A Supporting Lemmas

Lemma A.1. *(Lemma 11 in Abbasi-Yadkori et al. (2011)) Let $\{X_t\}_{t=1}^{\infty}$ be a sequence in $\mathbb{R}^d, V$ be a $d \times d$ positive definite matrix, and define $\bar{V}_t = V + \sum_{s=1}^{t} X_s X_s^\top$. Then, we have*

$$\log \left(\frac{\det \left(\bar{V}_n \right)}{\det(V)} \right) \leq \sum_{t=1}^{n} \|X_t\|_{\bar{V}_{t-1}^{-1}}^2 .$$

Further, if $\|X_t\|_2 \leq L$ for all t, then

$$\sum_{t=1}^{n} \min \left\{ 1, \|X_t\|_{\bar{V}_{t-1}^{-1}}^2 \right\} \leq 2 \left(\log \det \left(\bar{V}_n \right) - \log \det V \right) \leq 2 \left(d \log \left(\left(\operatorname{trace}(V) + nL^2 \right) / d \right) - \log \det V \right).$$

Finally, if $\lambda_{min}(V) \geq \max(1, L^2)$, then

$$\sum_{t=1}^{n} \|X_t\|_{\bar{V}_{t-1}^{-1}}^2 \leq 2 \log \frac{\det(\bar{V}_n)}{\det(V)}.$$

Lemma A.2. *(Lemma 12 in Abbasi-Yadkori et al. (2011)). Let A, B, and C be positive semi-definite matrices such that $A = B + C$. Then, we have*

$$\sup_{x \neq \mathbf{0}} \frac{x^\top A x}{x^\top B x} \leq \frac{\det(A)}{\det(B)}.$$

Theorem A.1. *(Theorem 2 in Abbasi-Yadkori et al. (2011)). Let $\{\mathcal{F}_i\}_{i=0}^{\infty}$ be a filtration. Let $\{x_i\}_{i=1}^{\infty}$ be an $\mathbb{R}^d$-valued stochastic process such that x_i is $\mathcal{F}_{i-1}$-measurable and $\|x_i\| \leq 1$ almost surely. Let $\{\epsilon_i\}_{i=1}^{\infty}$ be a real-valued stochastic process such that ε_i is $\mathcal{F}_i$-measurable and is sub-Gaussian with variance proxy 1 when conditioned on $\mathcal{F}_{i-1}$. Fix $\theta \in \mathbb{R}^d$ such that $\|\theta\| \leq 1$. Let $A_n = I + \sum_{i=1}^{n} x_i x_i^\top, r_i = x_i^\top \theta + \varepsilon_i$, and $\hat{\theta}_n = A_n^{-1} \sum_{i=1}^{n} r_i x_i$. For every $\delta > 0$, we have*

$$\mathbb{P} \left[\forall n \geq 0 : \left\| \hat{\theta}_n - \theta \right\|_{A_n} \leq 1 + \sqrt{d \ln \left(\frac{1+n}{\delta} \right)} \right] \geq 1 - \delta,$$

where we define $\|x\|_A = \sqrt{x^\top A x}$. Furthermore, when the above event holds, we have for every $n \geq 0$ and any vector $x \in \mathbb{R}^d$ that

$$\left| x^\top (\hat{\theta}_n - \theta) \right| \leq \left(1 + \sqrt{d \ln \left(\frac{1+n}{\delta} \right)} \right) \sqrt{x^\top A_n^{-1} x}.$$

Lemma A.3. *(Adapted from Lemma B.1 in He et al. (2022a)) Under the setting of Theorem 5.1, establish $C = 1/M^2, \alpha_0 = 1 + \sqrt{d \ln(2M^2 T/\delta)}$. In layer 0, with probability at least $1 - \delta$, the good event $\mathcal{E}_0$ happens:*

$$\mathcal{E}_0 \triangleq \left\{ \left| x_{t,a}^{i\top} \hat{\theta}_{t,s}^i - x_{t,a}^{i\top} \theta \right| \leq w_{t,s,a}^i, \forall i \in [M], a \in [K], t \in [T], s = 0 \right\}.$$

Lemma A.4. *(Lemma 31 in Ruan et al. (2021)). Given $\theta, x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ such that $\|\theta\| \leq 1$, for all $i \in [n]$, let $r_i = x_i^\top \theta + \epsilon_i$ where ϵ_i is an independent sub-Gaussian random variable with variance proxy 1. Let $A = I + \sum_{i=1}^{n} x_i x_i^\top$, and $\hat{\theta} = A^{-1} \sum_{i=1}^{n} r_i x_i$. For any $x \in \mathbb{R}^d$ and any $\alpha > 0$, we have*

$$\mathbb{P} \left[|x^\top (\theta - \hat{\theta})| > (\alpha + 1) \|x\|_{A^{-1}} \right] \leq 2 \exp\left(-\alpha^2 / 2 \right).$$

B Lemmas for the SupLinUCB Subroutine

We present several useful lemmas that are based on Algorithm 2. Recall that $\Psi_{t,s}$ represents the index set of rounds up to and including round t during which an action is taken in layer s. That is,

$$\Psi_{t,s} = \{ t' \in [t] : \exists i \in [M],\ a_{t'}^i \text{ is chosen in layer } s \}, \forall s \in [0 : S].$$

Similar to Lemma 4 in Chu et al. (2011), we claim that the rewards associated with rounds within each $\Psi_{t,s}, s \in [S]$ (excluding layer 0) are mutually independent.

Lemma B.1. *For each $t \in [T]$ each $s \in [S]$, given any fixed sequence of contexts $\{x^i_{t,a}, t \in \Psi_{t,s}\}$, the rewards $\{r^i_{t,s,a}, t \in \Psi_{t,s}\}$ are independent random variables with means $\mathbb{E}[r^i_{t,s,a}] = \theta^\top x^i_{t,s,a}$.*

Proof of Lemma B.1. For each $s \in [S]$ and each time t, the procedure of generating $\Psi_{t,s}$ only depends on the information in previous layers $\cup_{\sigma<s}\Psi_{t,\sigma}$ and confidence width $\{w^i_{t,s,a}, a \in [K]\}$. From its definition, $w^i_{t,s,a}$ only depends on $\{x_{\tau,a_\tau}, \tau \in \Psi_{t-1,s}\}$ and on the current context $x^i_{t,a}$. Thus the procedure of generating $\Psi_{t,s}$ does not depend on rewards $\{r_{\tau,a_\tau}, \tau \in \Psi_{t-1,s}\}$, and therefore the rewards are independent random variables when conditioned on $\Psi_{t,s}$. □

Given the above-mentioned statistical independence property, and by referring to Lemma A.4, we can establish the following lemma for each layer $s \in [S]$.

Lemma B.2. *Suppose the time index set $\Psi_{t,s}$ is constructed so that for fixed x_{τ,a_τ} with $\tau \in \Psi_{t,s}$, the rewards $\{r_{\tau,a_\tau}\}$ are independent random variables with mean $\mathbb{E}[r_{\tau,a_\tau}] = \theta^\top x_{\tau,a_\tau}$. For any round $t \in [T]$, if client $i_t = i$ is active and chooses arm a_t in layer $s \in [S]$, then with probability at least $1 - \frac{\delta}{MT\ln d}$, we have for any $a_t \in [K]$:*

$$\left|\hat{r}_{t,s,a_t} - \theta^\top x^i_{t,a_t}\right| \le w^i_{t,s,a_t} = \alpha_s \|x^i_{t,a_t}\|_{(A^i_{t,s})^{-1}}.$$

For layer 0, we employ the self-normalized martingale concentration inequality as outlined in He et al. (2022a). By resorting to Lemma A.3, we obtain the following:

Lemma B.3. *For any round $t \in [T]$, given that client $i_t = i$ is active in round t and arm a_t is chosen in layer 0, with probability at least $1 - \delta$, we have for any $a_t \in [K]$:*

$$\left|\hat{r}_{t,0,a_t} - \theta^\top x^i_{t,a_t}\right| \le w^i_{t,0,a_t} = \alpha_0 \|x^i_{t,a_t}\|_{(A^i_{t,0})^{-1}}.$$

Summarizing the discussions presented in Lemma B.2 and Lemma B.3, we now proceed to define the following good event:

Lemma B.4. *Define the good event $\mathcal{E}$ as:*

$$\mathcal{E} \triangleq \left\{\left|\hat{r}_{t,s,a} - x^{i\top}_{t,a}\theta\right| \le w^i_{t,s,a}, \forall i \in [M], a \in [K], t \in [T], s \in [0:S]\right\}. \tag{2}$$

We have $\mathbb{P}[\mathcal{E}] \ge 1 - \delta$.

