VII. Theoretical Analysis

This supplementary material provides additional details and theoretical proofs related to the FedSRC algorithm, as presented in the main paper. Federated Learning (FL) often faces challenges in maintaining the performance of the global model due to the varying quality and reliability of data generated by clients locally. Traditional approaches have attempted to mitigate this by discarding or limiting contributions from certain clients, but this can lead to wasted computational and communication resources. Moreover, while active client selection could help avoid such waste, it typically requires client-level profiling, which can compromise privacy.

FedSRC (**Fed**erated Learning with **S**elf-**R**egulating **C**lients) is a novel approach that addresses these issues by enabling clients to self-regulate their participation in the FL process. In FedSRC, each client independently assesses whether its local training is beneficial for the global model by using a lightweight checkpoint mechanism. This mechanism involves calculating the local test loss on the global model and evaluating the Refined Heterogeneity Index (RHI). Clients decide whether to participate in a given FL round based on these metrics, which not only preserves their anonymity but also optimizes resource usage by reducing unnecessary computation and communication.Through extensive evaluations on four datasets, FedSRC has demonstrated the ability to reduce communication costs by up to 30% and computation costs by up to 55%, without compromising the overall model performance.

This supplementary material is organized as follows: Section A presents the full theoretical analysis and proof of convergence for FedSRC. Section B provides detailed descriptions of the datasets and implementation details to assist in the reproducibility of our results. Section C includes extended experimental results that further validate our approach.

A. Main Theorem Proof

Our proof follows similar lines to that of [16] but with modifications based on our problem formulation of having good and bad clients as well as our different skewness metrics and local-global objective gap ρ_g, ρ_b, and Γ_g, respectively. We begin by recapping the key assumptions necessary for our theoretical analysis

Assumption VII.1. $F_1, \ldots, F_k$ are all L-smooth, i.e., for all v and w,

$$F_k(v) \le F_k(w) + (v-w)^T \nabla F_k(w) + \frac{L}{2}\|v-w\|_2^2.$$

Assumption VII.2. $F_1, \ldots, F_k$ are all μ-strongly convex, i.e., for all v and w,

$$F_k(v) \ge F_k(w) + (v-w)^T \nabla F_k(w) + \frac{\mu}{2}\|v-w\|_2^2.$$

Assumption VII.3. For the mini-batch ξ_k uniformly sampled at random from $\mathcal{D}_k$ of user k, the resulting stochastic gradient is unbiased; that is, $\mathbb{E}[g_k(w_k, \xi_k)] = \nabla F_k(w_k)$. Also, the variance of stochastic gradients is bounded: $\mathbb{E}[\|g_k(w_k, \xi_k) - \nabla F_k(w_k)\|^2] \le \sigma^2$ for all $k = 1, \ldots, K$.

Assumption VII.4. The stochastic gradients' expected squared norms are uniformly bounded, i.e., $\mathbb{E}[\|g_k(w_k, \xi_k)\|^2] \le G^2$ for $k = 1, \ldots, K$.

we present some preliminary lemmas that are useful for the proof of Theorem III.8.

B. Preliminary Lemmas

Lemma VII.5. *Assume F_k is L-smooth with global optimum at w_k^*. Then for any w_k in the domain of F_k,*

$$\|\nabla F_k(w_k)\|^2 \le 2L(F_k(w_k) - F_k(w_k^*)).$$

Proof. Since F_k is L-smooth,

$$F_k(w_k) - F_k(w_k^*) - \langle \nabla F_k(w_k^*), w_k - w_k^* \rangle \ge \frac{1}{2L}\|\nabla F_k(w_k) - \nabla F_k(w_k^*)\|^2$$

and $\nabla F_k(w_k^*) = 0$ since w_k^* is a minimizer, so this implies

$$F_k(w_k) - F_k(w_k^*) \ge \frac{1}{2L}\|\nabla F_k(w_k)\|^2$$

which yields the claim. □

Lemma VII.6. *For $w_k^{(t)}$ and $\bar{w}^{(t)} = \frac{1}{m}\sum\limits_{k \in S^{(t)}} w_k^{(t)}$,*

$$\frac{1}{m}\mathbb{E}\left[\sum_{k \in S^{(t)}} \|\bar{w}^{(t)} - w_k^{(t)}\|^2\right] \le 16\eta_t^2 \tau^2 G^2.$$

