Supplemental Material for *Asynchronous Personalized Federated Learning with Irregular Clients*

Editors: Emtiyaz Khan and Mehmet Gonen

1. Complete Proofs

1.1. Proof of Theorem 1

Suppose the virtual sequence of global models $\overline{\omega}^\tau$ for $\tau = 1, 2, ..., T$ is computed by

$$\overline{\omega}^\tau = \omega^0 - \sum_{l=1}^{M} p^l \sum_{i=0}^{\tau-1} \eta_i \nabla F_k(\omega_k^i, \xi_k^i). \tag{9}$$

We define $g^\tau = \sum_{k=1}^M p^k \nabla F_k(\omega_k^\tau, \xi_k^\tau)$ and $\overline{g}^\tau = \sum_{k=1}^M p^k \nabla F_k(\omega_k^\tau)$, and we have $\mathbb{E}[g^\tau] = \overline{g}^\tau$. Consequently, based on Equation 2, we have $\overline{\omega}^{\tau+1} = \overline{\omega}^\tau - \eta_\tau g^\tau$. Then it holds that

$$\begin{aligned}
||\overline{\omega}^{\tau+1} - \omega^*||^2 &= ||\overline{\omega}^\tau - \eta_\tau g^\tau - \omega^*||^2 \\
&= ||\overline{\omega}^\tau - \eta_\tau g^\tau - \omega^* - \eta_\tau \overline{g}^\tau + \eta_\tau \overline{g}^\tau||^2 \\
&= ||\overline{\omega}^\tau - \omega^* - \eta_\tau \overline{g}^\tau + \eta_\tau(\overline{g}^\tau - g^\tau)||^2 \\
&= \underbrace{||\overline{\omega}^\tau - \omega^* - \eta_\tau \overline{g}^\tau||^2}_{A_1} + \underbrace{\eta_\tau^2 ||\overline{g}^\tau - g^\tau||^2}_{A_2} + \underbrace{2\eta_\tau < \overline{\omega}^\tau - \omega^* - \eta_\tau \overline{g}^\tau, \overline{g}^\tau - g^\tau >}_{A_3}.
\end{aligned} \tag{10}$$

To bound Equation (10), we separately bound A_1, A_2 and A_3.

$$A_1 = ||\overline{\omega}^\tau - \omega^*||^2 + \underbrace{\eta_\tau^2 ||\overline{g}^\tau||^2}_{B_1} \underbrace{- 2\eta_\tau < \overline{\omega}^\tau - \omega^*, \overline{g}^\tau >}_{B_2}. \tag{11}$$

Given Assumption 1, for any ω and $\omega^{'}$, we have

$$\begin{aligned}
F(\omega) &= F(\omega^{'}) + \int_0^1 \frac{dF(\omega^{'} + t(\omega - \omega^{'}))}{dt} dt \\
&= F(\omega^{'}) + \int_0^1 \nabla F(\omega^{'} + t(\omega - \omega^{'}))^\top (\omega - \omega^{'}) dt \\
&= F(\omega^{'}) + \nabla F(\omega^{'})^\top (\omega - \omega^{'}) + \int_0^1 [\nabla F(\omega^{'} + t(\omega - \omega^{'})) - \nabla F(\omega^{'})]^\top (\omega - \omega^{'}) dt \\
&\leq F(\omega^{'}) + \nabla F(\omega^{'})^\top (\omega - \omega^{'}) + \int_0^1 L||t(\omega - \omega^{'})|| ||\omega - \omega^{'}|| dt \\
&= F(\omega^{'}) + \nabla F(\omega^{'})^\top (\omega - \omega^{'}) + \frac{1}{2} L ||\omega - \omega^{'}||^2,
\end{aligned} \tag{12}$$

and

$$||\nabla F_k(\omega_k^\tau)||^2 \leq 2L(F_k(\omega_k^\tau) - F_k(\omega^*)). \tag{13}$$

Thus, B_1 is bounded by

$$B_1 \leq \eta_\tau^2 \sum_{k=1}^{M} p^k ||\nabla F_k(\omega_k^\tau)||^2 \leq 2L\eta_\tau^2 \sum_{k=1}^{M} p^k (F_k(\omega_k^\tau) - F_k(\omega^*)). \tag{14}$$