Conditioned on the good event $\mathcal{E}$, the ensuing lemma illustrates that the optimal arm persists in the candidate set, and that the regret experienced in each layer aligns with the order of the confidence width.

Lemma B.5. *Conditioned on the good event $\mathcal{E}$, for $t \in [T]$, assume that client i is active and chooses an action $a_t \in \mathcal{A}_s$, and recall $(a^i_t)^*$ represents the optimal arm in the current round. For any $s' \le s$, we have:*

$$(a^i_t)^* = \arg\max_{a\in[K]} \theta^\top x^i_{t,a} = \arg\max_{a\in\mathcal{A}_{s'}} \theta^\top x^i_{t,a}.$$

Proof of Lemma B.5. For any time step $t \in [T]$, when the good event $\mathcal{E}$ holds, by the arm elimination rule in layer 0, we have

$$\hat{r}_{t,0,a^*} + w_{t,0,a^*} \ge \max_{a\in[K]} \theta^\top x_{t,a} \ge \max_{a\ne a^*} \theta^\top x_{t,a} \ge \max_{a\ne a^*}(\hat{r}_{t,0,a} - w_{t,0,a}).$$

Thus, $a^* \in \mathcal{A}_0$. For each layer $s' < s$, we have:

$$\hat{r}_{t,s',a^*} + w_{t,s',a^*} \ge \max_{a\in\mathcal{A}_{s'}} \theta^\top x_{t,a} \ge \max_{a\ne a^*, a\in\mathcal{A}_{s'}} \theta^\top x_{t,a} \ge \max_{a\ne a^*, a\in\mathcal{A}_{s'}} (\hat{r}_{t,s',a} - w_{t,s',a}).$$

Thus, we derive $\hat{r}_{t,s',a^*} \ge \max_{a\in\mathcal{A}_{s'}}(\hat{r}_{t,s',a}) - 2\overline{w}_{s'}$, which follows from $w_{t,s',a} \le \overline{w}_{s'}$ for all $a \in \mathcal{A}_{s'}$ by the arm elimination rule in Line 10 Algorithm 2. Therefore, arm eliminations will preserve the best arm. □

The forthcoming lemma demonstrates that, under the good event, the regret experienced in each layer aligns with the order of the corresponding confidence width.

Lemma B.6. *Conditioned on the good event $\mathcal{E}$, for $t \in [T]$ client $i \in [M]$ and $s \in [S]$, it holds that:*

$$\mathbb{I}[a_t \text{ is chosen in layer } 0](\max_{a\in\mathcal{A}_0} \theta^\top x_{t,a} - \theta^\top x_{t,a_t}) \le 4w_{t,0,a_t}, \tag{3}$$

$$\mathbb{I}[a_t \text{ is chosen in layer } s](\max_{a\in\mathcal{A}_s} \theta^\top x_{t,a} - \theta^\top x_{t,a_t}) \le 8\overline{w}_s. \tag{4}$$

Proof of Lemma B.6. If an action is taken in layer 0, we have that

$$a_t = \arg\max_{a\in\mathcal{A}_0, w_{t,0,a} > \overline{w_0}} w_{t,0,a},$$

and

$$\begin{aligned}
\max_{a\in\mathcal{A}_0} \theta^\top (x_{t,a} - \theta^\top x_{t,a_t}) &\le \max_{a\in\mathcal{A}_0} \theta^\top x_{t,a} - \min_{a\in\mathcal{A}_0} \theta^\top x_{t,a} \\
&\le \max_{a\in\mathcal{A}_0} (\hat{\theta}_0^\top x_{t,a} + w_{t,0,a}) - \min_{a\in\mathcal{A}_0} (\hat{\theta}_0^\top x_{t,a} - w_{t,0,a}) \\
&\le 4 \max_{a\in\mathcal{A}_0} w_{t,0,a} \\
&= 4w_{t,0,a_t}.
\end{aligned}$$

The second inequality is conditioned on the good event $\mathcal{E}$, and the third inequality arises from the arm elimination rule. If an action is taken in layer s, we establish the following:

$$a_t = \arg\max_{a\in\mathcal{A}_s, w_{t,s,a} > \overline{w}_s} w_{t,s,a},$$

and

$$\begin{aligned}
\max_{a\in\mathcal{A}_s} (\theta^\top x_{t,a} - \theta^\top x_{t,a_t}) &\le \max_{a\in\mathcal{A}_{s-1}} (\hat{\theta}_{s-1}^\top x_{t,a} + w_{t,s-1,a}) - \min_{a\in\mathcal{A}_{s-1}} (\hat{\theta}_{s-1}^\top x_{t,a} - w_{t,s-1,a}) \\
&\le 2 \max_{a\in\mathcal{A}_{s-1}} w_{t,s-1,a} + \max_{a\in\mathcal{A}_{s-1}} \hat{\theta}_{s-1}^\top - \min_{a\in\mathcal{A}_{s-1}} \hat{\theta}_{s-1}^\top x_{t,a} \\
&\le 2 \max_{a\in\mathcal{A}_{s-1}} w_{t,s-1,a} + 2\overline{w}_{s-1} \\
&\le 4\overline{w}_{s-1} \le 8\overline{w}_s.
\end{aligned}$$

The first inequality is based on the good event $\mathcal{E}$, the third inequality follows the arm elimination rule, and the fourth inequality is due to $w_{t,s-1,a} \le \overline{w}_{s-1}$ for all $a \in \mathcal{A}_{s-1}$. ☐

C Supporting Lemmas and Proofs for Async-`FedSupLinUCB`

Lemma C.1. *(Lemma 6.2 in He et al. (2022a)) In any epoch from round $T_{n,s}$ to round $T_{n+1,s} - 1$, the number of communications is at most $2(M + 1/C)$.*

Proof outline of Async-`FedSupLinUCB`. First, we reorganize the arrival pattern, demonstrating that the rearranged system parallels the original system, and present the requisite definitions for our analysis. Second, we deploy a virtual global model encapsulating information about all clients up to round t, subsequently interconnecting the local models with this global model. Lastly, we derive upper bounds on the regret and communication cost in each layer $s \in [0 : S]$ prior to aggregating them to yield the total regret and communication costs, respectively.

Suppose that client i communicates with the server at rounds t_1, t_2 with $t_1 < t_2$ and does not communicate during the rounds in between. The actions and information gained by client i at the rounds $t_1 < t < t_2$ do not impact other clients' decision-making, since the information is kept local without communication. Therefore, we can reorder the arrival of clients appropriately while keeping the reordered system equivalent to the original system.

More specifically, suppose client i communicates with the server at two rounds t_m and t_n and does not communicate in the rounds in between (even if she is active). We reorder all the active rounds of client i in $t_m < t < t_n$ and place them sequentially after the round t_m. Hence, the arrival of clients can be reordered such that each client communicates with the server and keeps active until the next client's communication begins. We assume that the sequence of communication rounds in the reordered arrival pattern is $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$, where in rounds $t_i \le t < t_{i+1}$, the active client is the same. Details of the reordering process are given in Definition C.2. Due to the equivalence between the original system and the reordered system, we carry out the proofs in the reordered system. Note that only one client i_t is active at round t, we will write $a_t = a_t^{i_t}$, $x_t = x_{t,a_t}^{i_t}$ and $r_t = r_{t,a_t}^{i}$ for simplicity.