Proof. We have

$$\frac{1}{m}\sum_{k\in S^{(t)}}\|\bar{w}^{(t)}-w_k^{(t)}\|^2 \leq \sum_{k\in S^{(t)}}\|\frac{1}{m}\sum_{\tilde{k}\in S^{(t)}} w_{\tilde{k}}^{(t)}-w_k^{(t)}\|^2 = \frac{1}{m^2}\sum_{k\in S^{(t)}}\sum_{\tilde{k}\in S^{(t)}}\|w_{\tilde{k}}^{(t)}-w_k^{(t)}\|^2$$
$$= \frac{1}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\|w_{\tilde{k}}^{(t)}-w_k^{(t)}\|^2,$$

where the inequality follows from $\|\sum_{i=1}^n \mathbf{x}_i\|^2 \leq n\sum_{i=1}^n\|\mathbf{x}_i\|^2$. For $k=\tilde{k}$, the right hand side of the above inequality is zero. Since the selected clients get updated at every τ for any t there exist t_0 such that $w_{\tilde{k}}^{(t_0)} = w_k^{(t)}$, where $0\leq t-t_0\leq\tau$. Hence for any t, $\|w_{\tilde{k}}^{(t)}-w_k^{(t)}\|^2$ is bounded above by τ updates. With non-increasing η_t over t and $\eta_{t_0}\leq 2\eta_t$, we can write the right hand side of the above inequality as

$$\frac{1}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\|w_{\tilde{k}}^{(t)}-w_k^{(t)}\|^2 \leq \frac{1}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\|\sum_{i=t_0}^{t_0+\tau-1}\eta_i\left(g_{\tilde{k}}(w_{\tilde{k}}^{(i)},\xi_{\tilde{k}}^{(i)})-g_k^{(i)}(w_k^{(i)},\xi_k^{(i)})\right)\|^2$$
$$\leq \frac{\eta_{t_0}^2\tau}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\sum_{i=t_0}^{t_0+\tau-1}\left[2\|\big(g_{\tilde{k}}(w_{\tilde{k}}^{(i)},\xi_{\tilde{k}}^{(i)})\|^2+2\|g_k^{(i)}(w_k^{(i)},\xi_k^{(i)})\big)\|^2\right].$$

Taking expectation and applying Assumption 4 gives

$$\mathbb{E}\Big[\frac{1}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\|w_{\tilde{k}}^{(t)}-w_k^{(t)}\|^2\Big] \leq \frac{2\eta_{t_0}^2\tau}{m^2}\mathbb{E}\Bigg[\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\sum_{i=t_0}^{t_0+\tau-1}\Big[\|\left(g_{\tilde{k}}(w_{\tilde{k}}^{(i)},\xi_{\tilde{k}}^{(i)})\|^2+\|g_k^{(i)}(w_k^{(i)},\xi_k^{(i)})\right)\|^2\Big]\Bigg]$$
$$\leq \frac{8\eta_t^2\tau}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}\sum_{i=t_0}^{t_0+\tau-1}(G^2+G^2) = \frac{8\eta_t^2\tau}{m^2}\sum_{\substack{k,\tilde{k}\in S^{(t)}\\ k\neq\tilde{k}}}2\tau G^2$$
$$= \frac{8\eta_t^2\tau}{m^2}m(m-1)2\tau G^2 \leq 16\eta_t^2\tau^2G^2.$$

$\square$

Lemma VII.7. *For any random selection strategy,* $\mathbb{E}\|\bar{w}^{(t)}-w^*\|^2$ *has the following upper bound:*

$$\mathbb{E}[\|\bar{w}^{(t)}-w^*\|^2] \leq \frac{1}{m}\mathbb{E}[\sum_{k\in S^{(t)}}\|w_k^{(t)}-w^*\|^2].$$

Proof.

$$\mathbb{E}[\|\bar{w}^{(t)}-w^*\|^2] = \mathbb{E}[\|\frac{1}{m}\sum_{k\in S^{(t)}}w_k^{(t)}-w^*\|^2] = \mathbb{E}[\|\frac{1}{m}\sum_{k\in S^{(t)}}(w_k^{(t)}-w^*)\|^2]$$
$$\leq \mathbb{E}[\frac{1}{m}\sum_{k\in S^{(t)}}\|w_k^{(t)}-w^*\|^2].$$