B_2 can be reformulated by

$$\begin{aligned} B_2 &= -2\eta_\tau \sum_{k=1}^{M} p^k < \overline{\omega}^\tau - \omega^*, g_k^\tau > \\ &= -2\eta_\tau \sum_{k=1}^{M} p^k < \overline{\omega}^\tau - \omega_k^\tau, g_k^\tau > -2\eta_\tau \sum_{k=1}^{M} p^k < \omega_k^\tau - \omega^*, g_k^\tau >, \end{aligned} \tag{15}$$

and we have

$$\begin{aligned} -2\mathbb{E}[< \overline{\omega}^\tau - \omega_k^\tau, g_k^\tau >] &= -2 < \overline{\omega}^\tau - \omega_k^\tau, \nabla F_k(\omega_k^\tau) > \\ &\leq 2||\overline{\omega}^\tau - \omega_k^\tau|| ||\nabla F_k(\omega_k^\tau)|| \\ &\leq \frac{1}{\eta_\tau} ||\overline{\omega}^\tau - \omega_k^\tau||^2 + \eta_\tau ||\nabla F_k(\omega_k^\tau)||^2. \end{aligned} \tag{16}$$

The first inequality holds for Cauchy-Schwarz inequality and the second holds for AM-GM inequality. According to Equations (13) and (16), we have

$$-2\mathbb{E}[< \overline{\omega}^\tau - \omega_k^\tau, g_k^\tau >] \leq \frac{1}{\eta_\tau} ||\overline{\omega}^\tau - \omega_k^\tau||^2 + 2L\eta_\tau (F_k(\omega_k^\tau) - F_k(\omega^*)), \tag{17}$$

and

$$-2\mathbb{E}[< \omega_k^\tau - \omega^*, g_k^\tau >] \leq -2(F_k(\omega_k^\tau) - F_k(\omega^*)) - \mu||\omega_k^\tau - \omega^*||^2. \tag{18}$$

Thus, B_2 can be bounded by

$$\mathbb{E}[B_2] \leq \sum_{k=1}^{M} p^k (||\overline{\omega}^\tau - \omega_k^\tau||^2 + 2L\eta_\tau^2 (F_k(\omega_k^\tau) - F_k(\omega^*))) - \eta_\tau \sum_{k=1}^{M} p^k (2(F_k(\omega_k^\tau) - F_k(\omega^*)) + \mu||\omega_k^\tau - \omega^*||^2). \tag{19}$$

According to the convexity property of quadratic function, we have

$$-\eta_\tau \mu \sum_{k=1}^{M} p^k ||\omega_k^\tau - \omega^*||^2 \leq -\eta_\tau \mu ||\overline{\omega}^\tau - \omega^*||^2. \tag{20}$$

Substituting Equation (20) into Equation (19), we bound B_2 by

$$\mathbb{E}[B_2] \leq \sum_{k=1}^{M} p^k (||\overline{\omega}^\tau - \omega_k^\tau||^2 + 2L\eta_\tau^2 (F_k(\omega_k^\tau) - F_k(\omega^*))) - 2\eta_\tau \sum_{k=1}^{M} p^k (F_k(\omega_k^\tau) - F_k(\omega^*)) - \eta_\tau \mu ||\overline{\omega}^\tau - \omega^*||^2. \tag{21}$$

Based on Equations (14) and (21), A_1 can be reformulated as

$$\begin{aligned}\mathbb{E}[A_1] \leq (1-\eta_\tau\mu)||\overline{\omega}^\tau-\omega^*||^2+\sum_{k=1}^{M}p^k||\overline{\omega}^\tau-\omega_k^\tau||^2 \\ +\underbrace{4L\eta_\tau^2\sum_{k=1}^{M}p^k(F_k(\omega_k^\tau)-F_k^*)-2\eta_\tau\sum_{k=1}^{M}p^k(F_k(\omega_k^\tau)-F_k(\omega^*))}_{C}.\end{aligned} \tag{22}$$

