Supplementary Material: "Communication-Efficient Federated Split Learning Relying on Low-Rank Approximation"

Huiqing Ao, Hui Tian, and Wanli Ni

Appendix A
Proof of Theorem 1

According to Assumption 1, the global loss function $F(\mathbf{w})$ is ℓ-smooth. Therefore, we have

$$\begin{aligned}
&\mathbb{E}[F(\mathbf{w}_{t+1})] - \mathbb{E}[F(\mathbf{w}_t)] \\
&\leq \mathbb{E}\langle \nabla F(\mathbf{w}_t), \mathbf{w}_{t+1} - \mathbf{w}_t\rangle + \frac{\ell}{2}\mathbb{E}[\|\mathbf{w}_{t+1} - \mathbf{w}_t\|^2] \\
&\leq \mathbb{E}\langle \nabla F(\mathbf{w}_t), \mathbf{w}_{t+1} - \tilde{\mathbf{w}}_t\rangle + \mathbb{E}\langle \nabla F(\mathbf{w}_t), \tilde{\mathbf{w}}_t - \mathbf{w}_t\rangle \\
&+ \ell\mathbb{E}[\|\mathbf{w}_{t+1} - \tilde{\mathbf{w}}_t\|^2] + \ell\mathbb{E}[\|\tilde{\mathbf{w}}_t - \mathbf{w}_t\|^2], \qquad (1)
\end{aligned}$$

The first term in (1) can be bounded as

$$\begin{aligned}
&\mathbb{E}\langle \nabla F(\mathbf{w}_t), \mathbf{w}_{t+1} - \tilde{\mathbf{w}}_t\rangle = -\eta\mathbb{E}\langle \nabla F(\mathbf{w}_t), \frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\rangle \\
&\overset{(i)}{=} -\frac{\eta}{2}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] - \frac{\eta}{2}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&+ \frac{\eta}{2}\mathbb{E}[\|\nabla F(\mathbf{w}_t) - \frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&\overset{(ii)}{\leq} -\frac{\eta}{2}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] - \frac{\eta}{2}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&+ \frac{\eta}{N}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F(\mathbf{w}_t) - \bar{\boldsymbol{g}}_{n,t}\|^2 + \|\bar{\boldsymbol{g}}_{n,t} - \sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&\overset{(iii)}{\leq} -\frac{\eta}{2}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] - \frac{\eta}{2}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&+ \frac{\eta}{N}\sum_{n=1}^{N}\zeta_n^2 + \eta H^2\vartheta^2 \qquad (2)
\end{aligned}$$

where (i) follows by using $\langle \boldsymbol{a}, \boldsymbol{b}\rangle = \frac{1}{2}\|\boldsymbol{a}\|^2 + \frac{1}{2}\|\boldsymbol{b}\|^2 - \frac{1}{2}\|\boldsymbol{a} - \boldsymbol{b}\|^2$, (ii) follows by using Jensen's inequality and $\|\boldsymbol{a}+\boldsymbol{b}\|^2 \leq 2\|\boldsymbol{a}\|^2 + 2\|\boldsymbol{b}\|^2$, and (iii) follows Assumptions 2 and 3.

The second term in (1) can be bounded as

$$\begin{aligned}
&\mathbb{E}\langle \nabla F(\mathbf{w}_t), \tilde{\mathbf{w}}_t - \mathbf{w}_t\rangle \\
&\overset{(i)}{\leq} \frac{1}{2C}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] + \frac{C}{2}\mathbb{E}[\|\tilde{\mathbf{w}}_t - \mathbf{w}_t\|^2], \qquad (3)
\end{aligned}$$

where (i) follows Cauchy-Schwarz inequality and AM-GM inequality, and $C > 0$ is a constant.

Plugging (2) and (3) into (1), we have

$$\begin{aligned}
&\mathbb{E}[F(\mathbf{w}_{t+1})] - \mathbb{E}[F(\mathbf{w}_t)] \\
&\leq -\frac{\eta - \frac{1}{C}}{2}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] + \frac{2\ell\eta^2 - \eta}{2N}\sum_{n=1}^{N}\mathbb{E}[\|\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&+ \frac{\eta}{N}\sum_{n=1}^{N}\zeta_n^2 + \eta H^2\vartheta^2 + (\frac{C}{2} + \ell)\mathbb{E}[\|\tilde{\mathbf{w}}_t - \mathbf{w}_t\|^2]. \qquad (4)
\end{aligned}$$

