A Appendix

A.1 Proof of Theorem 1

During the Procrustes alignment, we define $\mathbf{Z}^{t+1} = (\mathbf{U}^{t+1})^\intercal \mathbf{S}^{t+1} \mathbf{V}^{t+1} \in \mathbb{R}^{r' \times r'}$. Therefore, $\mathbf{S}^{t+1}$ satisfies

$$\mathbf{S}^{t+1} = \mathbf{U}^{t+1} \mathbf{Z}^{t+1} (\mathbf{V}^{t+1})^\intercal. \tag{1}$$

Note that $(\mathbf{S}^{t+1})^\intercal \mathbf{S}^{t+1} = \mathbf{I}_{r'}$. In addition, since matrices $\mathbf{U}^{t+1}$ and $\mathbf{V}^{t+1}$ have orthogonal columns, we have

$$\begin{aligned}
\mathrm{Tr}\left(\mathbf{A}^t (\tilde{\mathbf{A}}^{t+1})^\intercal (\mathbf{S}^t)^\intercal\right) &= \mathrm{Tr}\left(\mathbf{U}^{t+1} \mathbf{\Sigma}^{t+1} (\mathbf{V}^{t+1})^\intercal \mathbf{V}^{t+1} (\mathbf{Z}^{t+1})^\intercal (\mathbf{U}^{t+1})^\intercal\right) \\
&= \mathrm{Tr}\left(\mathbf{\Sigma}^{t+1} (\mathbf{Z}^{t+1})^\intercal\right) \\
&= \sum_{j=1}^{r'} \sigma_j^{t+1} z_{j,j}^{t+1}.
\end{aligned} \tag{2}$$

Since matrix $\mathbf{Z}^{t+1}$ satisfies $(\mathbf{Z}^{t+1})^\intercal \mathbf{Z}^{t+1} = \mathbf{I}_{r'}$, each element $z_{j,j}^{t+1}$ yields $z_{j,j}^{t+1} \leq 1$, we have $\mathrm{Tr}\left(\mathbf{A}^t (\tilde{\mathbf{A}}^{t+1})^\intercal (\mathbf{S}^t)^\intercal\right) \leq \sum_{j=1}^{r'} \sigma_j^{t+1}$, with the equality achieved at $\mathbf{Z}^t = \mathbf{I}_{r'}$. Therefore, when the maximum objective is achieved, matrix $\mathbf{S}^{t,\star}$ satisfies

$$\mathbf{S}^{t,\star} = \mathbf{U}^{t+1} \mathbf{I}_{r'} (\mathbf{V}^{t+1})^\intercal = \mathbf{U}^{t+1} (\mathbf{V}^{t+1})^\intercal. \tag{3}$$

This completes the proof of Theorem 1.

A.2 Algorithm for FLoRG

Algorithm 1 FLoRG

Input: Local fine-tuning datasets $\mathcal{D}_n$, $n \in \mathcal{N}$; learning rate η; pretrained model $\mathbf{W}^0$; initialized low-rank matrix $\mathbf{A}^1$.
The central server initializes $\mathbf{L}$ and $\mathbf{R}$ with orthogonal columns and broadcasts them to all clients.
$\mathbf{W}^1 := \mathbf{W}^0 + \mathbf{L}(\mathbf{A}^1)^\intercal \mathbf{A}^1 \mathbf{R}$.
For $t \in \mathcal{T}$ **do**
    The central server broadcasts $\mathbf{A}^t$ to all clients.
    For client $n \in \mathcal{N}$ **in parallel do**
        $\mathbf{A}_n^{t+\frac{1}{2}} := \mathbf{A}^t - \eta \nabla_{\mathbf{A}} F_n(\mathbf{W}^t; \boldsymbol{\xi}_n)$.
        Transmits $\mathbf{A}_n^{t+\frac{1}{2}}$ to the central server.
    End for
    $\mathbf{Q}^{t+1} := \frac{1}{N} \sum_{n \in \mathcal{N}} \left(\mathbf{A}_n^{t+\frac{1}{2}}\right)^\intercal \mathbf{A}_n^{t+\frac{1}{2}}$.
    Perform eigendecomposition to $\mathbf{Q}^{t+1}$ and obtain $\tilde{\mathbf{A}}^{t+1} := (\Lambda^{t+1})^{\frac{1}{2}} \mathbf{P}^{t+1}$.
    Perform Procrustes alignment to $\tilde{\mathbf{A}}^{t+1}$ and obtain $\mathbf{S}^{t,\star} := \mathbf{U}^{t+1} (\mathbf{V}^{t+1})^\intercal$.
    $\mathbf{A}^{t+1} := \mathbf{S}^{t,\star} \tilde{\mathbf{A}}^{t+1}$.
    $\bar{\mathbf{W}}^{t+1} := \mathbf{W}^0 + \mathbf{L}(\mathbf{A}^{t+1})^\intercal \mathbf{A}^{t+1} \mathbf{R}$.
End for
Output: Average fine-tuned model $\bar{\mathbf{W}}^{T+1}$.