Let $\gamma_\tau = 2\eta_\tau(1-2L\eta_\tau)$ for $\tau = 1,2,...,T$. C can be reformulated as

$$\begin{aligned}C &= -\gamma_\tau\sum_{k=1}^{M}p^k(F_k(\omega_k^\tau)-F(\omega^*))+(2\eta_\tau-\gamma_\tau)\sum_{k=1}^{M}p^k(F^*-F_k^*) \\ &= -\gamma_\tau\underbrace{\sum_{k=1}^{M}p^k(F_k(\omega_k^\tau)-F^*)}_{D}+4L\eta_\tau^2\Gamma,\end{aligned} \tag{23}$$

where

$$\begin{aligned}D &= \sum_{k=1}^{M}p^k(F_k(\omega_k^\tau)-F(\overline{\omega}^\tau))+\sum_{k=1}^{M}p^k(F_k(\overline{\omega}^\tau)-F^* \\ &\geq \underbrace{\sum_{k=1}^{M}p^k<\nabla F_k(\overline{\omega}^\tau),\overline{\omega}_k^\tau-\overline{\omega}^\tau>}_{D_1}+F(\overline{\omega}^\tau)-F^*.\end{aligned} \tag{24}$$

Given Assumption 1, we have

$$D_1 \geq -\frac{1}{2}\sum_{k=1}^{M}p^k(\eta_\tau||\nabla F_k(\overline{\omega}^\tau)||^2+\frac{1}{\eta_\tau}||\omega_k^\tau-\overline{\omega}^\tau||^2). \tag{25}$$

Substituting Equations (25) and (24) into (23), we have

$$\begin{aligned}C &= \gamma_\tau(L\eta_\tau-1)\sum_{k=1}^{M}p^k(F_k(\omega^\tau)-F_k(\omega^*))+(4L\eta_\tau^2+\gamma_\tau\eta_\tau L)\Gamma+\frac{\gamma_\tau}{2\eta_\tau}\sum_{k=1}^{M}p^k||\overline{\omega}^\tau-\omega_k^\tau||^2 \\ &\leq 6L\eta_\tau^2\Gamma+\sum_{k=1}^{M}p^k||\overline{\omega}^\tau-\omega_k^\tau||^2.\end{aligned} \tag{26}$$

According to Equations (22) and (26), A_1 can be bounded by

$$\mathbb{E}[A_1] \leq (1-\mu\eta_\tau)||\overline{\omega}^\tau-\omega^*||^2+\underbrace{2\sum_{k=1}^{M}p^k||\overline{\omega}^\tau-\omega_k^\tau||^2}_{D_2}+6L\eta_\tau^2\Gamma. \tag{27}$$

To bound D_2, we define the auxiliary inequality as

$$\begin{aligned} g_k^{\tau+1} &= ||g_k^{\tau+1} - g_k^\tau + g_k^\tau|| \leq ||g_k^{\tau+1} - g_k^\tau|| + ||g_k^\tau|| \\ &= ||\nabla F_k(\omega_K^\tau - \eta_\tau g_k^\tau) - \nabla F_k(\omega_K^\tau)|| + ||g_k^\tau|| \\ &\leq L||\omega_K^\tau - \eta_\tau g_k^\tau - \omega_K^\tau|| + ||g_k^\tau|| \\ &= (1 + L\eta_\tau)||g_k^\tau|| = (1 + L\eta_\tau)^\tau ||g_k^0||, \end{aligned} \tag{28}$$

and τ_{k_0} as the last time step when k-th client upload its model to the server. Given the fixed η, we have

$$\begin{aligned} ||\overline{\omega}^\tau - \omega_k^\tau|| = ||(\omega_k^\tau - \omega^{\tau_{k_0}}) - (\overline{\omega}^\tau - \omega^{\tau_{k_0}})|| &\leq \sum_{i=\tau_{k_0}}^{\tau} ||\eta(g_k^i - \sum_{j=1}^{M} p^j g_j^i)|| \\ &\leq \sum_{i=\tau_{k_0}}^{\tau} (1 + L\eta)^{i-\tau_{k_0}} ||\eta(g_k^{\tau_{k_0}} - \sum_{j=1}^{M} g_j^{\tau_{k_0}})|| \\ &\leq ||\frac{\sum_{i=\tau_{k_0}}^{\tau} (1 + L\eta)^{i-\tau_{k_0}} - 1}{L} (g_k^{\tau_{k_0}} - \sum_{j=1}^{M} g_j^{\tau_{k_0}})||. \end{aligned} \tag{29}$$

Equation (29) can be reformulated as

$$||\overline{\omega}^\tau - \omega_k^\tau|| \leq ||\frac{\sum_{i=\tau_{k_0}}^{\tau} (1 + L\eta)^{\pi_k E - 1}}{L} (g_k^{\tau_{k_0}} - \sum_{j=1}^{M} g_j^{\tau_{k_0}})||. \tag{30}$$