Definition C.1. ***Client information.*** *Recall for each client $i \in [M]$, we denote by $L_i(t)$ the last round when client i communicated with the server before and including round t. E.g., $L_i(t) = t$ if client i communicates at round t. For each round t each client i and each layer s, the information that has been uploaded by client i to the server is: $A_{t,s}^{i,up} = \sum_{t'=1}^{L_i(t)} x_{t'} x_{t'}^\top \mathbb{I}\{i_{t'} = i, a_t \text{ in layer } s\}, b_{t,s}^{i,up} = \sum_{t'=1}^{L_i(t)} r_{t'} x_{t'} \mathbb{I}\{i_{t'} = i, a_t \text{ in layer } s\}$, and the local information in the buffer that has not been uploaded to the server is: $\Delta A_{t,s}^{\mathrm{i}} = \sum_{t'=L_i(t)+1}^{t} x_{t'} x_{t'}^\top \mathbb{I}\{i_{t'} = i, a_t \text{ in layer } s\}, \Delta b_{t,s}^{i} = \sum_{t'=L_i(t)+1}^{t} r_{t'} x_j \mathbb{I}\{i_{t'} = i, a_t \text{ in layer } s\}$.*

Server information. *The information in the server is the data uploaded by all clients up to round t: $A_{t,s}^{ser} = I + \sum_{i=1}^{M} A_{t,s}^{i,up}, b_{t,s}^{ser} = \sum_{i=1}^{M} b_{t,s}^{i,up}$.*

Time index set. *Denote by $\Psi_{t,s}$ the time index set when the action a_t^i is chosen in layer s. It can be expressed as $\Psi_{t,s} = \{t' \in [t], a_{t'}^i \text{ in layer } s,\ i \in [M]\}, s \in [0:S]\}$.*

Virtual global information. *We define a virtual global model that contains all the information up to round t as: $A_{t,s}^{all} = I + \sum_{t' \in \Psi_{t,s}} x_{t'} x_{t'}^\top, b_{t,s}^{all} = \sum_{t' \in \Psi_{t,s}} r_{t'} x_{t'}$.*

The information that is stored on the server and all the information that has not yet been uploaded by clients are combined to generate the global information: $A_{t,s}^{all} = A_{t,s}^{ser} + \sum_{i=1}^{M} \Delta A_{t,s}^{i}, b_{t,s}^{all} = b_{t,s}^{ser} + \sum_{i=1}^{M} \Delta b_{t,s}^{i}$.

Before presenting the proof, we define *good event* $\mathcal{E}$ as

$$\mathcal{E} \triangleq \left\{ \left| x_{t,a}^{i\top} \hat{\theta}_{t,s}^{i} - x_{t,a}^{i\top} \theta \right| \le w_{t,s,a}^{i}, \forall i \in [M], a \in [K], t \in [T], s \in [0:S] \right\}.$$

Recall $\hat{\theta}_{t,s}^{i}$ is the estimate of θ by client i, and $x_{t,a}^{i}$ and $w_{t,s,a}^{i}$ is the corresponding context and confidence width of the action taken at round t. The following lemma shows the good event happens with high probability, similar to the result in Lemma B.4.

Lemma C.2. *It holds that $\mathbb{P}[\mathcal{E}] \ge 1 - \delta$.*

Conditioned on the good event, to upper bound the regret, we bound the confidence width in each layer via the size of each time index set in the lemma below.

Lemma C.3. *Conditioned on the good event $\mathcal{E}$, for each $s \in [0 : S-1]$ we have:*

$$\sum_{t \in \Psi_{T,s}} w_{t,s,a}^{i} \le \alpha_s \sqrt{2(1+MC)} \sqrt{2d |\Psi_{T,s}| \log |\Psi_{T,s}|} + \alpha_s d M \log(1 + T/d).$$

Noting that $|\Psi_{T,s}| \le T$ naturally holds, we give a tighter (dimension-dependent) bound on the size of $\Psi_{T,0}$ so as to mitigate the larger coefficient α_0 as follows.

Lemma C.4. *The size of $\Psi_{T,0}$ can be bounded by $|\Psi_{T,0}| \le T \log T \log(2MT/\delta)/d$.*

We postpone the proofs of Lemma C.3 and Lemma C.4 until the end of this section, and instead focus on presenting the regret analysis next. Equipped with the previous lemmas, we are ready to analyze the total regret.

Proof of Theorem 5.1. (**Regret analysis**) The total regret can be decomposed w.r.t. layers as follows:

$$R_T = \mathbb{E} \sum_{t \in \Psi_{T,0}} (r_{t,a_t^{i,*}}^{i} - r_{t,a_t}^{i}) + \sum_{s=1}^{S} \mathbb{E} \sum_{t \in \Psi_{T,s}} (r_{t,a_t^{i,*}}^{i} - r_{t,a_t}^{i}).$$

Conditioned on the good event $\mathcal{E}$, we first bound the regret in layer 0 by

$$\mathbb{E} \sum_{t \in \Psi_{T,0}} (r_{t,a_t^{i,*}}^{i} - r_{t,a_t}^{i}) \le \sum_{t \in \Psi_{T,0}} 4 w_{t,0,a_t}$$

$$\leq 4\alpha_0\sqrt{2(1+MC)}\sqrt{2d|\Psi_{T,0}|\log|\Psi_{T,0}|} + 4\alpha_0 dM\log(1+T/d)s \leq \tilde{O}(\sqrt{(1+MC)dT}).$$

The first inequality follows Lemma B.6, the second inequality is from Lemma C.3, and the last inequality is due to Lemma C.4. We next bound the regret in each layer $s \in [1 : S-1]$ similarly by

$$\sum_{t\in\Psi_{T,s}} \mathbb{E}\left[r^i_{t,a^{i,*}_t} - r^i_{t,a_t}\right] \leq \sum_{t\in\Psi_{T,s}} 8\overline{w}_s \leq \sum_{t\in\Psi_{T,s}} 8w_{t,s,a_t}$$
$$\leq 8\alpha_s\sqrt{2(1+MC)}\sqrt{2d|\Psi_{T,s}|\log|\Psi_{T,s}|} + 8\alpha_s dM\log(1+T/d) \leq \tilde{O}(\sqrt{(1+MC)dT})$$

where the first inequality follows Lemma B.6, the second inequality is from the arm selection rule in line 13 Algorithm 2, and the third inequality is from Lemma C.3. For the last layer S, we have:

$$\sum_{t\in\Psi_{T,S}} \mathbb{E}\left[r^i_{t,a^{i,*}_t} - r^i_{t,a_t}\right] \leq \sum_{t\in\Psi_{T,S}} 8\overline{w}_S \leq 8\overline{w}_S|\Psi_{T,S}| \leq 8\overline{w}_S T \leq 8\sqrt{dT}.$$

Finally, with Lemma C.2, we have $R_T \leq \tilde{O}(\sqrt{(1+MC)dT})$.

(**Communication cost analysis**) Next, we study the communication cost in an asynchronous setting. For each layer s, $i \geq 0$, we define $T_{n,s} = \min\{t \in [T] | \det(A^{ser}_{t,s}) \geq 2^i\}$. We divide rounds in each layer into epoch $\{T_{n,s}, T_{n,s}+1, .., \min(T, T_{n+1,s}-1)\}$, and the communication rounds in the epoch $T_{n,s} \leq t \leq T_{n+1,s}-1$ can be bound by Lemma C.1. Let N' be the largest integer such that $T_{N',s}$ is not empty. According to Lemma A.1 that $\log(\det(A^{all}_{t,s})) \leq d\log(1+|\Psi_{T,s}|/d)$, $N' \leq d\log(1+T/d)$. The total number of epochs of layer s is bounded by $d\log(1+T/d)$. By lemma C.1 the communication rounds in layer s is bounded by $O((M+1/C))d\log T$. There are $S = \lceil \log d \rceil$ in the FedSupLinUCB algorithm, the total communication cost is thus upper bound by $O(d(M+1/C)\log d\log T)$. Plugging in $C = 1/M^2$ proves the result.

□

Definition C.2. *(**Reorder function**) Without loss of generality, we assume all clients communicate with the server at round $t_0 = 0$, and the sequence of rounds that clients communicate with the server in the original system is $0 \leq t_0 < t_1 < t_2 < ... < t_N \leq T$. Define $I_{t,i} = \mathbb{I}$(client i communicates with the server at round t). Denote by $L_i(t)$ the last communication round of client i before and including round t:*

$$L_i(t) := \inf\{u : \sum_{t'=0}^{u} I_{t',i} = \sum_{t'=0}^{t} I_{t',i}\}.$$

Denote by $N_i(t)$ the next communication round of client i including and after round t:

$$N_i(t) := \inf\{u : \sum_{t'=t}^{u} I_{t',i} = 1\}.$$

The round $t \in [T]$ in the original system is placed in round $\phi(t)$ by the reordering function $\phi : [T] \to [T]$. We first reorder the communication round, suppose two consecutive communication rounds t_n and t_{n+1} with $t_n < t_{n+1}$, and client i is active at round t_n and client j is active at round t_{n+1}.

$$\phi(t_{n+1}) = \phi(t_n) + \sum\nolimits_{t'=t_n}^{N_i(t_n)} I(i_{t'} = i) - 1.$$

Then we reorder the no-communication rounds, assuming client i is active at round t and does not communicate at this round. We first find the last communication round of client i as $L_i(t)$, and place round t by $\phi(t)$:

$$\phi(t) = \phi(L_i(t)) + \sum\nolimits_{t'=L_i(t)}^{t} I(i_{t'} = i) - 1.$$

Lemma C.5. *(Adapted from Lemma 6.5 in He et al. (2022a)) For each round $t \in [T]$ each layer $s \in [0 : S]$ and each client $i \in [M]$, we have:*

$$A^{ser}_{t,s} = I + \sum_{i=1}^{M} A^{i,up}_{t,s} \succeq \frac{1}{C}\Delta A^i_{t,s}.$$