A.3 Proof of Lemma 1

$$\begin{aligned}
&\left\langle \mathbf{H}^t,\ (\mathbf{A}^t)^\top \mathbf{A}^t \left(\mathbf{H}^t + (\mathbf{H}^t)^\top\right) + \left(\mathbf{H}^t + (\mathbf{H}^t)^\top\right)(\mathbf{A}^t)^\top \mathbf{A}^t \right\rangle_F \\
\overset{\text{(a)}}{=}\ & \left\langle \tfrac{\mathbf{H}^t + (\mathbf{H}^t)^\top}{2},\ (\mathbf{A}^t)^\top \mathbf{A}^t \left(\mathbf{H}^t + (\mathbf{H}^t)^\top\right) + \left(\mathbf{H}^t + (\mathbf{H}^t)^\top\right)(\mathbf{A}^t)^\top \mathbf{A}^t \right\rangle_F \\
=\ & 2\left\langle \tfrac{\mathbf{H}^t + (\mathbf{H}^t)^\top}{2},\ (\mathbf{A}^t)^\top \mathbf{A}^t \tfrac{\mathbf{H}^t + (\mathbf{H}^t)^\top}{2} \right\rangle_F + 2\left\langle \tfrac{\mathbf{H}^t + (\mathbf{H}^t)^\top}{2},\ \tfrac{\mathbf{H}^t + (\mathbf{H}^t)^\top}{2} (\mathbf{A}^t)^\top \mathbf{A}^t \right\rangle_F
\end{aligned}$$

$$\overset{(b)}{\geq} 2\,\lambda_{\min}\big((\mathbf{A}^t)^\top \mathbf{A}^t\big)\left\|\tfrac{\mathbf{H}^t+(\mathbf{H}^t)^\top}{2}\right\|_F^2 + 2\,\lambda_{\min}\big((\mathbf{A}^t)^\top \mathbf{A}^t\big)\left\|\tfrac{\mathbf{H}^t+(\mathbf{H}^t)^\top}{2}\right\|_F^2$$
$$= 4\lambda_{\min}\big((\mathbf{A}^t)^\top \mathbf{A}^t\big)\left\|\tfrac{\mathbf{H}^t+(\mathbf{H}^t)^\top}{2}\right\|_F^2$$
$$= 4\lambda_{\min}\big((\mathbf{A}^t)^\top \mathbf{A}^t\big)\left\|\mathbf{H}^t\right\|_F^2, \tag{4}$$

where equality (a) is obtained due to the fact that the second factor within the inner product is symmetric. Thus, the skew-symmetric part of $\mathbf{H}^t$ is orthogonal. inequality (b) result from $(\mathbf{A}^t)^\intercal\mathbf{A}^t \succeq \lambda_{\min}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\mathbf{I}_k$.

This completes the proof of Lemma 1.

