |w_j^*\frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| - \sum_{k\in S*,k\neq j}|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| - |\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^t X_{ij}^t| > \lambda_{ij})\\ &\geqslant \mathbb{P}(|D_{ij}| > \lambda_{ij} + \delta_j + \sum_{k\in S*,k\neq j}\delta_k + \delta_e \wedge |w_j^*\frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| < \delta_j \wedge (\forall k\in S^*, k\neq j)\\ &|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| < \delta_k \wedge |\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^t X_{ij}^t| < \delta_e)\\ &\geqslant 1 - \mathbb{P}(|D_{ij}| \leqslant \lambda_{ij} + \delta_j + \sum_{k\in S*,k\neq j}\delta_k + \delta_e) - \mathbb{P}(|w_j^*\frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - {(\sigma_{jj}^i)}^2)| \geqslant \delta_j) -\\ &\sum_{k\in S*,k\neq j}\mathbb{P}(|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_k) - \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^t X_{ij}^t| \geqslant \delta_e)\end{aligned} \tag{48}$$

We take $\delta_k = 8\sqrt{2}|w_k^*|(1 + 4\max_j\frac{\rho_i^2}{{(\sigma_{jj}^i)}^2})\max_j{(\sigma_{jj}^i)}^2\delta$ in Lemma 10, then

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \frac{\delta_k}{|w_k^*|}) \leqslant 4\exp(-n_i\delta^2), \forall 0 \leqslant \delta \leqslant \frac{1}{\sqrt{2}} \tag{49}$$

Also taking $\delta_j = 8|w_j^*|\rho_i^2\delta$ and $\delta_e = 8|\eta_i\rho_i|\delta$ in Lemmas 7 and 9 respectively and noting that

$$\begin{aligned}0 < \lambda_{ij} < |(w_j^*{(\sigma_{jj}^i)}^2 + \sum_{k\in S*,k\neq j} w_k^*\sigma_{jk}^i)| - 8|w_j^*|\rho_i^2\delta - \sum_{k\in S*,k\neq j} 8\sqrt{2}|w_k^*|(1 + 4\max_j\frac{\rho_i^2}{{(\sigma_{jj}^i)}^2})\\ \max_j{(\sigma_{jj}^i)}^2\delta - 8|\eta_i\rho_i|\delta\ ,\end{aligned} \tag{50}$$

we can write:

$$\mathbb{E}(R_j) \geqslant 1 - \frac{4s}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2) \tag{51}$$

It follows that if we take $n_i > \frac{1}{\delta^2}\log 8s$, we get $\mathbb{E}(R_j) > \frac{1}{2}$. □

F Proof of Lemma 6

Lemma 6 *For $i \in [g]$, $j \in S_c^*$ and some $0 < \delta \leqslant \frac{1}{\sqrt{2}}$, if $\lambda_{ij} > |\sum_{k\in S^*} w_k^* \sigma_{jk}^i| + \sum_{k\in S^*} 8\sqrt{2}|w_k^*|(1 + 4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})\max_j {(\sigma_{jj}^i)}^2\delta + 8|\eta_i\rho_i|\delta$ then we have $\mathbb{E}(R_j) \leqslant \frac{4s+2}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2)$.*

Proof. Like before,

$$\mathbb{E}(R_j) = \frac{1}{g}\sum_{i=1}^{g}\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) \tag{52}$$

and

$$\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) = \mathbb{P}(|\sum_{k\in S^*}\frac{1}{n_i}\sum_{t=1}^{n_i}X_{ik}^tX_{ij}^tw_k^* + \frac{1}{n_i}\sum_{t=1}^{n_i}e_i^tX_{ij}^t| > \lambda_{ij}) \tag{53}$$

Let $D_{ij} = \sum_{k\in S^*} w_k^*\mathbb{E}(X_{ik}X_{ij}) = \sum_{k\in S^*} w_k^*\sigma_{jk}^i$. Then,

$$\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) = \mathbb{P}(|\sum_{k\in S^*}w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i) + \frac{1}{n_i}\sum_{t=1}^{n_i}e_i^tX_{ij}^t + D_{ij}| > \lambda_{ij}) \tag{54}$$