Further averaging the inequality above over M clients, we have:

$$A_{t,s}^{ser} = I + \sum_{i=1}^{M} A_{t,s}^{i,up} \succeq \frac{1}{MC} \sum_{i=1}^{M} \Delta A_{t,s}^{i}.$$

Proof of Lemma C.5. Without loss of generality, we consider client i and fix any round $t \in [T]$. Let $t_1 \leq t$ be the last round such that client i was active at round t_1. If client i communicated with the server at round t_1, and chose action a_{t_1} at layer s, then we have

$$A_{t,s}^{ser} = I + \sum_{i=1}^{M} A_{t,s}^{i,up} \succeq \frac{1}{C} \Delta A_{t_1,s}^{i} = 0$$

for other layers $s' \neq s$, according to the determinant-based communication criterion, we have:

$$\det\big(A_{t_1,s'}^{i} + \Delta A_{t_1,s'}^{i}\big) < (1+C)\det\big(A_{t_1,s'}^{i}\big).$$

By Lemma A.2 we have

$$A_{t,s'}^{i} = A_{t_1,s'}^{i} \succeq \frac{1}{C} \Delta A_{t_1,s'}^{i}.$$

Otherwise, if no communication happened at round t_1, by the communication criterion, at the end of round t_1, for each layer $s \in [0:S]$, we have $A_{t_1,s}^{i} \succeq \frac{1}{C} \Delta A_{t_1,s}^{i}$. Note that $\{A_{t_1,s}^{i},\ s \in [0:S]\}$ are the downloaded gram matrices from last communication before round t_1, so it must satisfy $A_{t_1,s}^{i} \preceq A_{t_1,s}^{ser}$ for all $s \in [0:S]$. For round t, since client i is inactive from round t_1 to t, we have for all $s \in [0:S]$:

$$A_{t,s}^{ser} \succeq A_{t_1,s}^{ser} \succeq A_{t_1,s}^{i} \succeq \frac{1}{C} \Delta A_{t_1,s}^{i} = \frac{1}{C} \Delta A_{t,s}^{i}$$

where the last equality holds for inactivation, which completes the proof of the first claim. Further average the above inequality over all clients $i \in [M]$, and we get:

$$A_{t,s}^{ser} = I + \sum_{i=1}^{M} A_{t,s}^{i,up} \succeq \frac{1}{MC} \sum_{i=1}^{M} \Delta A_{t,s}^{i}.$$

$\square$

Recall that client i utilizes $A_{t,s}^{i}$ and $b_{t,s}^{i}$ to make the decision at round t, which were received from the server during the last communication. The following lemma establishes a connection between the gram matrix of the virtual global model and the gram matrix in the active client at round t.

Lemma C.6. *In the reordered arrival pattern, for any $1 \leq t_1 < t_2 \leq T$, suppose client i communicates with the server at round t_1, and keep active during rounds $t_1 \leq t \leq t_2 - 1$. Then for rounds $t_1 + 1 \leq t \leq t_2 - 1$, it holds that for each $s \in [0:S]$:*

$$A_{t,s}^{i} \succeq \frac{1}{1+MC} A_{t,s}^{all}.$$

Proof of Lemma C.6. Client i is the only active client from round t_1 to $t_2 - 1$ and only communicated with the server at round t_1, which implies that for $t_1 + 1 \leq t \leq t_2 - 1 \forall s \in [0:S]$, we have

$$A_{t,s}^{i} = I + \sum_{i=1}^{M} A_{t_1,s}^{i,up} = I + \sum_{i=1}^{M} A_{t,s}^{i,up} \succeq \frac{1}{1+MC}(I + \sum_{i=1}^{M} A_{t,s}^{i,up} + \sum_{i=1}^{M} \Delta A_{t,s}^{i}) \succeq \frac{1}{1+MC} A_{t,s}^{all}$$

where the second equality holds due to the fact that no clients communicate with the server from round $t_1 + 1$ to $t_2 - 1$, and the first inequality follows Lemma C.5. $\square$

Proof of Lemma C.3. For $t \in \Psi_{T,s}$, if no communication happened at round t, under Lemma C.6 and Lemma A.2, we can connect confidence width at the local client with the global gram matrix as:

$$\|x_{t,a}^{i_t}\|_{(A_{t,s}^{i_t})^{-1}} \leq \sqrt{1+MC}\|x_{t,a}^{i}\|_{(A_{t,s}^{all})^{-1}}.$$

It remains to control the communication rounds in $\Psi_{T,s}$. We define

$$T_n = \min\left\{t \in \Psi_{T,s} \mid \det\left(A_{t,s}^{all}\right) \geq 2^n\right\},$$

and let N' be the largest integer such that $T_{N'}$ is not empty. According to Lemma A.1, we have:

$$\log\left(\det\left(A_{t,s}^{all}\right)\right) \leq d\log(1 + |\Psi_{T,s}|/d).$$

Thus, $N' \leq d\log(1 + T/d)$. For each time interval from T_n to T_{n+1} and each client $i \in [M]$, suppose client i communicates with the server more than once, and communication rounds sequentially are $T_{n,1}, T_{n,2}, \ldots, T_{n,k} \in [T_n, T_{n+1})$. Then for each $j = 2, \ldots, k$, since client i is active at rounds $T_{n,j-1}$ and $T_{n,j}$, we have

$$\|x_{T_{n,j}}\|_{(A^i_{T_{n,j},s})^{-1}} \leq \|x_{T_{n,j}}\|_{(A^i_{T_{n,j-1}+1,s})^{-1}} \leq \sqrt{1+MC}\|x_{T_{n,j}}\|_{((A^{all}_{T_{n,j-1}+1,s})^{-1}}.$$

Since $\det\left(A^{all}_{T_{n+1}-1,s}\right)/\det\left(A^{all}_{T_{n,j-1}+1,s}\right) \leq 2^{n+1}/2^n = 2$, by the definition of T_n, we have:

$$\|x_{T_{n,j}}\|_{(A^i_{T_{n,j},s})^{-1}} \leq \sqrt{2(1+MC)}\|x_{T_{n,j}}\|_{(A^{all}_{T_{n+1}-1,s})^{-1}} \leq \sqrt{2(1+MC)}\|x_{T_{n,j}}\|_{(A^{all}_{T_{n,j},s})^{-1}},$$

where the second inequality comes from $A^{all}_{T_{n+1}-1,s} \succeq A^{all}_{T_{n,j},s}$. Specifically, for round $T_{i,1}$ the first communication round, we can bound the confidence width by 1. Thus, for the communication rounds in $\Psi_{T,s}$, we have:

$$\sum_{t\in\Psi_{T,s}, round\ t\ comm} \|x^{i_t}_{t,a}\|_{(A^{i_t}_{t,s})^{-1}} \leq MN' + \sum_{t\in\Psi_{T,s}, round\ t\ comm} \sqrt{2(1+MC)}\|x^{i_t}_{t,a}\|_{(A^{all}_{t,s})^{-1}}.$$

Finally, we put all rounds in $\Psi_{T,s}$ together:

$$\begin{aligned}
\sum_{t\in\Psi_{T,s}} w_{t,s,a} &= \alpha_s \sum_{t\in\Psi_{T,s}} \|x^{i_t}_{t,a}\|_{(A^{i_t}_{t,s})^{-1}} \\
&\leq \alpha_s \sum_{t\in\Psi_{T,s}} \sqrt{2(1+MC)}\|x^{i_t}_{t,a}\|_{(A^{all}_{t,s})^{-1}} + \alpha_s MN' \\
&\leq \alpha_s\sqrt{2(1+MC)}\sqrt{2d|\Psi_{T,s}|\log|\Psi_{T,s}|} + \alpha_s dM\log(1+T/d)
\end{aligned}$$

where the second inequality follows Lemma A.1. □

Proof of Lemma C.4. Based on the algorithm, if we choose an action in layer 0, the selected arm is

$$a_t = \arg\max_{a\in\mathcal{A}_0, w_{t,0,a} > \overline{w}_0} w_{t,0,a},$$

and the corresponding confidence width satisfies $w_{t,0,a_t} > \overline{w}_0$. Furthermore,

$$\begin{aligned}
\overline{w}_0|\Psi_{T,0}| \leq \sum_{t\in\Psi_{T,0}} w_{t,0,a_t} &= \alpha_0 \sum_{t\in\Psi_{T,0}} \|x_{t,a_t}\|_{(A^{i_t}_{t,0})^{-1}} \\
&\leq \alpha_0\sqrt{2(1+MC)}\sqrt{2d|\Psi_{T,s}|\log|\Psi_{T,s}|} + \alpha_0 dM\log(1+T/d),
\end{aligned}$$

where the last inequality is by Lemma C.3. We can thus conclude that $\Psi_{T,0} \leq T\log T\log(2MT/\delta)/d$.