A.4 Proof of Lemma 2

We expand $\left\|\mathbf{S}^t\tilde{\mathbf{A}}^{t+1} - \mathbf{A}^t\right\|_F^2$ as follows:

$$\left\|\mathbf{S}^t\tilde{\mathbf{A}}^{t+1} - \mathbf{A}^t\right\|_F^2 = \left\|\tilde{\mathbf{A}}^{t+1}\right\|_F^2 + \left\|\mathbf{A}^t\right\|_F^2 - 2\,\mathrm{Tr}\left(\mathbf{S}^t\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right). \tag{5}$$

Similarly, we have

$$\left\|\mathbf{S}^{t,\star}\tilde{\mathbf{A}}^{t+1} - \mathbf{A}^t\right\|_F^2 = \left\|\tilde{\mathbf{A}}^{t+1}\right\|_F^2 + \left\|\mathbf{A}^t\right\|_F^2 - 2\,\mathrm{Tr}\left(\mathbf{S}^{t,\star}\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right). \tag{6}$$

Therefore, $\Delta_{\mathrm{proc}}^{t+1}$ satisfies

$$\Delta_{\mathrm{proc}}^{t+1} = 2\left(\mathrm{Tr}\left(\mathbf{S}^{t,\star}\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right) - \mathrm{Tr}\left(\mathbf{S}^t\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right)\right). \tag{7}$$

Let $\mathbf{M}^t = \tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal = \mathbf{U}_{\mathbf{M}}^t\boldsymbol{\Sigma}_{\mathbf{M}}^t\mathbf{V}_{\mathbf{M}}^t$, where matrices $\mathbf{U}_{\mathbf{M}}^t \in \mathbb{R}^{r'\times r'}$ and $\mathbf{V}_{\mathbf{M}}^t \in \mathbb{R}^{r'\times r}$ have orthogonal columns. $\boldsymbol{\Sigma}_{\mathbf{M}}^t \in \mathbb{R}^{r'\times r'}$. Let $\mathbf{P}^t = \mathbf{U}_{\mathbf{M}}^t(\mathbf{S}^t)^\intercal(\mathbf{V}_{\mathbf{M}}^t)^\intercal$ and $\mathbf{P}^{t,\star} = \mathbf{U}_{\mathbf{M}}^t(\mathbf{S}^{t,\star})^\intercal(\mathbf{V}_{\mathbf{M}}^t)^\intercal = \mathbf{I}_{r'}$. Then, we have

$$\mathrm{Tr}\left(\mathbf{S}^t\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right) = \mathrm{Tr}\left(\mathbf{P}^t\boldsymbol{\Sigma}_{\mathbf{M}}^t\right) = \sum_{j=1}^{r'}\sigma_j p_{j,j}. \tag{8}$$

Similarly, we have

$$\mathrm{Tr}\left(\mathbf{S}^{t,\star}\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right) = \sum_{j=1}^{r'}\lambda_j. \tag{9}$$

By combining eqn. (7), (8), and (9), we obtain

$$\Delta_{\mathrm{proc}}^{t+1} = 2\sum_{j=1}^{r'}\lambda_j(1 - p_{j,j}). \tag{10}$$

Since matrices $\mathbf{U}_{\mathbf{M}}^t$ and $\mathbf{V}_{\mathbf{M}}^t$ have orthogonal columns, we have

$$\begin{aligned}\left\|\mathbf{S}^t - \mathbf{S}^{t,\star}\right\|_F^2 &= \left\|\mathbf{P}^t - \mathbf{I}\right\|_F^2 \\ &= \mathrm{Tr}\left((\mathbf{P}^t - \mathbf{I})^\intercal(\mathbf{P}^t - \mathbf{I})\right) \\ &= 2r' - \mathrm{Tr}\big(\mathbf{P}^t\big) - \mathrm{Tr}\left((\mathbf{P}^t)^\intercal\right) \\ &= 2\sum_{j=1}^{r'}(1 - p_{j,j}).\end{aligned} \tag{11}$$