$\square$

C. Proof of Theorem III.8

Letting $\bar{g}(t) = \frac{1}{m}\sum_{k\in S(t)} g_k(w_k^{(t)}, \xi_k^{(t)})$, and using the condensed notation $\bar{g}_k = \bar{g}_k(\bar{w}_k^{(t)}, \xi_k^{(t)})$ for simplicity, we have

$$
\begin{aligned}
&\|\bar{w}^{(t+1)} - w^*\|^2 = \|\bar{w}^{(t)} - \eta_t \bar{g}^{(t)} - w^*\|^2 \\
&= \|\bar{w}^{(t)} - \eta_t \bar{g}^{(t)} - w^* - \frac{\eta_t}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)}) + \frac{\eta_t}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)})\|^2 \\
&= \|\bar{w}^{(t)} - w^* - \frac{\eta_t}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)})\|^2 + \eta_t^2\|\frac{1}{m}\sum_{k\in S^{(t)}} (\nabla F_k(w_k^{(t)}) - \bar{g}_k^{(t)})\|^2 \\
&+ 2\eta_t\langle \bar{w}^{(t)} - w^* - \frac{\eta_t}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)}), \frac{1}{m}\sum_{k\in S^{(t)}} (\nabla F_k(w_k^{(t)}) - \bar{g}_k^{(t)})\rangle \\
&= \|\bar{w}^{(t)} - w^*\|^2 \underbrace{- 2\eta_t\langle \bar{w}^{(t)} - w^*, \frac{1}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)})\rangle}_{A_1} + \\
&\underbrace{2\eta_t\langle \bar{w}^{(t)} - w^* - \frac{\eta_t}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)}), \frac{1}{m}\sum_{k\in S^{(t)}} (\nabla F_k(w_k^{(t)}) - g_k^{(t)})\rangle}_{A_2} + \underbrace{\eta_t^2\|\frac{1}{m}\sum_{k\in S^{(t)}} \nabla F_k(w_k^{(t)}\|^2}_{A_3} \\
&+ \underbrace{\eta_t^2\|\frac{1}{m}\sum_{k\in S^{(t)}} (\nabla F_k(w_k^{(t)}) - \bar{g}_k^{(t)})\|^2}_{A_4} \\
&= \|\bar{w}^{(t)} - w^*\|^2 + A_1 + A_2 + A_3 + A_4. \qquad (4)
\end{aligned}
$$

We first bound the quantity A_1 of inequality (4) as follows:

$$
\begin{aligned}
A_1 &= -\frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \langle \bar{w}^{(t)} - w^*, \nabla F_k(w_k^{(t)})\rangle \\
&= -\frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \langle \bar{w}^{(t)} - w_k^{(t)}, \nabla F_k(w_k^{(t)})\rangle - \frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \langle w_k^{(t)} - w^*, \nabla F_k(w_k^{(t)})\rangle \\
&\leq \frac{\eta_t}{m}\sum_{k\in S^{(t)}} \left(\frac{1}{\eta_t}\|\bar{w}^{(t)} - w_k^{(t)}\|^2 + \eta_t\|\nabla F_k(w_k^{(t)})\|^2\right) - \frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \langle w_k^{(t)} - w^*, \nabla F_k(w_k^{(t)})\rangle \\
&\quad \text{(using the AM-GM and Cauchy–Schwarz inequalities)} \\
&= \frac{1}{m}\sum_{k\in S^{(t)}} \|\bar{w}^{(t)} - w_k^{(t)}\|^2 + \frac{\eta_t^2}{m}\sum_{k\in S^{(t)}} \|\nabla F_k(w_k^{(t)})\|^2 - \frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \langle w_k^{(t)} - w^*, \nabla F_k(w_k^{(t)})\rangle \\
&\leq \frac{1}{m}\sum_{k\in S^{(t)}} \|\bar{w}^{(t)} - w_k^{(t)}\|^2 + \frac{2L\eta_t^2}{m}\sum_{k\in S^{(t)}} \Big(F_k(w_k^{(t)}) - F_k^*\Big) \\
&\quad - \frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \langle w_k^{(t)} - w^*, \nabla F_k(w_k^{(t)})\rangle \text{ (using Lemma VII.5)} \\
&\leq \frac{1}{m}\sum_{k\in S^{(t)}} \|\bar{w}^{(t)} - w_k^{(t)}\|^2 + \frac{2L\eta_t^2}{m}\sum_{k\in S^{(t)}} \Big(F_k(w_k^{(t)}) - F_k^*\Big) \\
&\quad - \frac{2\eta_t}{m}\sum_{k\in S^{(t)}} \Big[F_k(w_k^{(t)}) - F_k(w^*) \quad + \frac{\mu}{2}\|w_k^{(t)} - w^*\|^2\Big],
\end{aligned}
$$