Using the triangle inequality $|a+b| \leqslant |a| + |b|$,

$$\begin{aligned}
\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) &\leqslant \mathbb{P}(\sum_{k\in S^*}|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| + |\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^tX_{ij}^t| + |D_{ij}| > \lambda_{ij}) \\
&= 1 - \mathbb{P}(\sum_{k\in S^*}|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| + |\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^tX_{ij}^t| + |D_{ij}| \leqslant \lambda_{ij}) \\
&\leqslant 1 - \mathbb{P}((\forall k \in S^*)|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| \leqslant \delta_k \wedge |\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^tX_{ij}^t| \leqslant \delta_e \wedge \\
&\quad |D_{ij}| \leqslant \lambda_{ij} - \sum_{k\in S^*}\delta_k - \delta_e) \\
&\leqslant \sum_{k\in S^*}\mathbb{P}(|w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_k) + \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}e_i^tX_{ij}^t| \geqslant \delta_e) + \\
&\quad \mathbb{P}(|D_{ij}| \geqslant \lambda_{ij} - \sum_{k\in S^*}\delta_k - \delta_e)
\end{aligned} \tag{55}$$

We take $\delta_k = 8\sqrt{2}|w_k^*|(1 + 4\max_j \frac{\rho_i^2}{(\sigma_{jj}^i)^2})\max_j {(\sigma_{jj}^i)}^2\delta$ in Lemma 10, then

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| \geqslant \frac{\delta_k}{|w_k^*|}) \leqslant 4\exp(-n_i\delta^2), \forall 0 \leqslant \delta \leqslant \frac{1}{\sqrt{2}} \tag{56}$$

Taking $\delta_e = 8|\eta_i\rho_i|\delta$ in Lemma 9 and noticing that

$$\lambda_{ij} > |\sum_{k\in S^*}w_k^*\sigma_{jk}^i| + \sum_{k\in S^*}8\sqrt{2}|w_k^*|(1 + 4\max_j\frac{\rho_i^2}{{(\sigma_{jj}^i)}^2})\max_j{(\sigma_{jj}^i)}^2\delta + 8|\eta_i\rho_i|\delta\ , \tag{57}$$

we can write

$$\mathbb{E}(R_j) \leqslant \frac{4s+2}{g}\sum_{i=1}^{g}(\exp(-n_i\delta^2)) \tag{58}$$

$\square$

G Federated Sparse Regression With Correlated Gaussian Predictors

If predictors are correlated Gaussian random variables then our method works with an overall sample complexity of $\Omega(s \log d)$ which matches the well known result of Wainwright (2009b) for correlated Gaussian predictors in a centralized setting. Below we formally provide the proof of this special case.

Let $X_i \in \mathbb{R}^d$ be jointly Gaussian with mean $\mathbf{0} \in \mathbb{R}^d$ and covariance matrix $\Sigma^i \in \mathbb{R}^{d\times d}, \forall i \in [g]$. In this special case, we can achieve a tighter bound for Lemmas 5 and 6. In particular, we will prove Lemmas 11 and 12.

Lemma 11. *For $i \in [g]$, $j \in S^*$ and some $0 < \delta \leqslant \frac{1}{\sqrt{2}}$, if $0 < \lambda_{ij} < |w_j^*(\sigma_{jj}^i)^2 + \sum_{k\in S^*, k\neq j} w_k \sigma_{jk}^i| - 8|w_j^*|(\sigma_{jj}^i)^2\delta - 40\sqrt{2}\sigma_{jj}^i\sqrt{\sum_{k\in S^*,k\neq j} {w_k^*}^2(\sigma_{kk}^i)^2 + \sum_{k\in S^*}\sum_{l\in S^*,k\neq l} w_k^* w_l^* \sigma_{kl}^i}\,\delta - 8|\eta_i\sigma_{jj}^i|\delta$ then we have $\mathbb{E}(R_j) \geqslant 1 - \frac{8}{g}\sum_{i=1}^g \exp(-n_i\delta^2)$.*

Proof. We follow the proof of Lemma 5 until we reach the following step:

$$\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) = \mathbb{P}(|w_j^* \frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - (\sigma_{jj}^i)^2) + \sum_{k\in S^*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i) + \frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t + D_{ij}| > \lambda_{ij}) \tag{59}$$