By combining eqns. (10) and (11), we have

$$\begin{aligned}
\Delta_{\text{proc}}^{t+1} &= 2\sum_{j=1}^{r'} \lambda_j(1-p_{j,j}) \\
&\geq 2\lambda_{\min}\left(\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right)(1-p_{j,j}) \\
&= 2\lambda_{\min}\left(\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right)\left\|\mathbf{P}^t-\mathbf{I}\right\|_F^2 \\
&= 2\lambda_{\min}\left(\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right)\left\|\mathbf{S}^t-\mathbf{S}^{t,\star}\right\|_F^2. \qquad (12)
\end{aligned}$$

By rearranging this inequality, we have

$$\left\|\mathbf{S}^t-\mathbf{S}^{t,\star}\right\|_F^2 \leq \frac{\Delta_{\text{proc}}^{t+1}}{\lambda_{\min}\left(\tilde{\mathbf{A}}^{t+1}(\mathbf{A}^t)^\intercal\right)}. \qquad (13)$$

This completes the proof of Lemma 2.

A.5 Proof of Theorem 2

Based on Assumption 1, we expand $\mathbb{E}[f(\mathbf{W}^{t+1})]$ as

$$\mathbb{E}\left[f(\mathbf{W}^{t+1})\right] \leq \mathbb{E}\left[f(\mathbf{W}^t)\right] + \underbrace{\mathbb{E}\left[\left\langle\nabla f(\mathbf{W}^t), \mathbf{W}^{t+1}-\mathbf{W}^t\right\rangle_F\right]}_{T_1} + \frac{L}{2}\underbrace{\mathbb{E}\left[\left\|\mathbf{W}^{t+1}-\mathbf{W}^t\right\|_F^2\right]}_{T_2}. \qquad (14)$$

In particular, $\mathbf{W}^{t+1}-\mathbf{W}^t$ satisfies

$$\begin{aligned}
&\mathbf{W}^{t+1}-\mathbf{W}^t \\
&= \frac{1}{N}\sum_{n\in\mathcal{N}}\left(\mathbf{L}(\mathbf{A}^{t+1})^\intercal\mathbf{A}^{t+1}\mathbf{R}-\mathbf{L}(\mathbf{A}^t)^\intercal\mathbf{A}^t\mathbf{R}\right) \\
&= \mathbf{L}\frac{1}{N}\sum_{n\in\mathcal{N}}\left((\mathbf{A}_n^{t+\frac{1}{2}})^\intercal\mathbf{A}_n^{t+\frac{1}{2}}-(\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\mathbf{R} \\
&= \mathbf{L}\frac{1}{N}\sum_{n\in\mathcal{N}}\left(\left(\mathbf{A}^t-\eta\nabla_{\mathbf{A}}F_n(\mathbf{W}^t;\boldsymbol{\xi}_n)\right)^\intercal\left(\mathbf{A}^t-\eta\nabla_{\mathbf{A}}F_n(\mathbf{W}^t;\boldsymbol{\xi}_n)\right)-(\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\mathbf{R} \\
&= \mathbf{L}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t-\eta\left((\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t+(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\mathbf{A}^t\right)+\eta^2(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t-(\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\mathbf{R} \\
&= \mathbf{L}\left(\eta^2(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t-\eta\left((\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t+(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\mathbf{A}^t\right)\right)\mathbf{R}, \qquad (15)
\end{aligned}$$

where $\bar{\mathbf{G}}_{\mathbf{A}}^t = \frac{1}{N}\sum_{n\in\mathcal{N}}\nabla_{\mathbf{A}}F_n(\mathbf{W}^t;\boldsymbol{\xi}_n) = \mathbf{A}^t\left(\mathbf{H}^t+(\mathbf{H}^t)^\intercal\right)$. We first bound T_1 as follows:

