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

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

Based on Assumption 1 and the conclusion of Appendix B of [4], 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 both sides of (2) upon $\xi_{t,k}^i$ and combining Asssumption 3 we can immediately get

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

in which the two bounds T_1 and T_2 are derived respectively.

Bound T_1.

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

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

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

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

where the third item is because $-2 < a, b >= \|a - b\|^2 - \|a\|^2 - \|b\|^2$. 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

$$
\begin{aligned}
T_4 &= 2L^2\|\omega_t - \omega_{t-1}\|_2^2 \\
&= 2L^2\|\eta_{t-1}\frac{1}{P^2}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1} g(\omega_{t-2,k}^i, \xi_{t-1,k}^i)\|_2^2 \\
&= \frac{2\eta_{t-1}^2 L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}[g(\omega_{t-2,k}^i, \xi_{t-1,k}^i) - \nabla F(\omega_{t-2,k}^i)] + \sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&= \frac{2\eta_{t-1}^2 L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}[g(\omega_{t-2,k}^i, \xi_{t-1,k}^i) - \nabla F(\omega_{t-2,k}^i)]\|_2^2 + \frac{2\eta_{t-1}^2 L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&+ \frac{4\eta_{t-1}^2 L^2}{P^2}\left\langle\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}[g(\omega_{t-2,k}^i, \xi_{t-1,k}^i) - \nabla F(\omega_{t-2,k}^i)]\|_2^2, \frac{2\eta_{t-1}^2 L^2}{P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,k}^i)\right\rangle
\end{aligned} \tag{9}
$$

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

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

where the second equality is because the fact that if $a_1, a_2, \ldots, a_n$ are i.i.d. and $\mathbb{E}a_i = 0$ for any $i = 1, 2, \ldots, n$, then

$$
\begin{aligned}
Var(\sum_{i=1}^{n} a_i) &= \mathbb{E}\|\sum_{i=1}^{n} a_i\|_2^2 \\
&= \sum_{i=1}^{n} Var(a_i) \\
&= \sum_{i=1}^{n}\mathbb{E}\|a_i\|_2^2,
\end{aligned} \tag{11}
$$

and the third iequality is due to the Assumption 4. As for T_5, we have

$$
\begin{aligned}
T_5 &= 2L^2\|\omega_{t-1} - \omega_{t-1,k}^i\|_2^2 \\
&= 2L^2\|\eta_t\sum_{j=0}^{k-1} g(\omega_{t-1,j}^i, \xi_{t-1,j}^i)\|_2^2 \\
&= 2\eta_t^2 L^2\|\sum_{j=0}^{k-1}[g(\omega_{t-1,j}^i, \xi_{t-1,j}^i) - \nabla F(\omega_{t-1,j}^i)] + \sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2 \\
&= 2\eta_t^2 L^2\|\sum_{j=0}^{k-1}[g(\omega_{t-1,j}^i, \xi_{t-1,j}^i) - \nabla F(\omega_{t-1,j}^i)]\|_2^2 + 2\eta_t^2 L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2.
\end{aligned} \tag{12}
$$

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

$$
\mathbb{E}T_5 \le 2\eta_t^2 L^2 k\sigma^2 + +2\eta_t^2 L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2. \tag{13}
$$

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

$$
\begin{aligned}
T_1 \leq & -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\|\nabla F(\omega_t)\|_2^2 + \|\nabla F(\omega_{t-1,k}^i)\|_2^2] \\
& + \frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\Big[\frac{2\eta_{t-1}^2L^2\sigma^2}{P^2}\sum_{j=1}^{P}K_{t-1}^j + 2\eta_t^2L^2k\sigma^2 \\
& + \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{j=1}^{P}\sum_{l=0}^{K_{t-1}^j-1}\nabla F(\omega_{t-2,l}^j)\|_2^2 + 2\eta_t^2L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2\Big]
\end{aligned} \tag{14}
$$

Bound T_2.