Using the reverse triangle inequality $|a+b| \geqslant |a| - |b|$ recursively,

$$\begin{aligned}\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) &\geqslant \mathbb{P}(|D_{ij}| - |w_j^* \frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - (\sigma_{jj}^i)^2)| - |\sum_{k\in S^*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| - \\ &\quad |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| > \lambda_{ij}) \\ &\geqslant \mathbb{P}(|D_{ij}| > \lambda_{ij} + \delta_j + \delta_l + \delta_e \wedge |w_j^* \frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - (\sigma_{jj}^i)^2)| < \delta_j \wedge \\ &\quad |\sum_{k\in S^*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| < \delta_l \wedge |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| < \delta_e) \\ &\geqslant 1 - \mathbb{P}(|D_{ij}| \leqslant \lambda_{ij} + \delta_j + \delta_l + \delta_e) - \mathbb{P}(|w_j^* \frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - (\sigma_{jj}^i)^2)| \geqslant \delta_j) - \\ &\quad \mathbb{P}(|\sum_{k\in S^*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_l) - \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e)\end{aligned} \tag{60}$$

We already have good bounds for $\mathbb{P}(|w_j^* \frac{1}{n_i}\sum_{t=1}^{n_i}((X_{ij}^t)^2 - (\sigma_{jj}^i)^2)| \geqslant \delta_j)$ and $\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^t X_{ij}^t| \geqslant \delta_e)$ in Lemma 7 and Lemma 9 respectively. Below we provide a bound for the remaining term.

Let $y_i^t \triangleq \sum_{k\in S^*,k\neq j} w_k^* X_{ik}^t$ be a random variable. Then y_i^t is a normal random variable with 0 mean and variance $\tau^2 = \sum_{k\in S^*,k\neq j} {w_k^*}^2(\sigma_{kk}^i)^2 + \sum_{k\in S^*,k\neq j}\sum_{l\in S^*,k\neq j,l} w_k^* w_l^* \sigma_{kl}^i$. Then,

$$\mathbb{P}(|\sum_{k\in S^*,k\neq j} w_k^* \frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^t X_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_l) = \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(y_i^t X_{ij}^t - \sum_{k\in S^*,k\neq j} w_k^* \sigma_{jk}^i)| \geqslant \delta_l) \tag{61}$$

Note that $\frac{y_i^t}{\tau}$ and $\frac{X_{ij}^t}{\sigma_{jj}^i}$ are standard Gaussian variables(i.e., zero-mean, unit variance). Thus,

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(y_i^t X_{ij}^t - \sum_{k\in S^*,k\neq j} w_k^* \sigma_{jk}^i)| \geqslant \delta_l) = \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(\frac{y_i^t}{\tau}\frac{X_{ij}^t}{\sigma_{jj}^i} - \sum_{k\in S^*,k\neq j} w_k^* \frac{\sigma_{jk}^i}{\tau\sigma_{jj}^i})| \geqslant \frac{\delta_l}{\tau\sigma_{jj}^i}) \tag{62}$$

Using Lemma 1 from Ravikumar et al. (2011), we can write:

$$\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i}(\frac{y_i^t}{\tau}\frac{X_{ij}^t}{\sigma_{jj}^i} - \sum_{k\in S*,k\neq j} w_k^*\frac{\sigma_{jk}^i}{\tau\sigma_{jj}^i})| \geqslant \frac{\delta_l}{\tau\sigma_{jj}^i}) \leqslant 4\exp(-\frac{n_i\delta_l^2}{128\times 5^2\times (\sigma_{jj}^i)^2\tau^2}), \ \ \forall\delta_l \in (0, 40\tau\sigma_{jj}^i) \tag{63}$$

We take $\delta_l = 40\sqrt{2}\tau\sigma_{jj}^i\delta$ for $\delta \in (0, \frac{1}{\sqrt{2}})$, then

$$\mathbb{P}(|\sum_{k\in S*,k\neq j} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_l) \leqslant 4\exp(-n_i\delta^2), \ \ \forall\delta \in (0, \frac{1}{\sqrt{2}}) \tag{64}$$