The term $\mathbb{E}[\|\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2]$ in (4) can be bounded as

$$\begin{aligned}
&\mathbb{E}[\|\sum_{\tau=0}^{H-1}\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \leq H\sum_{\tau=0}^{H-1}\mathbb{E}[\|\tilde{\boldsymbol{g}}_{n,t,\tau}\|^2] \\
&\leq H\sum_{\tau=0}^{H-1}(2\mathbb{E}[\|\tilde{\boldsymbol{g}}_{n,t,\tau} - \boldsymbol{g}_{n,t,\tau}\|^2] + 2\mathbb{E}[\|\boldsymbol{g}_{n,t,\tau}\|^2]) \\
&= 2H\sum_{\tau=0}^{H-1}(\mathbb{E}[\|\boldsymbol{g}^{\mathrm{d}}_{n,t,\tau} - \tilde{\boldsymbol{g}}^{\mathrm{d}}_{n,t,\tau}\|^2] + \mathbb{E}[\|\boldsymbol{g}^{\mathrm{s}}_{n,t,\tau} - \tilde{\boldsymbol{g}}^{\mathrm{s}}_{n,t,\tau}\|^2]) \\
&+ 2H\sum_{\tau=0}^{H-1}\mathbb{E}[\|\boldsymbol{g}_{n,t,\tau}\|^2] \\
&\overset{(i)}{\leq} 2H^2(\Gamma^{\mathrm{d}}_{n,t} + \Gamma^{\mathrm{s}}_{n,t} + G^2), \qquad (5)
\end{aligned}$$

where (i) follows Assumption 4.

Suppose $\eta \geq \frac{1}{2\ell} > 0$ and subscribe (5) into (4), we have

$$\begin{aligned}
&\mathbb{E}[F(\mathbf{w}_{t+1})] - \mathbb{E}[F(\mathbf{w}_t)] \leq -\frac{\eta - \frac{1}{C}}{2}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] \\
&+ \frac{(2\ell\eta^2 - \eta)H^2}{N}\sum_{n=1}^{N}(\Gamma^{\mathrm{d}}_{n,t} + \Gamma^{\mathrm{s}}_{n,t} + G^2) \\
&+ \frac{\eta}{N}\sum_{n=1}^{N}\zeta_n^2 + \eta H^2\vartheta^2 + (\frac{C}{2} + \ell)\frac{1}{N}\sum_{n=1}^{N}\Gamma^{\mathrm{m}}_{n,t}. \qquad (6)
\end{aligned}$$

Recursively utilizing the above inequality from $t = 0$ to $t = T - 1$ and then averaging yields

$$\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla F(\mathbf{w}_t)\|^2] \\
&\leq \frac{2(\mathbb{E}[F(\mathbf{w}_0)] - \mathbb{E}[F(\mathbf{w}_T)])}{T(\eta - \frac{1}{C})} + \frac{2}{N(1 - \frac{1}{\eta C})}\sum_{n=1}^{N}\zeta_n^2 + \frac{2H^2\vartheta^2}{1 - \frac{1}{\eta C}} \\
&+ \frac{2H^2(2\ell\eta^2 - \eta)}{NT(\eta - \frac{1}{C})}\sum_{t=0}^{T-1}\sum_{n=1}^{N}(\Gamma^{\mathrm{d}}_{n,t} + \Gamma^{\mathrm{s}}_{n,t}) \\
&+ \frac{2H^2G^2(2\ell\eta^2 - \eta) + (C + 2\ell)\frac{1}{N}\sum_{n=1}^{N}\Gamma^{\mathrm{m}}_{n,t}}{\eta - \frac{1}{C}}, \qquad (7)
\end{aligned}$$

H. Ao and H. Tian are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: hqao@bupt.edu.cn; tianhui@bupt.edu.cn).

W. Ni is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: niwanli@tsinghua.edu.cn).

where $\eta > \max\{\frac{1}{2\ell}, \frac{1}{C}\}$. To this end, Theorem 1 is proved.