□

D Supporting Lemmas and Proofs for Sync-FedSupLinUCB

Proof outline of Sync-FedSupLinUCB. To prove a high-probability regret bound, we first define the good event $\mathcal{E}$ in the following lemma, under which the regret bound is derived.

Lemma D.1. *Define* $\mathcal{E} \triangleq \{\left|x^{i\top}_{t,a}\hat{\theta}^i_{t,s} - x^{i\top}_{t,a}\theta\right| \leq w^i_{t,s,a}, \forall i \in [M], a \in [K], t \in [T_c], 0 \leq s \leq S\}$. *Then,* $\mathbb{P}[\mathcal{E}] \geq 1-\delta$.

Define client i's one-step regret at round t as $\mathsf{reg}_t^i = \theta^\top(x_{t,a_t^{i*}}^i - x_{t,a_t}^i)$. Let $\mathsf{reg}_{t,s}^i = \mathsf{reg}_t^i$ if action a_t is chosen in layer s; otherwise $\mathsf{reg}_{t,s}^i = 0$. The total regret can be written as

$$R_T = \sum_{i=1}^{M}\sum_{t=1}^{T_c} \mathsf{reg}_t^i = \sum_{s=0}^{S}\sum_{i=1}^{M}\sum_{t=1}^{T_c} \mathsf{reg}_{t,s}^i.$$

Fix an arbitrary $s \in \{0, 1, \ldots, S\}$, we analyze the total regret induced by the actions taken in layer s, i.e., $R_{s,T_c} = \sum_{i=1}^{M}\sum_{t=1}^{T_c} \mathsf{reg}_{t,s}^i$. The analysis can be carried over to different s in the same manner.

We call the chunk of consecutive rounds without communicating information in layer s (except the last round) an *epoch*. In other words, information in layer s is collected locally by each client and synchronized at the end of the epoch, following which the next epoch starts. The set of rounds that at least one client is pulling an arm in layer s can then be divided into multiple consecutive epochs, and we further dichotomize these epochs into good and bad epochs in the following definition.

Definition D.1. ***(Good epoch)*** *Suppose the set of rounds that at least one client is pulling an arm in layer s are divided into P epochs and denoted by $A_{p,s}^{all}, b_{p,s}^{all}$ the synchronized gram matrix and reward-action vector at the end of the p-th epoch. P epochs can then be dichotomized into $\mathcal{P}_s^{good} \triangleq \left\{p \in [P] : \frac{\det\left(A_{p,s}^{all}\right)}{\det\left(A_{p-1,s}^{all}\right)} \le 2\right\}, \mathcal{P}_s^{bad} \triangleq [P] \setminus \mathcal{P}_s^{good}$, where $A_{0,s}^{all} \triangleq I$. We say round t is* good *if the epoch containing round t belongs to $\mathcal{P}_s^{good}$; otherwise t is* bad*.*

We bound regrets in layer s induced by the good and bad epochs separately in the following lemmas. Recall $\Psi_{t,s}$ is the time index set when the action a_t^i is chosen in the s layer.

Lemma D.2. *Conditioned on the good event $\mathcal{E}$, for each layer $s \in [0 : S]$, the regret induced by good epochs of layer s is bounded as $\sum_{t\in\Psi_{T_c,s}, t \text{ is good}} \mathsf{reg}_{t,s}^i \le \tilde{O}\left(\alpha_s\sqrt{d|\Psi_{T_c,s}|\log(MT_c)}\right)$.*

Lemma D.3. *Define $D = \frac{T_c \log T_c}{d^2 M}$ and $R_s = d\log\left(1 + \frac{|\Psi_{T_c,s}|}{d}\right)$. Conditioned on the good event $\mathcal{E}$, for each layer $s \in [0 : S]$, the regret induced by bad epochs of layer s is bounded as $\sum_{t\in\Psi_{T_c,s}, t \text{ is bad}} \mathsf{reg}_{t,s}^i \le O\left(\alpha_s M\sqrt{D}R_s\right)$.*

Lemma D.4. *We have $|\Psi_{T_c,s}| \le \tilde{O}(\frac{MT_c}{d})$.*

Proof of Theorem 6.1. (**Regret analysis**) For each $s \in [0 : S]$, the regret induced in layer s is bounded by:

$$\begin{aligned} R_{s,T_c} &\le \sum_{t\in\Psi_{T_c,s}, t \text{ is good}} \mathsf{reg}_{t,s}^i + \sum_{t\in\Psi_{T_c,s}, t \text{ is bad}} \mathsf{reg}_{t,s}^i \\ &\le O(\alpha_s\sqrt{d|\Psi_{T_c,s}|\log(MT)} + \alpha_s M\sqrt{D}R_s) \le \tilde{O}(\sqrt{dMT_c}) \end{aligned}$$

where the second inequality is from Lemmas D.2 and D.3, and the last inequality is due to Lemma D.4. The total regret can thus be bounded as $R_T = \sum_{s=0}^{S} R_{s,T_c} = \tilde{O}(\sqrt{dMT_c})$. ☐

Proof of Lemma D.2. If t is good and belongs to the p-th epoch, we have by Lemma A.2 that

$$w_{t,s,a}^i = \alpha_s\|x_{t,a}^i\|_{(A_{t,s}^i)^{-1}} \le \sqrt{2}\alpha_s\|x_{t,a}^i\|_{(A_{p,s}^{all})^{-1}} \le 2\alpha_s\|x_{t,a}^i\|_{(A_{p-1,s}^{all})^{-1}}. \tag{5}$$

Within p-th good epoch, we have

$$A_{p-1,s}^{all} + \sum_{i=1}^{M}\sum_{t\in p\text{-th good epoch}} x_{t,a_t^i}^i(x_{t,a_t^i}^i)^\top = A_{p,s}^{all},$$

which together with inequality (5) and the last inequality in the elliptical potential lemma (Lemma A.1) imply that

$$\sum_{i=1}^{M}\sum_{t\in p\text{-th good epoch}} \|x_{t,a_t^i}^i\|_{(A_{t,s}^i)^{-1}}^2 \le 4\log\frac{\det\left(A_{p,s}^{all}\right)}{\det\left(A_{p-1,s}^{all}\right)}.$$

Thus under event $\mathcal{E}$, the regret induced by good epochs of layer s is

$$\begin{aligned}\sum_{(i,t)\in\Psi_{T,s},t\text{ is good}} reg^i_{t,s} &\leq \sum_{(i,t)\in\Psi_{T,s},t\text{ is good}} 8w^i_{t,s,a^i_t}\\ &\leq 8\sqrt{|\Psi_{T,s}|\sum_{(i,t)\in\Psi_{T,s},t\text{ is good}}(w^i_{t,s,a^i_t})^2}\\ &=\tilde{O}\left(\alpha_s\sqrt{d|\Psi_{T,s}|\log(MT)}\right),\end{aligned}$$

where the first inequality is from Lemma B.6, the second inequality is by Cauchy-Schwartz inequality, and the last relation is from

$$\sum_{p=1}^{P}\log\frac{\det\left(A^{all}_{p,s}\right)}{\det\left(A^{all}_{p-1,s}\right)}=\log\det\left(A^{all}_{P,s}\right)\leq d\log\left(1+\frac{|\Psi_{T,s}|}{d}\right)=R_s. \tag{6}$$

□

Proof of Lemma D.3. Denote by $R_s = d\log\left(1+\frac{|\Psi_{T,s}|}{d}\right)$. It follows that the number of bad epochs is at most $O(R_s)$. Moreover, the regret within a bad epoch of length n can be upper bounded as $O(M+\alpha_s M\sqrt{D})$ by applying the elliptical potential lemma for each client i and the communication condition, where the extra 1 in the upper bound is due to that at most M clients trigger the communication condition at the end of the p-th epoch. We thus have

$$\sum_{t\text{ is bad}} reg^i_{t,s}\leq\sum_{t\in\Psi_{T,s}\text{ is bad}}8w^i_{t,s}=O\left(MR_s+\alpha_s M\sqrt{D}R_s\right)=O\left(\alpha_s M\sqrt{D}R_s\right).$$

□

Proof of Lemma D.4. Recall $D=\frac{T\log(T)}{d^2M}$. Note that if $\alpha_s\sqrt{d|\Psi_{T,s}|\log(T)}=O(\alpha_s M\sqrt{D}R_s)$, we have $|\Psi_{T,s}|=\tilde{O}(M^2Dd)=\tilde{O}(\frac{MT}{d})$. Otherwise $|\Psi_{T,s}|\overline{w}^s_t=O(\alpha_s\sqrt{d|\Psi_{T,s}|\log(T)})$, which implies $|\Psi_{T,s}|=\tilde{O}(\frac{\alpha_s^2 d}{(\overline{w}^s_t)^2})=\tilde{O}(\frac{MT4^s}{d})$. □

E Variance-adaptive Async-`FedSupLinUCB`

The variance-adaptive SupLinUCB subroutine is presented in Alg. 5, while the complete variance-adaptive Async-`FedSupLinUCB` is given in Alg. 6.