Taking $\delta_j = 8|w_j^*|(\sigma_{jj}^i)^2\delta$ and $\delta_e = 8|\eta_i\sigma_{jj}^i|\delta$ in Lemmas 7 and 9 respectively and noting that

$$\begin{aligned} &0 < \lambda_{ij} < |w_j^*(\sigma_{jj}^i)^2 + \sum_{k\in S*,k\neq j} w_k\sigma_{jk}^i| - 8|w_j^*|(\sigma_{jj}^i)^2\delta - \\ &40\sqrt{2}\sigma_{jj}^i\sqrt{\sum_{k\in S*,k\neq j} {w_k^*}^2(\sigma_{kk}^i)^2 + \sum_{k\in S*,k\neq j}\sum_{l\in S*,k\neq j,l} w_k^*w_l^*\sigma_{kl}^i}\delta - 8|\eta_i\sigma_{jj}^i|\delta \end{aligned} \tag{65}$$

we can write:

$$\mathbb{E}(R_j) \geqslant 1 - \frac{8}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2) \tag{66}$$

□

Next, we provide the second main lemma to obtain tighter results for correlated Gaussian predictors.

Lemma 12. *For $i \in [g]$, $j \in S_c^*$ and some $0 < \delta \leqslant \frac{1}{\sqrt{2}}$, if $\lambda_{ij} > |\sum_{k\in S*} w_k^*\sigma_{jk}^i| + 40\sqrt{2}\sqrt{\sum_{k\in S*} {w_k^*}^2(\sigma_{kk}^i)^2 + \sum_{k\in S*}\sum_{l\in S*,k\neq l} w_k^*w_l^*\sigma_{kl}^i}\sigma_{jj}^i\delta + 8|\eta_i\sigma_{jj}^i|\delta$ then we have $\mathbb{E}(R_j) \leqslant \frac{6}{g}\sum_{i=1}^{g}\exp(-n_i\delta^2)$.*

Proof. We follow the proof of Lemma 6 until we reach the following step:

$$\mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) = \mathbb{P}(|\sum_{k\in S*} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i) + \frac{1}{n_i}\sum_{t=1}^{n_i} e_i^tX_{ij}^t + D_{ij}| > \lambda_{ij}) \tag{67}$$

Using the triangle inequality $|a+b| \leqslant |a| + |b|$,

$$\begin{aligned} \mathbb{P}(|\hat{\alpha}_{ij}| > \lambda_{ij}) &\leqslant \mathbb{P}(|\sum_{k\in S*} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| + |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^tX_{ij}^t| + |D_{ij}| > \lambda_{ij}) \\ &= 1 - \mathbb{P}(|\sum_{k\in S*} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| + |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^tX_{ij}^t| + |D_{ij}| \leqslant \lambda_{ij}) \\ &\leqslant 1 - \mathbb{P}(|\sum_{k\in S*} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| \leqslant \delta_l \wedge |\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^tX_{ij}^t| \leqslant \delta_e \wedge |D_{ij}| \leqslant \lambda_{ij} - \delta_l - \delta_e) \\ &\leqslant \mathbb{P}(|\sum_{k\in S*} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij}^t - \sigma_{jk}^i)| \geqslant \delta_l) + \mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^tX_{ij}^t| \geqslant \delta_e) + \mathbb{P}(|D_{ij}| \geqslant \lambda_{ij} - \delta_l - \delta_e) \end{aligned} \tag{68}$$

We already have good bounds for $\mathbb{P}(|\frac{1}{n_i}\sum_{t=1}^{n_i} e_i^tX_{ij}^t| \geqslant \delta_e)$ in Lemma 9. The remaining term $\mathbb{P}(|\sum_{k\in S*} w_k^*\frac{1}{n_i}\sum_{t=1}^{n_i}(X_{ik}^tX_{ij} - \sigma_{jk}^i)| \geqslant \delta_l)$ can be bound similarly as in Lemma 11.