$$\begin{aligned}
&T_1 \\
&= -\eta\mathbb{E}\left[\left\langle\nabla f(\mathbf{W}^t),\mathbf{L}\left((\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t+(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\mathbf{A}^t\right)\mathbf{R}\right\rangle_F\right]+\eta^2\mathbb{E}\left[\left\langle\nabla f(\mathbf{W}^t),\mathbf{L}(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\mathbf{R}\right\rangle_F\right] \\
&= \underbrace{\eta\mathbb{E}\left[\left\langle\nabla f(\mathbf{W}^t),\mathbf{L}\left((\mathbf{A}^{t,\star}-\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t+(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal(\mathbf{A}^{t,\star}-\mathbf{A}^t)\right)\mathbf{R}\right\rangle_F\right]}_{T_3} \\
&\quad -\eta\mathbb{E}\left[\left\langle\nabla f(\mathbf{W}^t),\mathbf{L}\left((\mathbf{A}^{t,\star})^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t+(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\mathbf{A}^t\right)\mathbf{R}\right\rangle_F\right] \\
&\quad +\eta^2\mathbb{E}\left[\left\langle\nabla f(\mathbf{W}^t),\mathbf{L}(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\mathbf{R}\right\rangle_F\right] \\
&\overset{\text{(a)}}{=} T_3-\eta\mathbb{E}\left[\left\langle\mathbf{H}^t,\left((\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t+(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\mathbf{A}^t\right)\right\rangle_F\right]+\eta^2\mathbb{E}\left[\left\langle\mathbf{H}^t,(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\rangle_F\right] \\
&= T_3-\eta\mathbb{E}\left[\left\langle\mathbf{H}^t,(\mathbf{A}^t)^\intercal\mathbf{A}^t\left(\mathbf{H}^t+(\mathbf{H}^t)^\intercal\right)+\left(\mathbf{H}^t+(\mathbf{H}^t)^\intercal\right)(\mathbf{A}^t)^\intercal\mathbf{A}^t\right\rangle_F\right] \\
&\quad +\eta^2\mathbb{E}\left[\left\langle\mathbf{H}^t,(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\rangle_F\right], \qquad (16)
\end{aligned}$$

where equality (a) is obtained due to the fact that for an arbitrary matrix $\mathbf{X}$, we have $\langle \mathbf{G}^t, \mathbf{L}\mathbf{X}\mathbf{R}\rangle = \langle \mathbf{H}^t, \mathbf{X}\rangle$. T_3 can be bounded as follows:

$$\begin{aligned}
T_3 &\leq \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{\eta}{2}\mathbb{E}\left[\left\|\mathbf{L}\left((\mathbf{A}^{t,\star}-\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t + (\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal(\mathbf{A}^{t,\star}-\mathbf{A}^t)\right)\mathbf{R}\right\|_F^2\right] \\
&\overset{\text{(a)}}{\leq} \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + 2\eta\mathbb{E}\left[\left\|(\mathbf{A}^{t,\star}-\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^2\right] \\
&\overset{\text{(b)}}{\leq} \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + 2\eta\left\|\mathbf{A}^{t,\star}-\mathbf{A}^t\right\|_F^2\mathbb{E}\left[\left\|\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^2\right] \\
&\overset{\text{(c)}}{\leq} \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{2\eta}{N^2}\left\|\mathbf{A}^{t,\star}-\mathbf{A}^t\right\|_F^2\sum_{n\in\mathcal{N}}\left\|\nabla_{\mathbf{A}}F_n(\mathbf{W}^t;\boldsymbol{\xi}_n)\right\|_F^2 \\
&\overset{\text{(d)}}{\leq} \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{2\eta\psi}{N}\left\|\mathbf{A}^{t,\star}-\mathbf{A}^t\right\|_F^2 \\
&= \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{2\eta\psi}{N}\left\|\left(\mathbf{S}^{t,\star}-\mathbf{S}^t\right)\tilde{\mathbf{A}}^t\right\|_F^2 \\
&\overset{\text{(e)}}{\leq} \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{2\eta\psi}{N}\left\|\mathbf{S}^{t,\star}-\mathbf{S}^t\right\|_F^2\left\|\tilde{\mathbf{A}}^t\right\|_F^2 \\
&\overset{\text{(f)}}{\leq} \frac{\eta}{2}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{2\eta\psi\tilde{C}_{\mathbf{A}}^2\Delta_{\text{proc}}^t}{N\lambda_{\min}\left(\tilde{\mathbf{A}}^t(\mathbf{A}^{t-1})^\intercal\right)},
\end{aligned} \tag{17}$$