E.1 Algorithm

E.2 Supporting Lemmas and Proofs

Theorem E.1. *(Theorem 4.3 in Zhou and Gu (2022)) Let $\{\mathcal{F}_t\}_{t=1}^{\infty}$ be a filtration, and $\{x_t,\eta_t\}_{t\geq 1}$ be a stochastic process such that $x_t\in\mathbb{R}^d$ is $\mathcal{F}_t$-measurable and $\eta_t\in\mathbb{R}$ is $\mathcal{F}_{t+1}$-measurable. Let $\sigma,\epsilon>0,\theta^*\in\mathbb{R}^d$. For $t\geq 1$, let $y_t=\langle\theta^*,x_t\rangle+\eta_t$ and suppose that η_t,x_t also satisfy*

$$\mathbb{E}[\eta_t\mid\mathcal{F}_t]=0,\mathbb{E}[\eta_t^2\mid\mathcal{F}_t]\leq\sigma^2,|\eta_t|\leq R,\|x_t\|_2\leq 1.$$

For $t\geq 1$, let $Z_t=I+\sum_{i=1}^t x_i x_i^\top, b_t=\sum_{i=1}^t y_i x_i, \theta_t=Z_t^{-1}b_t$, and

$$\begin{aligned}\beta_t =&12\sqrt{\sigma^2 d\log\left(1+tL^2/(d)\right)\log\left(32(\log(R/\epsilon)+1)t^2/\delta\right)}\\ &+24\log\left(32(\log(R/\epsilon)+1)t^2/\delta\right)\max_{1\leq i\leq t}\left\{|\eta_i|\min\left\{1,\|\mathbf{x}_i\|^{-1}_{\mathbf{Z}^{-1}_{i-1}}\right\}\right\}+6\log\left(32(\log(R/\epsilon)+1)t^2/\delta\right)\epsilon.\end{aligned}$$

Then, for any $0<\delta<1$, we have with probability at least $1-\delta$ that,

$$\forall t\geq 1,\left\|\sum_{i=1}^t x_i\eta_i\right\|_{Z_t^{-1}}\leq\beta_t,\quad\|\theta_t-\theta^*\|_{Z_t}\leq\beta_t+\|\theta^*\|_2.$$

Algorithm 5 Variance-adaptive SupLinUCB subroutine: VS-LUCB

Initialization: $S \leftarrow \lceil \log R + \log T \rceil$, $\overline{w}_0 = dR^2$, $\overline{w}_s \leftarrow 2^{-s}\overline{w}_0, \forall s \in [1:S]$,
$\alpha_0 = \tilde{O}(\sqrt{d}), \alpha_s = 1 + \sqrt{2\ln(2KMT\ln d/\delta)}, \rho = 1/\sqrt{T}, \gamma = R^{1/2}/d^{1/4}$.
Input: Client i (with local information $A^i, b^i, \Delta A^i, \Delta b^i$), contexts set $\{x^i_{t,1}, \ldots, x^i_{t,K}\}$
$A^i_{t,s} \leftarrow A^i_s, b^i_{t,s} \leftarrow b^i_s$ for `lazy update`
$\hat{\theta}_s \leftarrow (A^i_{t,s})^{-1} b^i_{t,s}$, $\hat{r}^i_{t,s,a} = \hat{\theta}_s^\top x^i_{t,a}$, $w^i_{t,s,a} \leftarrow \alpha_s \|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}, \forall s \in [0:S], \forall a \in [K]$.
$s \leftarrow 0$; $\mathcal{A}_0 \leftarrow \{a \in [K] \mid \hat{r}^i_{t,0,a} + w^i_{t,0,a} \geq \max_{a \in [K]}(\hat{r}^i_{t,0,a} - w^i_{t,0,a})\}$ ▷ Initial screening
repeat ▷ Layered successive screening
  if $s = S$ **then**
    Choose action a^i_t arbitrarily from $\mathcal{A}_S$
  else if $w^i_{t,s,a} \leq \overline{w}_s$ for all $a \in \mathcal{A}_s$ **then**
    $\mathcal{A}_{s+1} \leftarrow \{a \in \mathcal{A}_s \mid \hat{r}^i_{t,s,a} \geq \max_{a' \in \mathcal{A}_s}(\hat{r}^i_{t,s,a'}) - 2\overline{w}_s\}$; $s \leftarrow s+1$
  else
    Choose $a_t = \arg\max_{\{a \in \mathcal{A}_s, w^i_{t,s,a} > \overline{w}_s\}} w^i_{t,s,a}$
  end if
until action a_t is found
Take action a_t and and receive reward r^i_{t,a_t} and variance σ_t
$\overline{\sigma}_t = \max\{\sigma_t, \rho, \gamma \|x^i_{t,a_t}\|^{1/2}_{(A^i_{t,s})^{-1}}\}$
$\Delta A^i_s \leftarrow \Delta A^i_s + x^i_{t,a_t} x^{i\top}_{t,a_t}/\overline{\sigma}_t^2$, $\Delta b^i_s \leftarrow \Delta b^i_s + r^i_{t,a_t} x^i_{t,a_t}/\overline{\sigma}_t^2$ ▷ Update local information
Return layer index s

Algorithm 6 Variance-adaptive Async-`FedSupLinUCB`

Initialization: $T, C, S = \lceil \log R + \log T \rceil$
$\{A^{ser}_s \leftarrow I_d, b^{ser}_s \leftarrow 0 \mid s \in [0:S]\}$ ▷ Server initialization
$\{A^i_s \leftarrow I_d, \Delta A^i_s, b^i_s, \Delta b^i_s \leftarrow 0 \mid s \in [0:S], i \in [M]\}$ ▷ Clients initialization
for $t = 1, 2, ..., T$ **do**
  Client $i_t = i$ is active, and observes K contexts $\{x^i_{t,1}, x^i_{t,2}, \cdots, x^i_{t,K}\}$
  $s = \texttt{VS-LUCB}\left(client i, \{x^i_{t,1}, x^i_{t,2}, \cdots, x^i_{t,K}\}\right)$ with the lazy update
  if $\frac{\det\left(A^i_s + \Delta A^i_s\right)}{\det(A^i_s)} > (1+C)$ **then**
    Sync(s, server, clients i) for each $s \in [0:S]$
  end if
end for

Lemma E.1. *(Adapted from Lemma B.1 in Zhou and Gu (2022)). Let $\{\sigma_t, \beta_t\}_{t \geq 1}$ be a sequence of non-negative numbers, $\rho, \gamma > 0, \{x_t\}_{t \geq 1} \subset \mathbb{R}^d$ and $\|x_t\|_2 \leq 1$. Let $\{Z_t\}_{t \geq 1}$ and $\{\bar{\sigma}_t\}_{t \geq 1}$ be recursively defined as follows:*

$$Z_1 = I; \quad Z_{t+1} = Z_t + x_t x_t^\top / \bar{\sigma}_t^2, \quad \forall t \geq 1, \bar{\sigma}_t = \max\{\sigma_t, \rho, \gamma \|x_t\|^{1/2}_{Z_t^{-1}}\}.$$

Let $\iota = \log\bigl(1 + T/(d\rho^2)\bigr)$. Then we have

$$\sum_{t=1}^T \min\{1, \beta_t \|x_t\|_{Z_t^{-1}}\} \leq 2d\iota + 2\beta_T \gamma^2 d\iota + 2\sqrt{d\iota}\sqrt{\sum_{t=1}^T \beta_t^2(\sigma_t^2 + \rho^2)}.$$

Following a similar proof structure to Async-`FedSupLinUCB`, we employ a novel Bernstein-type self-normalized martingale inequality, proposed by Zhou and Gu (2022), for layer 0 to manage the variance information. We define $\alpha_0 = \beta_T$ as specified in Theorem E.1, and establish the following lemma, analogous to Lemma B.3.