Taking $\delta_e = 8|\eta_i \sigma_{jj}^i|\delta$ in Lemma 9 and noting that

$$\lambda_{ij} > |\sum_{k \in S*} w_k^* \sigma_{jk}^i| + 40\sqrt{2}\sqrt{\sum_{k \in S*} {w_k^*}^2 (\sigma_{kk}^i)^2 + \sum_{k \in S*} \sum_{l \in S*, k \neq l} w_k^* w_l^* \sigma_{kl}^i} \sigma_{jj}^i \delta + 8|\eta_i \rho_i|\delta \tag{69}$$

we can write:

$$\mathbb{E}(R_j) \leqslant \frac{6}{g} \sum_{i=1}^{g} \exp(-n_i \delta^2) \tag{70}$$

□

We can always take $n_i > \frac{\log 16}{\delta^2}$ to show that $\mathbb{E}(R_j) > \frac{1}{2}, \forall j \in S^*$ and $\mathbb{E}(R_j) < \frac{1}{2}, \forall j \in S_c^*$. This allows us to state a modified theorem for correlated Gaussian predictors.

Theorem 3 (Correlated Gaussian Predictors)**.** *For federated support learning in linear regression, as described in Section 3, with at least $g = \Omega(\log d)$ clients and correlated Gaussian predictor variables, if each client has $n_i = \Omega(s), s > 1$ i.i.d. data samples and the following condition holds:*

$$\begin{aligned}
&\max_{j \in S_c^*, i \in [g]} |\sum_{k \in S*} w_k^* \sigma_{jk}^i| + \frac{C}{\sqrt{s}} \left(40\sqrt{2}\sqrt{\sum_{k \in S*} {w_k^*}^2 (\sigma_{kk}^i)^2 + \sum_{k \in S*} \sum_{l \in S*, k \neq l} w_k^* w_l^* \sigma_{kl}^i} \sigma_{jj}^i + 8|\eta_i \sigma_{jj}^i| \right) \\
&< \lambda < \min_{j \in S*, i \in [g]} |w_j^* (\sigma_{jj}^i)^2 + \sum_{k \in S*, k \neq j} w_k \sigma_{jk}^i| \\
&- \frac{C}{\sqrt{s}} \left(8|w_j^*|(\sigma_{jj}^i)^2 + 40\sqrt{2}\sigma_{jj}^i \sqrt{\sum_{k \in S*, k \neq j} {w_k^*}^2 (\sigma_{kk}^i)^2 + \sum_{k \in S*} \sum_{l \in S*, k \neq l} w_k^* w_l^* \sigma_{kl}^i} + 8|\eta_i \sigma_{jj}^i| \right)
\end{aligned} \tag{71}$$

where $C > 0$ is an absolute constant independent of n_i, s and d, then Algorithm 1 recovers the exact support of the shared parameter vector w^* with probability at least $1 - \mathcal{O}(\frac{1}{d})$.

The proof for Theorem 3 follows by choosing $n_i = \Omega(s)$ and $\delta = \frac{C}{\sqrt{s}}$. Doing an analysis similar to Section 6, we can show that if $\lambda > \mathcal{O}(1) + \mathcal{O}(\frac{1}{\sqrt{s}})$, Algorithm 1 recovers the exact support of w^* with probability at least $1 - \mathcal{O}(\frac{1}{d})$. This allows for an overall sample complexity of $\Omega(s \log d)$ for correlated Gaussian predictors and matches the results of Wainwright (2009b) for correlated Gaussian predictors in a centralized setting.

G.1 A note on sub-Gaussian vectors

The most important part of the above proof is to show that $\sum_{k \in S*, k \neq j} w_k^* X_{ik}^t$ is also a Gaussian random variable. If we assume X_i to be a sub-Gaussian vector (even if predictors are correlated) as defined in Hsu et al. (2012), i.e., there exists a $\tau > 0$, such that for all $\alpha \in \mathbb{R}^d$ the following holds:

$$\mathbb{E}(\exp(\alpha^\intercal X_i)) \leqslant \exp(\frac{\|\alpha\|_2^2 \tau^2}{2}) \tag{72}$$

This readily implies that $\sum_{k \in S*, k \neq j} w_k^* X_{ik}^t$ is sub-Gaussian and the other parts of the above proof also follow. Again doing a similar analysis to Section 6, we can show that for $n_i = \Omega(s)$ and $k_{13} + \frac{k_{14}}{\sqrt{s}} < \lambda_i < k_{15}$ for some positive constants k_{13}, k_{14} and k_{15}, Algorithm 1 recovers the exact support of w^* with probability at least $1 - \mathcal{O}(\frac{1}{d})$. Thus, we can achieve an overall sample complexity of $\Omega(s \log d)$ for the case when the predictors form a sub-Gaussian vector.