Given Assumption 4, we have

$$||\overline{\omega}^\tau - \omega_k^\tau|| \leq ||\frac{\sum_{i=\tau_{k_0}}^{\tau} (1 + L\eta)^{i-\tau_{k_0}} - 1}{L} (g_k^{\tau_{k_0}} - \sum_{j=1}^{M} g_j^{\tau_{k_0}})||, \tag{31}$$

and according to Taylor expansion,

$$||\overline{\omega}^\tau - \omega_k^\tau|| \leq \eta^2 (\pi_k E - 1)^2 \chi^2 \leq \eta^2 E(\pi_k^2 E - \pi_k)\chi^2. \tag{32}$$

Thus, D_2 can be bounded by

$$D_2 \leq 2\eta_\tau^2 E \sum_{k=1}^{M} p^k (\pi_k^2 E - \pi_k)\chi^2. \tag{33}$$

Substituting Equation (33) into (27), A_1 can be counded by

$$\mathbb{E}[A_1] \leq (1 - \mu\eta_\tau)||\overline{\omega}^\tau - \omega^*||^2 + 2\eta_\tau^2 E \sum_{k=1}^{M} p^k (\pi_k^2 E - \pi_k)\chi^2 + 6L\eta_\tau^2 \Gamma. \tag{34}$$

Then, $A_2 = \eta_\tau^2 ||\overline{g}^\tau - g^\tau||^2$ in Equation (10) can be bounded by

$$\begin{aligned} \mathbb{E}[||g^\tau - \overline{g}^\tau||^2] &= \mathbb{E}[||\sum_{k=1}^{M} p^k (\nabla F_k(\omega_k^\tau) - g_k^\tau)||^2] \\ &\leq \sum_{k=1}^{M} p^{2k} \mathbb{E}[||\nabla F_k(\omega_k^\tau) - g_k^\tau||^2] \leq \sum_{k=1}^{M} p^{2k} \sigma_k^2, \end{aligned} \tag{35}$$

and $A_3 = 2\eta_\tau < \overline{\omega}^\tau - \omega^* - \eta_\tau \overline{g}^\tau, \overline{g}^\tau - g^\tau >$ can be bounded by $\mathbb{E}[A_3] = 0$ for $\mathbb{E}[g^\tau] = \overline{g}^\tau$.

Let $\Delta_{\tau+1} = \mathbb{E}[||\overline{\omega}^{\tau+1} - \omega^*||^2]$ and combining the obtained bounds of A_1, A_2 and A_3, we have

$$\Delta_{\tau+1} \leq (1 - \mu\eta_\tau)\Delta_\tau + \eta_\tau^2 A^{'}, \tag{36}$$

where $A^{'} = \sum_{k=1}^{M} p^{2k}\sigma_k^2 + 6L\Gamma + 2E\sum_{k=1}^{M} p^k(\pi_k^2 E - \pi_k)\chi^2$. Equation (36) can be further reformulated as

$$\Delta_{\tau+1} - \frac{\eta_\tau A^{'}}{\mu} \leq (1 - \mu\eta_\tau)(\Delta_\tau - \frac{\eta_\tau A^{'}}{\mu}). \tag{37}$$

Thus we have

$$\Delta_{\tau+1} - \frac{\eta A^{'}}{\mu} \leq (1 - \mu\eta)^{\tau+1}(\Delta_0 - \frac{\eta A^{'}}{\mu}). \tag{38}$$

Given Assumption 1, the convergence bound of AsyPFL with fixed learning rate is

$$\mathbb{E}[F(\overline{\omega}^\tau)] - F^* \leq \frac{L}{2}\Delta_\tau. \tag{39}$$

1.2. Proof of Theorem 2

The proof of Theorem 2 follows from the proof of Theorem 1. Starting from inequality (28), using the decayed learning rate in D_2, we have

$$\begin{aligned} ||\overline{\omega}^\tau - \omega_k^\tau|| = ||(\omega_k^\tau - \omega^{\tau_{k_0}}) - (\overline{\omega}^\tau - \omega^{\tau_{k_0}})|| &\leq \sum_{i=\tau_{k_0}}^{\tau} ||\eta_\tau(g_k^i - \sum_{j=1}^{M} g_j^i)|| \\ &\leq ||\frac{(1 + L\eta_{\tau_{k_0}})^{\pi_k E - 1} - 1}{L}(g_k^{\tau_{k_0}} - \sum_{j=1}^{M} g_j^{\tau_{k_0}})||. \end{aligned} \tag{40}$$

