A PROOF OF THEOREM 1

Proof. To prove the covergence of *OSP*,we firstly bound the update of one step $F(\omega_{t+1}) - F(\omega_t)$ and then summarize all steps from 1 to N to achieve the overall convergence.

According to algorithms, the one-step updated fomula of global parameter ω_{t+1} is

$$\omega_{t+1} = \omega_t - \eta_t \frac{1}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} g(\omega_{t-1,k}^i, \xi_{t,k}^i) \tag{1}$$

Combined with Assumption 1, the bound of one iteration is

$$\begin{aligned} F(\omega_{t+1}) - F(\omega_t) &\leq \langle \nabla F(\omega_t), \omega_{t+1} - \omega_t \rangle + \frac{L}{2} \|\omega_{t+1} - \omega_t\|_2^2 \\ &= \left\langle \nabla F(\omega_t), -\eta_t \frac{1}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} g(\omega_{t-1,k}^i, \xi_{t,k}^i) \right\rangle \\ &+ \frac{L}{2} \|\eta_t \frac{1}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} g(\omega_{t-1,k}^i, \xi_{t,k}^i)\|_2^2 \end{aligned} \tag{2}$$

Because the $\xi_{t,k}^i$ in our algorithm are i.i.d. for all t, k, i, by taking expectation for equation (2) related to $\xi_{t,k}^i$ we can immediately get

$$\begin{aligned} \mathbb{E}F(\omega_{t+1}) - F(\omega_t) &\leq \underbrace{\left\langle \nabla F(\omega_t), -\eta_t \frac{1}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} \nabla F(\omega_{t-1,k}^i) \right\rangle}_{T_1} \\ &+ \underbrace{\frac{L}{2} \mathbb{E} \|\eta_t \frac{1}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} g(\omega_{t-1,k}^i, \xi_{t,k}^i)\|_2^2}_{T_2} \end{aligned} \tag{3}$$

We here below bound T_1 and T_2 respectively.

Bound T_1.

$$T_1 = \left\langle \nabla F(\omega_t), -\eta_t \frac{1}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} \nabla F(\omega_{t-1,k}^i) \right\rangle \tag{4}$$

$$= -\frac{\eta_t}{P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} \left\langle \nabla F(\omega_t), \nabla F(\omega_{t-1,k}^i) \right\rangle \tag{5}$$

$$= -\frac{\eta_t}{2P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} [\|\nabla F(\omega_t)\|_2^2 + \|\nabla F(\omega_{t-1,k}^i)\|_2^2] \tag{6}$$

$$+ \frac{\eta_t}{2P} \sum_{i=1}^{P} \sum_{k=0}^{K_t^i - 1} \underbrace{\|\nabla F(\omega_t) - \nabla F(\omega_{t-1,k}^i)\|_2^2}_{T_3} \tag{7}$$

According to Assumption 1, T_3 can be bounded as

$$\begin{aligned} T_3 &= \|\nabla F(\omega_t) - \nabla F(\omega_{t-1,k}^i)\|_2^2 \\ &\leq L^2 \|\omega_t - \omega_{t-1,k}^i\|_2^2 \\ &\leq \underbrace{2L^2 \|\omega_t - \omega_{t-1}\|_2^2}_{T_4} + \underbrace{2L^2 \|\omega_{t-1} - \omega_{t-1,k}^i\|_2^2}_{T_5} \end{aligned} \tag{8}$$

For clarity, we bound T_4 and T_5 respectively. For T_4

$$
\begin{aligned}
T_4 &= 2L^2\|\omega_t - \omega_{t-1}\|_2^2 \\
&= 2L^2\|\eta_{t-1}\frac{1}{P^2}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1} g(\omega_{t-2,k}^i, \xi_{t-1,k}^i)\|_2^2 \\
&= \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}[g(\omega_{t-2,k}^i, \xi_{t-1,k}^i) - \nabla F(\omega_{t-2,k}^i)] \\
&+ \sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&= \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}[g(\omega_{t-2,k}^i, \xi_{t-1,k}^i) - \nabla F(\omega_{t-2,k}^i)]\|_2^2 \\
&+ \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2
\end{aligned} \tag{9}
$$