where inequalities (a) and (c) result from Jensen's inequality. Inequalities (b) and (e) is obtained by using the fact $\|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\|_F\|\mathbf{B}\|_F$. Inequality (d) results from Assumption 2. Inequality (f) results from Assumption 3 and Lemma 2. Then, we further bound the second term of T_1 as

$$\begin{aligned}
&-\eta\mathbb{E}\left[\left\langle \mathbf{H}^t, (\mathbf{A}^t)^\intercal\mathbf{A}^t\left(\mathbf{H}^t+(\mathbf{H}^t)^\intercal\right) + \left(\mathbf{H}^t+(\mathbf{H}^t)^\intercal\right)(\mathbf{A}^t)^\intercal\mathbf{A}^t\right\rangle_F\right] \\
&\overset{\text{(a)}}{\leq} -4\eta\lambda_{\min}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\left\|(\mathbf{L})^\intercal\nabla f(\mathbf{W}^t)(\mathbf{R})^\intercal\right\|_F^2,
\end{aligned} \tag{18}$$

where inequality (a) results from Lemma 2. Then, we bound $\eta^2\mathbb{E}\left[\left\langle \mathbf{H}^t, (\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\rangle_F\right]$ as follows:

$$\begin{aligned}
\eta^2\mathbb{E}\left[\left\langle \mathbf{H}^t, (\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\rangle_F\right] &\overset{\text{(a)}}{\leq} \eta^2\left(\frac{1}{2}\left\|\mathbf{H}^t\right\|_F^2 + \frac{1}{2}\left\|\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^4\right) \\
&= \eta^2\left(\frac{1}{2}\left\|(\mathbf{L})^\intercal\nabla f(\mathbf{W}^t)(\mathbf{R})^\intercal\right\|_F^2 + \frac{1}{2}\left\|\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^4\right) \\
&\overset{\text{(b)}}{\leq} \frac{\eta^2}{2}\left\|(\mathbf{L})^\intercal\nabla f(\mathbf{W}^t)(\mathbf{R})^\intercal\right\|_F^2 + \frac{\eta^2\psi^2}{2},
\end{aligned} \tag{19}$$

where inequality (a) results from Jensen's inequality. Inequality (b) is obtained by using Lemma 2. By combining inequalities (16), (17), (18), and (19), we have

$$T_1 \leq \frac{2\eta\psi\tilde{C}_{\mathbf{A}}^2\Delta_{\text{proc}}^t}{N\lambda_{\min}\left(\tilde{\mathbf{A}}^t(\mathbf{A}^{t-1})^\intercal\right)} + \left(-4\eta\lambda_{\min}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t\right) + \frac{\eta^2}{2} + \frac{\eta}{2}\right)\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{\eta^2}{2}\psi^2. \tag{20}$$

Now, we bound T_2. In particular, it satisfies

$$\begin{aligned}
T_2 &= \mathbb{E}\left[\left\|\mathbf{L}\Big(\eta^2(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t - \eta\left((\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t + (\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\mathbf{A}^t\right)\Big)\mathbf{R}\right\|_F^2\right] \\
&\overset{\text{(a)}}{\leq} 3\eta^4\mathbb{E}\left[\left\|\mathbf{L}\eta^2(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\mathbf{R}\right\|_F^2\right] + 6\eta^2\mathbb{E}\left[\left\|\mathbf{L}\eta(\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\mathbf{R}\right\|_F^2\right] \\
&\overset{\text{(b)}}{\leq} 3\eta^4\mathbb{E}\left[\left\|(\bar{\mathbf{G}}_{\mathbf{A}}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^2\right] + 6\eta^2\mathbb{E}\left[\left\|(\mathbf{A}^t)^\intercal\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^2\right] \\
&\overset{\text{(c)}}{\leq} 3\eta^4\mathbb{E}\left[\left\|\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^4\right] + 6\eta^2\mathbb{E}\left[\left\|\mathbf{A}^t\right\|_F^2\right]\mathbb{E}\left[\left\|\bar{\mathbf{G}}_{\mathbf{A}}^t\right\|_F^2\right] \\
&\overset{\text{(d)}}{\leq} 3\eta^2\psi\left(\eta^2\psi + 2C_{\mathbf{A}}^2\right).
\end{aligned} \tag{21}$$