where the last inequality follows from μ strong convexity of F_k (Assumption 2). Hence, by Lemma VII.6, the expected value of A_1 satisfies

$$
\begin{aligned}
\mathbb{E}[A_1] \leq& 16\eta_t^2\tau^2G^2 - \frac{\eta_t\mu}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}} \|w_k^{(t)} - w^*\|^2\big] + \frac{2L\eta_t^2}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}} \Big(F_k(w_k^{(t)}) - F_k^*\Big)\big] \\
&- \frac{2\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}} (F_k(w_k^{(t)}) - F_k(w^*))\big]. \qquad (5)
\end{aligned}
$$

Leaving this bound aside for the moment, next notice that $\mathbb{E}[A_2] = 0$ because of the unbiased gradient assumption (Assumption 3). We may then bound A_3 by Lemma VII.6 as follows:

$$\begin{aligned}\mathbb{E}[A_3] = \mathbb{E}\left[\frac{\eta_t^2}{m^2}\|\sum_{k\in S^{(t)}}\nabla F_k(w_k^{(t)})\|^2\right] &\leq \frac{\eta_t^2}{m}\sum_{k\in S^{(t)}}\mathbb{E}\left[\|\nabla F_k(w_k^{(t)}\|^2\right] \\ &\leq \frac{2L\eta_t^2}{m}\mathbb{E}\left[\sum_{k\in S^{(t)}}(F_k(w_k^{(t)}) - F_k^*)\right].\end{aligned} \tag{6}$$

Finally, the bound for A_4 is as follows:

$$\begin{aligned}\mathbb{E}[A_4] =&\mathbb{E}\left[\frac{\eta_t^2}{m^2}\|\sum_{k\in S^{(t)}}\left(\nabla F_k(w_k^{(t)}) - g_k^{(t)}\right)\|^2\right] = \frac{\eta_t^2}{m^2}\mathbb{E}_{S^{(t)}}\left[\sum_{k\in S^{(t)}}\mathbb{E}\|(\nabla F_k(w_k^{(t)}) - g_k^{(t)})\|^2\right] \\ \leq& \frac{\eta_t^2 m\sigma^2}{m^2} = \frac{\eta_t^2\sigma^2}{m},\end{aligned} \tag{7}$$

where the second equality and inequality use Assumption 3.

Using the bounds (5), (6), and (7) in (4), we have

$$\begin{aligned}&\mathbb{E}[\|\bar{w}^{(t+1)} - w^*\|^2] \leq \mathbb{E}[\|\bar{w}^{(t)} - w^*\|^2] + \sum_{i=1}^{4}\mathbb{E}[A_i] \leq \mathbb{E}\big[\|\bar{w}^{(t)} - w^*\|^2\big] - \frac{\eta_t\mu}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}\|w_k^{(t)} - w^*\|^2\big] \\ &+ 16\eta_t^2\tau^2G^2 + \frac{\eta_t^2\sigma^2}{m} + \frac{4L\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(w_k^{(t)}) - F_k^*)\big] - \frac{2\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(w_k^{(t)}) - F_k(w^*))\big] \\ &\leq (1-\eta_t\mu)\mathbb{E}[\|\bar{w}^{(t)} - w^*\|^2] + 16\eta_t^2\tau^2G^2 + \frac{\eta_t^2\sigma^2}{m} + \underbrace{\frac{4L\eta_t^2}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(w_k^{(t)}) - F_k^*)\big]}_{A_5} \\ &\underbrace{-\frac{2\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(w_k^{(t)}) - F_k(w^*))\big]}_{A_5}.\end{aligned} \tag{8}$$

The final inequality above utilizes Lemma VII.7.