Taking expectation in terms of $\xi_{t-1,k}^i$ for T_4, we have

$$
\begin{aligned}
\mathbb{E}T_4 &= \frac{2\eta_{t-1}^2L^2}{P^2}\mathbb{E}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}[g(\omega_{t-2,k}^i, \xi_{t-1,k}^i) - \nabla F(\omega_{t-2,k}^i)]\|_2^2 \\
&+ \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&\le \frac{2\eta_{t-1}^2L^2}{P^2}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\sigma_2 + \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&\le \frac{2\eta_{t-1}^2L^2\sigma^2}{P^2}\sum_{i=1}^{P}K_{t-1}^i + \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2
\end{aligned} \tag{10}
$$

As for T_5, we have

$$
\begin{aligned}
T_5 &= 2L^2\|\omega_{t-1} - \omega_{t-1,k}^i\|_2^2 \\
&= 2L^2\|\eta_t\sum_{j=0}^{k-1} g(\omega_{t-1,j}^i, \xi_{t-1,j}^i)\|_2^2 \\
&= 2\eta_t^2L^2\|\sum_{j=0}^{k-1}[g(\omega_{t-1,j}^i, \xi_{t-1,j}^i) - \nabla F(\omega_{t-1,j}^i)] + \sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2 \\
&= 2\eta_t^2L^2\|\sum_{j=0}^{k-1}[g(\omega_{t-1,j}^i, \xi_{t-1,j}^i) - \nabla F(\omega_{t-1,j}^i)]\|_2^2 \\
&+ 2\eta_t^2L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2
\end{aligned} \tag{11}
$$

Similar to $T4$, taking expectation in terms of $\xi_{t-1,j}^i$ for T_5, we can get

$$
\mathbb{E}T_5 \le 2\eta_t^2L^2k\sigma_2 + +2\eta_t^2L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2 \tag{12}
$$

Putting T_4, T_5 into T_3 and then putting T_3 back into T_1, we get

$$
\begin{aligned}
T_1 \le & -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\|\nabla F(\omega_t)\|_2^2 + \|\nabla F(\omega_{t-1,k}^i)\|_2^2] \\
& + \frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\frac{2\eta_{t-1}^2 L^2\sigma^2}{P^2}\sum_{i=1}^{P}K_{t-1}^i + 2\eta_t^2 L^2 k\sigma_2 \\
& + \frac{2\eta_{t-1}^2 L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 + 2\eta_t^2 L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2]
\end{aligned} \tag{13}
$$

Bound T_2.

$$
\begin{aligned}
T_2 &= \frac{L}{2}\mathbb{E}\|\eta_t\frac{1}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}g(\omega_{t-1,k}^i,\xi_{t,k}^i)\|_2^2 \\
&= \frac{\eta_t^2 L}{2P^2}\mathbb{E}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[g(\omega_{t-1,k}^i,\xi_{t,k}^i) - \nabla F(\omega_{t-1,k}^i)] \\
&+ \sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&= \frac{\eta_t^2 L}{2P^2}\mathbb{E}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[g(\omega_{t-1,k}^i,\xi_{t,k}^i) - \nabla F(\omega_{t-1,k}^i)]\|_2^2 \\
&+ \frac{\eta_t^2 L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&\le \frac{\eta_t^2 L\sigma^2}{2P^2}\sum_{i=1}^{P}K_t^i + \frac{\eta_t^2 L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2
\end{aligned} \tag{14}
$$