Lemma E.2. *For any round $t \in [T]$, if client $i_t = i$ is active in round t and arm a_t is chosen in layer* 0*, with probability at least $1 - \delta$, with $\alpha_0 = \tilde{O}(\sqrt{d})$ we have for any $a_t \in [K]$:*

$$\left|\hat{r}_{t,0,a_t} - \theta^\top x^i_{t,a_t}\right| \leq w^i_{t,0,a_t} = \alpha_0 \|x^i_{t,a_t}\|_{(A^i_{t,0})^{-1}}.$$

We define *good event* $\mathcal{E}$ as $\mathcal{E} \triangleq \left\{ \left| x_{t,a}^{i\top} \hat{\theta}_{t,s}^{i} - x_{t,a}^{i\top} \theta \right| \leq w_{t,s,a}^{i}, \forall i \in [M], a \in [K], t \in [T], s \in [0:S] \right\}$. In a manner similar to the proof of Lemma B.4, we have that $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$.

Lemma E.3. *Conditioned on the event $\mathcal{E}$, the regret in layer 0 can be bounded by* $\mathsf{reg}_{\textit{layer}\ 0} \leq \tilde{O}(d)$.

Proof of Lemma E.3. We set $\overline{w}_0 = dR^2$ to provide a tighter bound for the size of $\Psi_{T,0}$. Mirroring the proof methodology in Lemma C.4, we establish the following:

$$\overline{w}_0 |\Psi_{T,0}| \leq \alpha_0 \sum_{t \in \Psi_{T,0}} \|x_{t,i}\|_{(A_{t,s}^{i_t})^{-1}} \leq 2d\iota + 2\alpha_0 \gamma^2 d\iota + 2\alpha_0 \sqrt{d\iota} \sqrt{\sum_{t \in \Psi_{T,0}} (\sigma_t^2 + \rho^2)}$$
$$\leq 2d\iota + 2\alpha_0 \gamma^2 d\iota + 2\alpha_0 \sqrt{d\iota} \sqrt{|\Psi_{T,0}|(R^2 + \rho^2)}.$$

The first inequality results from the arm selection rule of layer 0, the second is derived from Lemma E.1, and the third arises due to the constraint $\sigma_t^2 \leq R^2$. Consequently, we infer that $\Psi_{T,0} \leq O(d^2 R^2 / \overline{w}_0^2)$. We can then bound the regret in layer 0 as follows:

$$\mathsf{reg}_{\text{layer } 0} \leq 4\alpha_0 \sum_{t \in \Psi_{T,0}} \|x_{t,i}\|_{(A_{t,s}^{i_t})^{-1}} \leq 8d\iota + 8\alpha_0 \gamma^2 d\iota + 8\alpha_0 \sqrt{d\iota} \sqrt{|\Psi_{T,0}|(R^2 + \rho^2)} \leq \tilde{O}(d).$$

□

Lemma E.4. *Conditioned on the event $\mathcal{E}$, the regret of each layer $s \in [1 : S-1]$ can be bounded by* $\mathsf{reg}_{\textit{layer } s} \leq \tilde{O}\sqrt{d \sum_t \sigma_t^2}$.

Proof of Lemma E.4. For $s \in \{1, 2, ..., S-1\}$, the rewards in each layer s are mutually independent, as proven in Lemma B.1. We deduce:

$$\mathsf{reg}_{\text{layer } s} \leq 8\overline{w}_s |\Psi_{T,s}| \leq 8\alpha_s \sum_{t \in \Psi_{T,s}} \|x_{t,i}\|_{(A_{t,s}^{i_t})^{-1}}$$
$$\leq \alpha_s \sum_{t \in \Psi_{T,s}} \|x_{t,a}^{i_t}\|_{(A_{t,s}^{all})^{-1}} + \alpha_s dM \log(1 + T/d)$$
$$\leq \tilde{O}(\sqrt{d \sum_{t \in \Psi_{T,s}} \sigma_t^2}).$$

The first inequality arises from Lemma B.6, the second is a result of the arm selection rule in Line 13, the third derives from Lemma C.3, and the final inequality is attributable to Lemma E.1. □

For the final layer S, applying Lemma B.6 and setting $\overline{w}_S = d/T$, we have $\mathsf{reg}_{\text{layer } S} \leq 8\overline{w}_s |\Psi_S| \leq \tilde{O}(d)$.

Proof of the communication bound in Theorem 7.1. Having established the bound for regret in each layer, we have demonstrated that $R_T \leq \tilde{O}\left(\sqrt{d \sum_{t=1}^{T} \sigma_t^2}\right)$. Given that we set $\overline{w}_0 = dR^2$ and $\overline{w}_S = d/T$, it requires $S = \log(\overline{w}_0 / \overline{w}_S) = \Theta(\log R + \log T)$ layers to achieve the desired accuracy. The number of communications triggered by layer s can be upper bounded by $O(dM^2 \log(T)$ (Lemma C.1). Consequently, we are able to constrain the overall communication cost to $\tilde{O}(dM^2 \log^2 T)$.

F Corruption Robust Async-`FedSupLinUCB`

The corruption robust SupLinUCB subroutine is presented in Alg. 7, while the complete corruption robust Async-`FedSupLinUCB` is given in Alg. 8.

Algorithm 7 Corruption Robust SupLinUCB subroutine: CS-LUCB

Initialization: $S = \lceil \log d \rceil$, $\overline{w}_0 = d^{1.5}/\sqrt{T}$, $\overline{w}_s \leftarrow 2^{-s}\overline{w}_0$, $\gamma = \sqrt{d}/C_p$.
$\alpha_0 = 1 + \sqrt{d \ln(2M^2T/\delta)} + \gamma C_p, \alpha_s \leftarrow 1 + \sqrt{2\ln(2KMT \ln d/\delta)} + \gamma C_p, \forall s \in [1:S]$
Input: Client i (with local information $A^i, b^i, \Delta A^i, \Delta b^i$), contexts set $\{x^i_{t,1}, \ldots, x^i_{t,K}\}$
$A^i_{t,s} \leftarrow A^i_s, b^i_{t,s} \leftarrow b^i_s$ for `lazy update`
$\hat{\theta}_s \leftarrow (A^i_{t,s})^{-1} b^i_{t,s}$, $\hat{r}^i_{t,s,a} = \hat{\theta}_s^\top x^i_{t,a}$, $w^i_{t,s,a} \leftarrow \alpha_s \|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}, \forall s \in [0:S], \forall a \in [K]$.
$s \leftarrow 0$; $\mathcal{A}_0 \leftarrow \{a \in [K] \mid \hat{r}^i_{t,0,a} + w^i_{t,0,a} \geq \max_{a \in [K]}(\hat{r}^i_{t,0,a} - w^i_{t,0,a})\}$. ▷ Initial screening
repeat ▷ Layered successive screening
  if $s = S$ **then**
    Choose action a^i_t arbitrarily from $\mathcal{A}_S$
  else if $w^i_{t,s,a} \leq \overline{w}_s$ for all $a \in \mathcal{A}_s$ **then**
    $\mathcal{A}_{s+1} \leftarrow \{a \in \mathcal{A}_s \mid \hat{r}^i_{t,s,a} \geq \max_{a' \in \mathcal{A}_s}(\hat{r}^i_{t,s,a'}) - 2\overline{w}_s\}$; $s \leftarrow s+1$
  else
    $a^i_t \leftarrow \arg\max_{\{a \in \mathcal{A}_s, w^i_{t,s,a} > \overline{w}_s\}} w^i_{t,s,a}$
  end if
until action a^i_t is found
Take action a^i_t and and receive reward r^i_{t,a^i_t}
$\eta_t = \min\{1, \gamma/\|x^i_{t,a_t}\|_{(A^i_{t,s})^{-1}}\}$
$\Delta A^i_s \leftarrow \Delta A^i_s + \eta_t x^i_{t,a^i_t} x^{i\top}_{t,a^i_t}$, $\Delta b^i_s \leftarrow \Delta b^i_s + \eta_t r^i_{t,a^i_t} x^i_{t,a^i_t}$ ▷ Update local information
Return layer index s