Now we bound A_5 as follows:

$$\begin{aligned}A_5 =&\mathbb{E}\big[\frac{4L\eta_t^2}{m}\sum_{k\in S^{(t)}}F_k(w_k^{(t)}) - \frac{2\eta_t}{m}\sum_{k\in S(t)}F_k(w_k^{(t)}) - \frac{2\eta_t}{m}\sum_{k\in S^{(t)}}(F_k^* - F_k(w^*)) \\ &+ \frac{2\eta_t}{m}\sum_{k\in S^{(t)}}F_k^* - \frac{4L\eta_t^2}{m}\sum_{k\in S^{(t)}}F_k^*\big] \\ &= \mathbb{E}\big[\underbrace{\frac{2\eta_t(2L\eta_t - 1)}{m}\sum_{k\in S^{(t)}}(F_k(w_k^{(t)}) - F_k^*)\big]}_{A_6} + 2\eta_t E\big[\frac{1}{m}\sum_{k\in S(t)}(F_k(w^*) - F_k^*)\big].\end{aligned}$$

Take $\eta_t < 1/(4L)$ and define $\upsilon_t = 2\eta_t(1-2L\eta_t) \geq 0$; then we can bound A_6 as

$$
\begin{aligned}
&-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}(F_k(w_k^{(t)})-F_k^*)\\
&=-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}(F_k(w_k^{(t)})-F_k(\bar{w}^{(t)})+F_k(\bar{w}^{(t)})-F_k^*)\\
&=-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}\big[F_k(w_k^{(t)})-F_k(\bar{w}^{(t)})\big]-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}\big[F_k(\bar{w}^{(t)})-F_k^*\big]\\
&\leq-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}\big[\langle\nabla F_k(w^{(t)}),w_k^{(t)}-\bar{w}^{(t)}\rangle+\frac{\mu}{2}\|w_k^{(t)}-\bar{w}^{(t)}\|^2\big]-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}\big[F_k(\bar{w}^{(t)})-F_k^*\big]\\
&\leq\frac{\nu_t}{m}\sum_{k\in S^{(t)}}\big[\eta_t L(F_k(\bar{w}^{(t)})-F_k^*)+(\frac{1}{2\eta_t}-\frac{\mu}{2})\|w_k^{(t)}-\bar{w}^{(t)}\|^2\big]-\frac{\upsilon_t}{m}\sum_{k\in S^{(t)}}\big[F_k(\bar{w}^{(t)})-F_k^*\big]\\
&\text{(using the Cauchy-Schwarz inequality, the AM-GM inequality, and Lemma VII.5)}\\
&=-\frac{\nu_t}{m}(1-\eta_t L)\sum_{k\in S^{(t)}}(F_k(\bar{w}^{(t)})-F_k^*)+\Big(\frac{\nu_t}{2\eta_t m}-\frac{\nu_t\mu}{2m}\Big)\sum_{k\in S^{(t)}}\|w_k^{(t)}-\bar{w}^{(t)}\|^2\\
&\leq-\frac{\nu_t}{m}(1-\eta_t L)\sum_{k\in S^{(t)}}(F_k(\bar{w}^{(t)})-F_k^*)+\frac{1}{m}\sum_{k\in S^{(t)}}\|w_k^{(t)}-\bar{w}^{(t)}\|^2. \qquad (9)
\end{aligned}
$$

The first inequality above uses μ strong convexity of F_k, the subsequent inequality uses L–smoothness of F_k, and the final inequality follows because $\frac{\nu_t(1-\eta_t\mu)}{2\eta_t} \leq 1$. Hence, we can bound A_5 as follows:

$$
\begin{aligned}
\mathbb{E}[A_5] &\leq -\frac{\nu_t}{m}(1-\eta_t L)\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(\bar{w}^{(t)})-F_k^*)\big]+\frac{1}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}\|w_k^{(t)}-\bar{w}^{(t)}\|^2\big]\\
&\quad+\frac{2\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(w^*)-F_k^*)\big]\\
&\leq -\frac{\nu_t}{m}(1-\eta_t L)\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(\bar{w}^{(t)})-F_k^*)\big]+16\eta_t^2\tau^2G^2+\frac{2\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}}(F_k(w^*)-F_k^*)\big]\\
&= -\frac{\nu_t}{m}(1-\eta_t L)\mathbb{E}\big[\sum_{k\in S^{(t)}\cap\mathcal{G}}(F_k(\bar{w}^{(t)})-F_k^*)+\sum_{k\in S^{(t)}\cap\mathcal{B}}(F_k(\bar{w}^{(t)})-F_k^*)\big]+16\eta_t^2\tau^2G^2\\
&\quad+\frac{2\eta_t}{m}\mathbb{E}\big[\sum_{k\in S^{(t)}\cap\mathcal{G}}(F_k(w^*)-F_k^*)+\sum_{k\in S^{(t)}\cap\mathcal{B}}(F_k(w^*)-F_k^*)\big] \qquad (10)\\
&= 16\eta_t^2\tau^2G^2-\frac{\nu_t(1-\eta_t L)}{m}\mathbb{E}\big[\big(p\rho_g(S(\pi,\bar{w}^{(\tau\lfloor t/\tau\rfloor)}),\bar{w}^{(t)})\\
&\quad+q\rho_b(S(\pi,\bar{w}^{(\tau\lfloor t/\tau\rfloor)}),\bar{w}(t))\big)(F_g(\bar{w}^{(t)})-\sum_{k\in\mathcal{G}}p_kF_k^*\big]+\frac{2\eta_t}{m}\mathbb{E}\big[\big(p\rho_g(S(\pi,\bar{w}^{(\tau\lfloor t/\tau\rfloor)}),w^*)\\
&\quad+q\rho_b(S(\pi,\bar{w}^{(\tau\lfloor t/\tau\rfloor)}),w^*)(F_g(w^*)-\sum_{k\in\mathcal{G}}p_kF_k^*)\big]\\
&\leq 16\eta_t^2\tau^2G^2\underbrace{-\frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g+q\bar{\rho}_b\big](\mathbb{E}[F_g(\bar{w}^{(t)}-\sum_{k\in\mathcal{G}}p_kF_k^*)}_{A_7}+\frac{2\eta_t}{m}(p\tilde{\rho}_g+q\tilde{\rho}_b)\Gamma_g
\end{aligned}
$$

We used the definition of $\rho(S(\pi,w),w')$ and Γ_g to arrive at (10). We can get a bound for A_7 in (10) as follows:

$$
\begin{aligned}
A_7 &= -\frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\sum_{k\in\mathcal{G}} p_k\big(\mathbb{E}[F_k(\bar{w}^{(t)})] - F^* + F^* - F_k^*\big)\\
&= -\frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\sum_{k\in\mathcal{G}} p_k\left(\mathbb{E}[F_k(\bar{w}^{(t)})] - F^* + F^* - F_k^*\right)\\
&= -\frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\sum_{k\in\mathcal{G}} p_k\big(\mathbb{E}[F_k(\bar{w}^{(t)})] - F^*\big)\\
&\quad -\frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\sum_{k\in\mathcal{G}} p_k\big(F^* - F_k^*\big)\\
&= -\frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\big(\mathbb{E}[F_g(\bar{w}^{(t)})] - F^*\big) - \frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\Gamma_g\\
&\text{(using the definition of } \Gamma_g)\\
&\leq -\frac{\nu_t(1-\eta_t L)\mu\big[p\bar{\rho}_g + q\bar{\rho}_b\big]}{2m}\mathbb{E}\big[\|\bar{w}^{(t)} - w^*\|^2\big] - \frac{\nu_t(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\Gamma_g\\
&\text{(using } \mu \text{ strongly convexity)}\\
&= -\frac{2\eta_t(1-2L\eta_t)(1-\eta_t L)\mu\big[p\bar{\rho}_g + q\bar{\rho}_b\big]}{2m}\mathbb{E}\big[\|\bar{w}^{(t)} - w^*\|^2\big]\\
&\quad -\frac{2\eta_t(1-2L\eta_t)(1-\eta_t L)}{m}\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\Gamma_g\\
&\leq -\frac{3\eta_t\mu\big[p\bar{\rho}_g + q\bar{\rho}_b\big]}{8m}\mathbb{E}\big[\|\bar{w}^{(t)} - w^*\|^2\big] - \frac{2\eta_t\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\Gamma_g}{m} + \frac{6\eta_t^2\big[p\bar{\rho}_g + q\bar{\rho}_b\big]L\Gamma_g}{m}
\end{aligned}
\tag{11}
$$