Given Assumption 4, we have

$$||\overline{\omega}^\tau - \omega_k^\tau||^2 \leq (\frac{(1 + L\eta_{\tau_{k_0}})^{\pi_k E - 1} - 1}{L})^2\chi^2. \tag{41}$$

According to the Taylor expansion, Equation (41) can be reformulated as

$$||\overline{\omega}^\tau - \omega_k^\tau||^2 \leq 4\eta_\tau^2(\pi_k E - 1)^2\chi^2 \leq 4\eta_\tau^2 E(\pi_k^2 E - \pi_k)\chi^2. \tag{42}$$

Thus D_2 can be bounded by

$$D_2 \leq 8\eta_\tau^2 E\sum_{k=1}^{M} p^k(\pi_k^2 E - \pi_k)\chi^2. \tag{43}$$

Substituting the bound of D_2 into Equation (34), we have

$$\Delta_{\tau+1} \leq (1 - \mu\eta_\tau)\Delta_\tau + \eta_\tau^2 A^{'}. \tag{44}$$

1.3. Proof of Theorem 3

Suppose the error bound of AsyPFL satisfies

$$\mathbb{E}[\omega^{T_{\min}})] - F^* \leq \frac{2L}{\mu(\gamma + T_{\min})^{2/3}}(\frac{A}{\mu} + 2L||\omega^0 - \omega^*||) \leq \epsilon. \tag{45}$$

Equation (45) can be reformulated as

$$T_{\min} + \gamma \geq \frac{2L}{\mu\epsilon}(\frac{A}{\mu} + 2L||\omega^0 - \omega^*||), \tag{46}$$

which is equivalent to

$$T_{\min} \geq \frac{2L}{\mu\epsilon}(\frac{\sum_{k=1}^{M} p^{2k}\sigma_k^2 + 6L\Gamma}{\mu} + 2L||\omega^0 - \omega^*||) + \frac{16L}{\mu^2\epsilon}E\sum_{k=1}^{M} p^k(\pi_k^2 E - \pi_k)\chi^2 - \gamma. \tag{47}$$

Thus, the required communication rounds are

$$R_{\min}(E) = \frac{T_{\min}}{E} \geq \frac{2L}{E\mu\epsilon}(\frac{\sum_{k=1}^{M} p^{2k}\sigma_k^2 + 6L\Gamma}{\mu} + 2L||\omega^0 - \omega^*||) + \frac{16L}{\mu^2\epsilon}\sum_{k=1}^{M} p^k(\pi_k^2 E - \pi_k)\chi^2 - \gamma. \tag{48}$$

According to Equation (48), we define the required communication rounds as a function of local epochs E by

$$R(E) = (\frac{16L}{\epsilon\mu^2}\chi^2\sum_{k=1}^{M} p^k\pi_k^2)E + \frac{V_\epsilon + \frac{4L^2}{\epsilon\mu}||\omega^0 - \omega^*||}{E} - (\frac{16L}{\epsilon\mu^2}\sum_{k=1}^{M} p^k\pi_k\chi^2 + \gamma), \tag{49}$$

which is in the form of

$$R(E) = aE + \frac{b}{E} - c. \tag{50}$$

Taking the derivative w.r.t. E and let the resulting function equal to zero, we have

$$\frac{dR(E)}{dE} = a - \frac{b}{E^2} = 0, \tag{51}$$

and $E{=}\sqrt{\frac{b}{a}}$.

1.4. Proof of Theorem 4

ΔT is minimized if C_{π_k} achieves its minimum. Thus, according to the definition of C_{π_k} and equation (47), we have

$$C_{\pi_k} = \Delta T_k^g Z_k + \frac{16L}{\epsilon\mu^2}\Delta T_k^g \chi^2 p^k (E\pi_k)^2 + \Delta T_k^c \frac{Z_k}{E\pi_k} - \frac{16L}{\epsilon\mu^2}\Delta T_k^c \chi^2 p^k. \tag{52}$$

Taking the derivative w.r.t. π_k and let the resulting equation equal to zero, we have

$$2\frac{\Delta T_k^g}{\Delta T_k^c}\frac{16L}{\epsilon\mu^2}E^2\chi^2 p^k \pi_k = \frac{Z_k}{E\pi_k^2}. \tag{53}$$

Rearranging Equation (53) to obtain the result in Theorem 4.