where inequality (a) is obtained by using Jensen's inequality. Inequality (b) is obtained since $\mathbf{L}$ and $\mathbf{R}$ have orthogonal columns. Inequality (c) results from the fact $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F\|\mathbf{B}\|_F$. Inequality (d) is obtained by using Lemma 4 and Assumption 3.

Then, we combine inequalities (14), (20), and (21), we have

$$\mathbb{E}\left[f(\bar{\mathbf{W}}^{t+1})\right] \leq \mathbb{E}\left[f(\bar{\mathbf{W}}^{t})\right] + \left(\frac{\eta^2}{2} + \frac{\eta}{2} - 4\eta\lambda_{\min}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\right)\left\|(\nabla f(\mathbf{W}^t)\right\|_F^2 + \frac{\eta^2\psi^2}{2} + \frac{3L\eta^2\psi\left(\eta^2\psi + 2C_{\mathbf{A}}^2\right)}{2} + \frac{2\eta\psi\tilde{C}_{\mathbf{A}}^2\Delta_{\text{proc}}^t}{N\lambda_{\min}\left(\tilde{\mathbf{A}}^t(\mathbf{A}^{t-1})^\intercal\right)}. \tag{22}$$

We denote $\Omega = 4\eta\min_{t\in\mathcal{T}}\{\lambda_{\min}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\} - \frac{\eta^2}{2} - \frac{\eta}{2}$. By rearranging inequality (22) and summing over all T fine-tuning rounds, we have

$$\begin{aligned}
&\Omega\sum_{t\in\mathcal{T}}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 \\
&\leq \sum_{t\in\mathcal{T}}\left(f(\bar{\mathbf{W}}^t) - f(\bar{\mathbf{W}}^{t+1})\right) + T\left(\frac{\eta^2\psi^2}{2} + \frac{3L\eta^2\psi\left(\eta^2\psi + 2C_{\mathbf{A}}^2\right)}{2}\right) + \sum_{t\in\mathcal{T}}\frac{2\eta\psi\tilde{C}_{\mathbf{A}}^2\Delta_{\text{proc}}^t}{N\lambda_{\min}\left(\tilde{\mathbf{A}}^t(\mathbf{A}^{t-1})^\intercal\right)} \\
&\overset{(a)}{\leq} f(\bar{\mathbf{W}}^1) - f(\bar{\mathbf{W}}^\star) + T\left(\frac{\eta^2\psi^2}{2} + \frac{3L\eta^2\psi\left(\eta^2\psi + 2C_{\mathbf{A}}^2\right)}{2}\right) + \sum_{t\in\mathcal{T}}\frac{2\eta\psi\tilde{C}_{\mathbf{A}}^2\Delta_{\text{proc}}^t}{N\lambda_{\min}\left(\tilde{\mathbf{A}}^t(\mathbf{A}^{t-1})^\intercal\right)},
\end{aligned} \tag{23}$$

where inequality (a) is obtained due to the fact that $f(\mathbf{W}^\star) \leq f(\mathbf{W}^{T+1})$. If $\eta < 8\min_{t\in\mathcal{T}}\{\lambda_{\min}\left((\mathbf{A}^t)^\intercal\mathbf{A}^t\right)\} - 1$, we have $\Omega > 0$. Then, we multiply $\frac{1}{T\Omega}$ on both sides of inequality (23) and obtain