After achieving T_1 and T_2, the bound of $\mathbb{E}F(\omega_{t+1}) - F(\omega_t)$ in equation (3) can transformed into

$$
\begin{aligned}
&\mathbb{E}F(\omega_{t+1}) - F(\omega_t) \leq T_1 + T_2 \\
&\leq -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\|\nabla F(\omega_t)\|_2^2 + \|\nabla F(\omega_{t-1,k}^i)\|_2^2] \\
&+ \frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\frac{2\eta_{t-1}^2 L^2\sigma^2}{P^2}\sum_{i=1}^{P}K_{t-1}^i + 2\eta_t^2 L^2 k\sigma^2 \\
&+ \frac{2\eta_{t-1}^2 L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 + 2\eta_t^2 L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2] \\
&+ \frac{\eta_t^2 L\sigma^2}{2P^2}\sum_{i=1}^{P}K_t^i + \frac{\eta_t^2 L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&= -\frac{\bar{K}\eta_t}{2}\|\nabla F(\omega_t)\|_2^2 + [\frac{\eta_t^2 L\sigma_2}{2P^2}\sum_{i=1}^{P}K_t^i \\
&+ \frac{\eta_t^3 L^2\sigma^2}{2P}\sum_{i=1}^{P}(K_t^i - 1)K_t^i + \frac{\eta_t\eta_{t-1}^2 L^2\sigma^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\sum_{i=1}^{P}K_{t-1}^i] \\
&+ [-\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 + \frac{\eta_t^2 L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&+ \frac{\eta_t^3 L^2}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2 \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2]
\end{aligned} \tag{15}
$$

We further simplify the bound of $\mathbb{E}F(\omega_{t+1}) - F(\omega_t)$. According to Assumption 5 and 6, we have

$$
\begin{aligned}
T_6 &= \frac{\eta_t^2 L\sigma_2}{2P^2}\sum_{i=1}^{P}K_t^i + \frac{\eta_t^3 L^2\sigma^2}{2P}\sum_{i=1}^{P}(K_t^i - 1)K_t^i \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2\sigma^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\sum_{i=1}^{P}K_{t-1}^i \\
&\leq \frac{\eta_t^3 L^2\sigma^2 M}{2} + \frac{\eta_t^3 L^2\sigma^2\bar{K}^2}{2} - \frac{\eta_t^3 L^2\sigma^2\bar{K}}{2} + \frac{\eta_t\eta_{t-1}^2 L^2\sigma^2\bar{K}^2}{P} + \frac{\eta_t^2 L\sigma^2\bar{K}}{2P}
\end{aligned} \tag{16}
$$

After bounding T_6, we further bound

$$
\begin{aligned}
&T_7 \\
&= -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 + \frac{\eta_t^2 L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&+ \frac{\eta_t^3 L^2}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2 \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&\leq -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 + \frac{\eta_t^2 L}{2P}\sum_{i=1}^{P}K_t^i\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&+ \frac{\eta_t^3 L^2}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}k\sum_{j=0}^{k-1}\|\nabla F(\omega_{t-1,j}^i)\|_2^2 \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2}{P^2}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\sum_{i=1}^{P}K_{t-1}^i\sum_{l=0}^{K_{t-1}^i-1}\|\nabla F(\omega_{t-2,l}^i)\|_2^2 \\
&\leq -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 + \frac{\eta_t^2 L}{2P}\sum_{i=1}^{P}K_t^i\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&+ \frac{\eta_t^3 L^2}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}(K_t^i-1)K_t^i\|\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2\bar{K}}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}K_{t-1}^i\|\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&= \sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2[-\frac{\eta_t}{2P} + \frac{\eta_t^2 L K_t^i}{2P} + \frac{\eta_t^3 L^2(K_t^i-1)K_t^i}{2P}] \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2\bar{K}}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}K_{t-1}^i\|\nabla F(\omega_{t-2,k}^i)\|_2^2
\end{aligned} \tag{17}
$$