Algorithm 8 Corruption Robust Async-`FedSupLinUCB`

Initialization: $T, C, S = \lceil \log d \rceil$
$\{A^{ser}_s \leftarrow I_d, b^{ser}_s \leftarrow 0 \mid s \in [0:S]\}$ ▷ Server initialization
$\{A^i_s \leftarrow I_d, \Delta A^i_s, b^i_s, \Delta b^i_s \leftarrow 0 \mid s \in [0:S], i \in [M]\}$ ▷ Clients initialization
for $t = 1, 2, \cdots, T$ **do**
  Client $i_t = i$ is active, and observes K contexts $\{x^i_{t,1}, x^i_{t,2}, \cdots, x^i_{t,K}\}$
  $s \leftarrow$ `CS-LUCB` $\big($client $i, \{x^i_{t,1}, x^i_{t,2}, \cdots, x^i_{t,K}\}\big)$ with lazy update
  if $\frac{\det(A^i_s + \Delta A^i_s)}{\det(A^i_s)} > (1 + C)$ **then**
    Sync(s, server, clients i) for each $s \in [0:S]$
  end if
end for

F.1 Algorithm

F.2 Supporting Lemmas and Proof

When confronted with adversarial corruption, we utilize a weighted ridge regression in which the weight assigned to each selected action depends on its confidence. Further, we expand the confidence width to accommodate this corruption, with $\alpha_0 = 1 + \sqrt{d\ln(2M^2T/\delta)} + \gamma C_p$ and $\alpha_s = 1 + \sqrt{2\ln(2KMT \ln d/\delta)} + \gamma C_p$ as proposed in He et al. (2022b). In our analysis of layer 0, we adapt Lemma B.1 from He et al. (2022b) to fit a federated scenario, yielding the following lemma:

Lemma F.1. *(Adapted from Lemma B.1 in He et al. (2022b)) Under the setting of Theorem 5.1, in the layer* 0*, with probability at least* $1 - \delta$*, the following event* $\mathcal{E}_0$ *happens:*

$$\mathcal{E}_0 \triangleq \{\left| x^{i\top}_{t,a} \hat{\theta}^i_{t,s} - x^{i\top}_{t,a}\theta \right| \leq w^i_{t,s,a}, \forall i \in [M], a \in [K], t \in [T], s = 0\}.$$

For each layer $s \in [S]$, the rewards are mutually independent, analogous to the proof of Lemma B.1. We can restate the lemma as follows:

Lemma F.2. *Suppose the time index set* $\Psi_{t,s}$ *is constructed so that for fixed* x_{τ,a_τ} *with* $\tau \in \Psi_{t,s}$*, the rewards* $\{r_{\tau,a_\tau}\}$ *are independent random variables with means* $\mathbb{E}[r_{\tau,a_\tau}] = \theta^\top x_{\tau,a_\tau} + c_\tau$*. For any round* $t \in [T]$*, if client* $i_t = i$ *is active and chooses arm* a_t *in layer* $s \in [S]$*, with probability at least*

$1 - \frac{\delta}{MT \ln d}$, *we have for any* $a_t \in [K]$*:*

$$\left|\hat{r}_{t,s,a_t} - \theta^\top x^i_{t,a_t}\right| \le w^i_{t,s,a_t} = \alpha_s \|x^i_{t,a_t}\|_{(A^i_{t,s})^{-1}}.$$

After combining the aforementioned events, we redefine the good event in the presence of corruption as follows:

$$\mathcal{E} \triangleq \left\{\left|x^{i\top}_{t,a}\hat{\theta}^i_{t,s} - x^{i\top}_{t,a}\theta\right| \le w^i_{t,s,a}, \forall i \in [M], a \in [K], t \in [T], s \in [0:S]\right\}.$$

Similar to proof of Lemma B.4, we have that $\mathbb{P}[\mathcal{E}] \ge 1 - \delta$.

Lemma F.3. *Conditioned on the good event* $\mathcal{E}$*, the regret of layer* $s \in [0 : S-1]$ *can be bounded as follows:* $\mathsf{reg}layers \le \tilde{O}(\sqrt{dT} + dC_p)$*.*

Proof of Lemma F.3. Under the condition of the good event $\mathcal{E}$, we adopt a similar approach to the regret decomposition analysis presented in He et al. (2022b) to bound the regret in each layer $s \in [0 : S-1]$.

$$\mathbb{E}\sum_{t\in\Psi_{T,s}}(r^i_{t,a^{i,*}_t} - r^i_{t,a_t}) \le \sum_{t\in\Psi_{T,s}} 8w_{t,s,a_t} = \sum_{t\in\Psi_{T,s}} 8\alpha_s\|x^i_{t,a}\|_{(A^i_{t,s})^{-1}} \tag{7}$$

$$= \underbrace{8\alpha_s \sum_{t\in\Psi_{T,s},\eta_t=1}\|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}}_{I_1} + \underbrace{8\alpha_s\sum_{t\in\Psi_{T,s},\eta_t<1}\|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}}_{I_2}. \tag{8}$$

The first inequality is derived from Lemma B.6, while Equation (8) follows from the definition of η_t. For the term I_1, we consider the rounds with $\eta_t = 1$, assuming these rounds can be listed as $\{k_1, k_2, ..., k_n\}$. To analyze this, we construct the auxiliary matrix $B_{t,s} = I + \sum_{j=1}^n x_{k_j}x_{k_j}^\top I\{k_j \le t\}$. Using the definition of $A^i_{t,s}$, we can establish the inequality $A^i_{t,s} \succeq \frac{1}{1+MC}A^{all}_{t,s} \succeq \frac{1}{1+MC}B_{t,s}$.

Then we have

$$\begin{aligned}I_1 &= \sum_{t\in\Psi_{T,s},\eta_t=1} 8\alpha_s\|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}\\ &\le 8\alpha_s\sqrt{2(1+MC)}\sqrt{2d|\Psi_{T,s}|\log|\Psi_{T,s}|} + 8\alpha_s dM\log(1+T/d) \le \tilde{O}(\sqrt{dT}),\end{aligned}$$

where the first inequality follows from Lemma C.3, and the second inequality is obtained by noting that the size of $\Psi_{T,0}$ is bounded by $\widetilde{O}(T/d)$, as stated in Lemma C.4 particularly for layer 0.

For the term I_2, using the property $\eta_t < 1$, we can express η_t as $\eta_t = \gamma/\|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}$, which implies:

$$\begin{aligned}I_2 &= \sum_{t\in\Psi_{T,s},\eta_t<1} 8\alpha_s\|x^i_{t,a}\|_{(A^i_{t,s})^{-1}}\\ &\le \sum_{t\in\Psi_{T,s},\eta_t<1} 8\frac{\alpha_s}{\gamma}\eta_t x^{i\top}_{t,a}(A^i_{t,s})^{-1}x^i_{t,a} \le \frac{\alpha_s}{\gamma}d\log(T) \le \tilde{O}(dC_p),\end{aligned}$$

where the first inequality is derived from the definition of η_t, the second inequality is obtained from the elliptical potential lemma, as referenced in Lemma A.1, and the third inequality stems from the definition of α_s.

By combining I_1 and I_2, we can ultimately bound the regret in each layer $s \in [0 : S-1]$ as $\mathsf{reg}_{\mathsf{layers}} \le \tilde{O}(\sqrt{dT} + dC_p)$. □

For the regret that occurs in the last layer S, we can derive the following bound:

$$\sum_{t\in\Psi_{T,S}}\mathbb{E}\left[r^i_{t,a^{i,*}_t} - r^i_{t,a_t}\right] \le \sum_{t\in\Psi_{T,S}} 8\overline{w}_S \le 8\overline{w}_S|\Psi_{T,S}| \le 8\overline{w}_S T \le 8\sqrt{dT}.$$

The first inequality is from Lemma B.6, and the last inequality follows from $\overline{w}_S = \sqrt{d/T}$.

Proof of the communication bound in Theorem 7.2. By combining the regret in each layer, we can conclude that $R_T \leq \tilde{O}(\sqrt{dT} + dC_p)$. Note that, based on the definition of $\eta_t \leq 1$ and Lemma A.1, it follows that $\log\bigl(\det\bigl(A_{t,s}^{all}\bigr)\bigr) \leq d\log(1 + |\Psi_{T,s}|/d)$. Additionally, by following a similar proof as in Lemma C.1, we can bound the number of communication rounds in layer s by $O(dM^2 \log T)$. Considering that the FedSupLinUCB algorithm has $S = \lceil \log d \rceil$ layers, the total communication cost is therefore upper bounded by $O(dM^2 \log d \log T)$.