$$\frac{1}{T}\sum_{t\in\mathcal{T}}\left\|\nabla f(\mathbf{W}^t)\right\|_F^2 \leq \frac{f(\mathbf{W}^1) - f(\mathbf{W}^\star)}{T\Omega} + \frac{\eta^2\psi^2}{2\Omega} + \frac{3L\eta^2\psi\left(\eta^2\psi + 2C_{\mathbf{A}}^2\right)}{2\Omega} + \frac{1}{T\Omega}\sum_{t\in\mathcal{T}}\frac{2\eta\psi\tilde{C}_{\mathbf{A}}^2\Delta_{\text{proc}}^t}{N\lambda_{\min}\left(\tilde{\mathbf{A}}^t(\mathbf{A}^{t-1})^\intercal\right)}. \tag{24}$$

This completes the proof of Theorem 2.

A.6 Additional Experimental Results

Table 1: Comparison of the testing accuracy.

Base Model	Dataset	FLoRG	FedIT	FeDeRA	FFA-LoRA	FedSA-LoRA
OPT-125M	MRPC	**88.30**	82.09	85.70	84.78	86.51
	QQP	**89.72**	84.18	87.09	86.30	87.92
RoBERTa-large	MRPC	**90.80**	85.16	88.30	87.50	89.21
	QQP	**91.52**	87.89	89.30	88.51	90.01

A.7 The Use of Large Language Models (LLMs)

We used large language models (e.g., ChatGPT) for the general purpose of writing assistants to revise and polish the manuscript's prose: improving grammar, wording, clarity, and fixing minor LaTeX/formatting issues. The models were not used for research ideation, designing algorithms/experiments, deriving proofs, selecting citations, or writing substantive technical content.

Table 2: Comparison of the testing accuracy under different ranks.

Rank	Dataset	FLoRG	FedIT	FeDeRA	FFA-LoRA	FedSA-LoRA
$r = 2$	MRPC	**85.91**	79.66	83.17	82.32	83.00
	QQP	**87.62**	81.88	84.90	84.09	85.65
	MNLI	**86.14**	80.70	83.44	82.64	84.22
	QNLI	**89.70**	86.67	87.22	86.40	88.10
$r = 4$	MRPC	**90.80**	85.16	88.30	87.50	89.21
	QQP	**91.52**	87.89	89.30	88.51	90.01
	MNLI	**91.39**	84.76	88.20	89.12	90.90
	QNLI	**92.44**	87.63	89.91	90.82	91.69
$r = 8$	MRPC	**91.70**	85.55	87.12	88.20	89.91
	QQP	**91.80**	86.39	89.26	88.45	90.10
	MNLI	**91.70**	86.40	89.29	90.44	90.41
	QNLI	**92.50**	87.77	90.20	90.41	91.13

Table 3: Comparison of the testing accuracy under different degrees of data heterogeneity.

Non-IIDness	Dataset	FLoRG	FedIT	FeDeRA	FFA-LoRA	FedSA-LoRA
$\rho = 0.1$	MRPC	**85.06**	78.19	81.91	81.00	83.01
	QQP	**86.95**	78.80	84.15	83.21	85.00
	MNLI	**81.35**	75.44	78.42	77.57	79.36
	QNLI	**88.88**	83.31	86.20	85.33	87.19
$\rho = 0.5$	MRPC	**90.80**	85.16	88.30	87.50	89.21
	QQP	**91.52**	87.89	89.30	88.51	90.01
	MNLI	**91.39**	84.76	88.20	89.12	90.90
	QNLI	**92.44**	87.63	89.91	90.82	91.69
$\rho = 1$	MRPC	**90.76**	85.77	89.26	88.91	88.20
	QQP	**91.88**	88.61	89.93	89.37	90.39
	MNLI	**91.75**	86.47	90.05	91.70	91.33
	QNLI	92.17	88.03	90.52	90.79	**92.72**