where equation (11) is due to μ strong convexity and we used $-2\eta_t(1-2L\eta_t)(1-L\eta_t) \leq -\frac{3}{4}\eta_t$ and $-(1-2L\eta_t)(1-L\eta_t) \leq -(1-3L\eta_t)$. Hence, the bound of A_5 is as follows:

$$
\begin{aligned}
&\frac{4L\eta_t}{m}\mathbb{E}\big[\sum_{k\in S(t)}\big[(F_k(w_k^{(t)}) - F_k^*) - \frac{2\eta_t}{m}(F_k(w_k^{(t)} - F_k(w^*))\big]\big]\\
&\leq -\frac{3\eta_t\mu\big[p\bar{\rho}_g + q\bar{\rho}_b\big]}{8m}\mathbb{E}\big[\|\bar{w}^{(t)} - w^*\|^2 + \eta_t^2\left(\frac{6\big[p\bar{\rho}_g + q\bar{\rho}_b\big]L\Gamma_g}{m} + 16\tau^2G^2\right)\\
&\quad -\frac{2\eta_t\big[p\bar{\rho}_g + q\bar{\rho}_b\big]\Gamma_g}{m} + \frac{2\eta_t\big[p\tilde{\rho}_g + q\tilde{\rho}_b\big]\Gamma_g}{m}.
\end{aligned}
\tag{12}
$$

Finally, using equation (8), and (12) we can bound $\|\bar{w}^{(t+1)} - w^*\|$ as follows:

$$
\begin{aligned}
\mathbb{E}\big[\|\bar{w}^{(t+1)} - w^*\|\big] &\leq \big[1 - \eta_t\mu\big[1 + \frac{3(p\bar{\rho}_g + q\bar{\rho}_b)}{8m}\big]\big]\mathbb{E}\big[\bar{w}^{(t)} - w^*\|^2\big]\\
&\quad + \eta_t^2\big[32\tau^2G^2 + \frac{\sigma^2}{m} + \frac{6(p\bar{\rho}_g + q\bar{\rho}_b)L\Gamma_g}{m}\big] + \frac{2\eta_t\Gamma_g}{m}(p\tilde{\rho}_g + q\tilde{\rho}_b - p\bar{\rho}_g - q\bar{\rho}_b)\\
&\leq \big[1 - \eta_t\mu\big[1 + \frac{3(p\bar{\rho}_g + q\bar{\rho}_g)}{8m}\big]\big]\mathbb{E}\big[\bar{w}^{(t)} - w^*\|^2\big] + \eta_t^2\big[32\tau^2G^2 + \frac{\sigma^2}{m} + \frac{6(p\bar{\rho}_b + q\bar{\rho}_b)L\Gamma_g}{m}\big]\\
&\quad + \frac{2\eta_t\Gamma}{m}(p\tilde{\rho}_g + q\tilde{\rho}_b - p\bar{\rho}_g - q\bar{\rho}_g).
\end{aligned}
\tag{13}
$$

Equation (13) is obtained using $\bar{\rho}_g \leq \bar{\rho}_b$, which gives

$$
\begin{aligned}
\mathbb{E}\big[\|\bar{w}^{(t+1)} - w^*\|\big] &\leq \big[1 - \eta_t\mu\big[1 + \frac{3\bar{\rho}_g}{8}\big]\big]\mathbb{E}\big[\bar{w}^{(t)} - w^*\|^2\big]\\
&\quad + \eta_t^2\big[32\tau^2G^2 + \frac{\sigma^2}{m} + 6\bar{\rho}_bL\Gamma_g\big] + \frac{2\eta_t\Gamma_g}{m}(p\tilde{\rho}_g + q\tilde{\rho}_b - m\bar{\rho}_g).
\end{aligned}
$$