where the second-to-last iequality is due to Assumption 5 and 6. Sumarize T_7 from 1 to N

$$
\begin{aligned}
&\sum_{t=1}^{N}T_7 \\
&\leq \sum_{t=1}^{N}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2[-\frac{\eta_t}{2P} + \frac{\eta_t^2 L K_t^i}{2P} + \frac{\eta_t^3 L^2(K_t^i-1)K_t^i}{2P}] \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2\bar{K}}{P}\sum_{t=1}^{N}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}K_{t-1}^i\|\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&\leq \sum_{t=1}^{N}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2[-\frac{\eta_t}{2P} + \frac{\eta_t^2 L K_t^i}{2P} + \frac{\eta_t^3 L^2(K_t^i-1)K_t^i}{2P} \\
&+ \frac{\eta_t\eta_{t-1}^2 L^2\bar{K}K_t^i}{P}]
\end{aligned} \tag{18}
$$

Obviously, if

$$
LK_{max}(\eta_t + \eta_t^2 LK_{max} + 2\eta_{t-1}^2 L\bar{K} - \eta_t^2 L) \leq 1 \tag{19}
$$

then

$$\sum_{t=1}^{N} T_7 < 0 \tag{20}$$

So, sumarizing $F(\omega_{t+1}) - F(\omega_t)$ from $t = 1$ to $t = N$, we could achieve

$$\begin{aligned}
&\mathbb{E}F(\omega_{N+1}) - F(\omega_1) \\
&\leq -\frac{\bar{K}}{2}\sum_{t=1}^{N}\eta_t\|\nabla F(\omega_t)\|_2^2 + \sum_{t=1}^{N} T_6 + \sum_{t=1}^{N} T_7 \\
&\leq -\frac{\bar{K}}{2}\sum_{t=1}^{N}\eta_t\|\nabla F(\omega_t)\|_2^2 + \sum_{t=1}^{N}[\frac{\eta_t^3 L^2\sigma^2 M}{2} + \frac{\eta_t^3 L^2\sigma^2\bar{K}^2}{2} \\
&- \frac{\eta_t^3 L^2\sigma^2\bar{K}}{2} + \frac{\eta_t\eta_{t-1}^2 L^2\sigma^2\bar{K}^2}{P} + \frac{\eta_t^2 L\sigma^2\bar{K}}{2P}]
\end{aligned} \tag{21}$$

Moving sum of squared gradient to left

$$\begin{aligned}
&\sum_{t=1}^{N}\eta_t\mathbb{E}\|\nabla F(\omega_t)\|_2^2 \leq \frac{1}{\bar{K}}[2(F(\omega_1) - F(\omega^\star)) + \sum_{t=1}^{N}\eta_t[\eta_t^2 L^2\sigma^2 M \\
&+ \eta_t^2 L^2\sigma^2\bar{K}^2 - \eta_t^2 L^2\sigma^2\bar{K} + \frac{2\eta_{t-1}^2 L^2\sigma^2\bar{K}^2}{P} + \frac{2\eta_t L\sigma^2\bar{K}}{P}]]
\end{aligned} \tag{22}$$

As the learning rate is fixed $\eta_t = \bar{\eta}$, the condition 19 is equivalent to $\bar{\eta}LK_{max}(\bar{\eta}LK_{max} + 2\bar{\eta}L\bar{K} + 1 - \bar{\eta}L) \leq 1$, and then the bound of average squared gradient is

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\|\nabla F(\omega_t)\|_2^2 \\
&\leq \frac{2[F(\omega_1) - F(\omega^\star)]}{N\bar{\eta}\bar{K}} + (\frac{\bar{\eta}LM}{\bar{K}} + \bar{\eta}L\bar{K} - \bar{\eta}L + \frac{2\bar{\eta}L\bar{K}}{P} + \frac{1}{P})\bar{\eta}L\sigma^2
\end{aligned} \tag{23}$$

which completes the proof.

□