APPENDIX B
DETAILED PROCESS FOR POWER ALLOCATION

Given computation capacity, decoding order, and split layer selection, the power allocation sub-problem can be expressed as

$$\begin{aligned}\min_{\mathbf{\Phi}_1} \quad & \rho(\max_n\{\tilde{T}_n^{\mathrm{D}}\} + H(\max_n\{T_n^{\mathrm{SD}}\} + \max_n\{T_n^{\mathrm{G}}\}) \\ & + \max_n\{T_n^{\mathrm{U}}\}) + (1-\rho)(\sum\nolimits_{n=1}^{N}(p_n T_n^{\mathrm{D}} + \sum\nolimits_{k=1}^{2} p'_{n,k} T_{n,k}^{\mathrm{U}} \\ & + H(\sum\nolimits_{k=1}^{2} p_{n,k} T_{n,k}^{\mathrm{SD}} + p'_n T_n^{\mathrm{G}})) + p_0 T_0^{\mathrm{D}}) \quad (8a)\\ \text{s.t.}\ & (23d) - (23g), \quad (8b)\end{aligned}$$

where $\mathbf{\Phi}_1 = \{p_n, p_{n,k}, p'_n, p'_{n,k} \mid \forall n, k\}$. Problem (8) is non-convex owing to the non-convexity of objective function. Fortunately, the variables are independent and their expressions have similar structures. Due to space constraints, we only address one variable as an example and the others follow similarly. For $\{p_n\}$, some variables δ_1 and δ_2 are introduced to replace $\max_n\{\tilde{T}_n^{\mathrm{D}}\}$ and $\sum_{n=1}^{N} p_n T_n^{\mathrm{D}}$. Afterwards, we adopt successive convex approximation to tackle them as convex constraints. Specifically, we can introduce slack variable set $\{\beta_{1n}\}$ and obtain

$$\sum_{n=1}^{N} \frac{p_n \sum_{l_n=1}^{L-1} m_{n,l_n} \xi_{n,l_n}^{\mathrm{D}}}{\beta_{1n}} \leq \delta_2, \quad (9)$$

$$\frac{\sum_{l_n=1}^{L-1} m_{n,l_n} \xi_{n,l_n}^{\mathrm{D}}}{\beta_{1n}} \leq \delta_1, \quad (10)$$

$$B\log_2(1 + \frac{p_n|h_n|^2}{\sum_{k=1,k\neq n}^{N} p_k|h_n|^2 + \sigma_n^2}) \geq \beta_{1n}. \quad (11)$$

Clearly, (10) and the right hand side of (9) are convex. By applying the first-order Taylor expansion to replace the left hand side of (9), (9) can be approximated as

$$\begin{aligned}\delta_2 \geq \sum_{n=1}^{N}\sum_{l_n=1}^{L-1} m_{n,l_n}\xi_{n,l_n}^{\mathrm{D}}(&\frac{p_n^{(j)}}{\beta_{1n}^{(j)}} + \frac{1}{\beta_{1n}^{(j)}}(p_n - p_n^{(j)}) \\ &- \frac{p_n^{(j)}}{(\beta_{1n}^{(j)})^2}(\beta_{1n} - \beta_{1n}^{(j)})), \quad (12)\end{aligned}$$

where the superscript (j) is the value of the variable in iteration j. By introducing an auxiliary variable set $\{\beta_{2n}\}$, (11) can be transformed as

$$B\log_2(1+\beta_{2n}) \geq \beta_{1n}, \quad (13)$$

$$\frac{p_n|h_n|^2}{\sum_{k=1,k\neq n}^{N} p_k|h_n|^2 + \sigma_n^2} \geq \beta_{2n}. \quad (14)$$

For (14), it can be equivalent to

$$\frac{p_n|h_n|^2}{\beta_{3n}} \geq \beta_{2n}, \quad (15)$$

$$\sum_{k=1,k\neq n}^{N} p_k|h_n|^2 + \sigma_n^2 \leq \beta_{3n}, \quad (16)$$

where β_{3n} is a non-negative slack variable. The (15) can be further equivalent to

$$|h_n|^2 \geq \frac{\beta_{2n}\beta_{3n}}{p_n}. \quad (17)$$

By applying the first-order Taylor expansion to replace the right hand side of (17), it can be approximated as

$$\begin{aligned}|h_n|^2 \geq \frac{\beta_{2n}^{(j)}\beta_{3n}^{(j)}}{p_n^{(j)}} + \frac{\beta_{3n}^{(j)}}{p_n^{(j)}}(\beta_{2n} - \beta_{2n}^{(j)}) + \frac{\beta_{2n}^{(j)}}{p_n^{(j)}}(\beta_{3n} - \beta_{3n}^{(j)}) \\ - \frac{\beta_{2n}^{(j)}\alpha_{3n}^{(j)}}{(p_n^{(j)})^{(2)}}(p_n - p_n^{(j)}). \quad (18)\end{aligned}$$

Now, we complete the non-convex process for optimization variable $\{p_n\}$. The other variables can be tackled in the same manner.