By setting $\Delta_{t+1} = \mathbb{E}\big[\|\bar{w}^{(t+1)} - w^*\|^2\big], B = 1 + \frac{3\bar{\rho}_g}{8}, C = 32\tau^2G^2 + \frac{\sigma^2}{m} + 6\bar{\rho}_bL\Gamma_g,\ D = \frac{2\Gamma_g}{m}(p\tilde{\rho}_g + q\tilde{\rho}_b - m\bar{\rho}_g)$, we get

$$
\Delta_{t+1} \leq (1 - \eta_t\mu B)\Delta_t + \eta_t^2C + D\eta_t.
$$

For a decreasing stepsize, $\eta_t = \frac{\beta}{t+\gamma}$ for some $\beta > \frac{1}{\mu B}$, $\gamma > 0$, we have that $\Delta_t \le \frac{\psi}{t+\gamma}$, where

$$\psi = \max\left\{(\gamma+1)\|\bar{w}^{(1)} - w^*\|^2, \frac{\beta^2 C + \beta D(t+\gamma)}{\beta\mu B - 1}\right\}.$$

This can be shown by induction on t (see Lemma VII.8 below). Then using the $L-$smoothness of $F(\cdot)$ we get

$$\mathbb{E}\big[F(\bar{w}^{(t)}] - F^* \le \frac{L}{2}\Delta_t \le \frac{L}{2}\frac{\psi}{\gamma+t}.$$

Now for $\beta = \frac{1}{\mu}$, we get

$$\mathbb{E}[F(\bar{w}^{(T)})] - F^* \le \frac{1}{(T+\gamma)}\left[\frac{4L(32\tau^2G^2 + \sigma^2/m)}{3\mu^2\bar{\rho}_g} + \frac{8L^2\Gamma_g}{\mu^2}\frac{\bar{\rho}_b}{\bar{\rho}_g} + \frac{L(\gamma+1)(\|\bar{w}^{(1)} - w^*\|^2)}{2}\right] + \frac{8L\Gamma_g}{3\mu}\left(\frac{p\tilde{\rho}_g + q\tilde{\rho}_b}{m\bar{\rho}_g} - 1\right),$$

which completes the proof of the theorem. □

Lemma VII.8. *For a decreasing stepsize, $\eta_t = \frac{\beta}{t+\gamma}$ for some $\beta > \frac{1}{\mu B}$, $\gamma > 0$,*

$$\Delta_t \le \frac{\psi}{t+\gamma} \tag{14}$$

where,

$$\psi = \max\big\{(\gamma+1)\|\bar{w}^{(1)} - w^*\|^2, \frac{1}{\beta\mu B - 1}(\beta^2 C + D\beta(t+\gamma))\big\} \tag{15}$$

and

$$\Delta_{t+1} \le (1 - \eta_t\mu B)\Delta_t + \eta_t^2 C + \eta_t D.$$

Proof. For $t = 1$, equation (14) holds clearly as (using (15))

$$\Delta_1 \le \frac{\psi}{\gamma+1} \le \|\bar{w}^{(1)} - w^*\|^2 = \Delta_1$$

Assume that it holds for some t, then

$$\begin{aligned}
\Delta_{t+1} &\le (1 - \eta_t\mu B)\Delta_t + \eta_t^2 C + \eta_t D \\
&\le (1 - \frac{\beta}{t+\gamma}\mu B)\frac{\psi}{t+\gamma} + \frac{\beta^2}{(t+\gamma)^2}C + \frac{\beta}{t+\gamma}D \\
&= \frac{t+\gamma-\beta\mu B}{(t+\gamma)^2}\psi + \frac{\beta^2 C + \beta D(t+\gamma)}{(t+\gamma)^2} \\
&= \frac{t+\gamma-1}{(t+\gamma)^2}\psi + \frac{\beta^2 C + \beta D(t+\gamma)}{(t+\gamma)^2} - \frac{\beta\mu B - 1}{(t+\gamma)^2}\psi \\
&= \frac{t+\gamma-1}{(t+\gamma)^2}\psi \qquad \text{(Using (15))} \\
&\le \frac{t+\gamma-1}{(t+\gamma)^2 - 1}\psi = \frac{\psi}{t+\gamma+1}
\end{aligned}$$

□