$$
\begin{aligned}
T_2 &= \frac{L}{2}\mathbb{E}\|\eta_t\frac{1}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}g(\omega_{t-1,k}^i,\xi_{t,k}^i)\|_2^2 \\
&= \frac{\eta_t^2L}{2P^2}\mathbb{E}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[g(\omega_{t-1,k}^i,\xi_{t,k}^i) - \nabla F(\omega_{t-1,k}^i)] + \sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&= \frac{\eta_t^2L}{2P^2}\mathbb{E}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[g(\omega_{t-1,k}^i,\xi_{t,k}^i) - \nabla F(\omega_{t-1,k}^i)]\|_2^2 + \frac{\eta_t^2L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&\leq \frac{\eta_t^2L\sigma^2}{2P^2}\sum_{i=1}^{P}K_t^i + \frac{\eta_t^2L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2,
\end{aligned} \tag{15}
$$

where the derive of third equality is similar to that of (9) to first item of (10) and the last iequality is the last two iequalities of (10). Now putting T_1 and T_2 back into (5)

$$
\begin{aligned}
&\mathbb{E}F(\omega_{t+1}) - F(\omega_t) \leq T_1 + T_2 \\
&\leq -\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\|\nabla F(\omega_t)\|_2^2 + \|\nabla F(\omega_{t-1,k}^i)\|_2^2] \\
&+ \frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}[\frac{2\eta_{t-1}^2L^2\sigma^2}{P^2}\sum_{j=1}^{P}K_{t-1}^j + 2\eta_t^2L^2k\sigma^2 \\
&+ \frac{2\eta_{t-1}^2L^2}{P^2}\|\sum_{j=1}^{P}\sum_{l=0}^{K_{t-1}^j-1}\nabla F(\omega_{t-2,l}^j)\|_2^2 + 2\eta_t^2L^2\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2] \\
&+ \frac{\eta_t^2L\sigma^2}{2P^2}\sum_{i=1}^{P}K_t^i + \frac{\eta_t^2L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 \\
&= -\frac{\bar{K}\eta_t}{2}\|\nabla F(\omega_t)\|_2^2 \\
&+ \underbrace{[\frac{\eta_t^2L\sigma^2}{2P^2}\sum_{i=1}^{P}K_t^i + \frac{\eta_t^3L^2\sigma^2}{2P}\sum_{i=1}^{P}(K_t^i-1)K_t^i + \frac{\eta_t\eta_{t-1}^2L^2\sigma^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\sum_{j=1}^{P}K_{t-1}^j]}_{T_6} \\
&+ \underbrace{[-\frac{\eta_t}{2P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\nabla F(\omega_{t-1,k}^i)\|_2^2 + \frac{\eta_t^2L}{2P^2}\|\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\nabla F(\omega_{t-1,k}^i)\|_2^2 + \frac{\eta_t^3L^2}{P}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\sum_{j=0}^{k-1}\nabla F(\omega_{t-1,j}^i)\|_2^2 + \frac{\eta_t\eta_{t-1}^2L^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\|\sum_{j=1}^{P}\sum_{l=0}^{K_{t-1}^i-1}\nabla F(\omega_{t-2,l}^j)\|_2^2]}_{T_7}
\end{aligned} \tag{16}
$$

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

$$
\begin{aligned}
T_6 &= \frac{\eta_t^2 L\sigma^2}{2P^2}\sum_{i=1}^{P} K_t^i + \frac{\eta_t^3 L^2\sigma^2}{2P}\sum_{i=1}^{P}(K_t^i - 1)K_t^i + \frac{\eta_t\eta_{t-1}^2 L^2\sigma^2}{P^3}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1}\sum_{j=1}^{P} K_{t-1}^j \\
&\leq \frac{\eta_t^3 L^2\sigma^2 M}{2} + \frac{\eta_t^3 L^2\sigma^2\bar{K}^2}{2} - \frac{\eta_t^3 L^2\sigma^2\bar{K}}{2} + \frac{\eta_t\eta_{t-1}^2 L^2\sigma^2\bar{K}^2}{P} + \frac{\eta_t^2 L\sigma^2\bar{K}}{2P}.
\end{aligned}
\tag{17}
$$

Similarly, we show bound of T_7

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

where the second iequality is due to Cauchy–Schwarz inequality and the second-to-last iequality is due to Assumption 5 and 6. When considering all N iterations, we have

$$\begin{aligned}
&\sum_{t=1}^{N} T_7 \\
&\leq \sum_{t=1}^{N}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1} \|\nabla F(\omega_{t-1,k}^i)\|_2^2[-\frac{\eta_t}{2P} + \frac{\eta_t^2 L K_t^i}{2P} + \frac{\eta_t^3 L^2 (K_t^i - 1)K_t^i}{2P}] \\
&+ \frac{\eta_t \eta_{t-1}^2 L^2 \bar{K}}{P} \sum_{t=1}^{N}\sum_{i=1}^{P}\sum_{k=0}^{K_{t-1}^i-1} K_{t-1}^i \|\nabla F(\omega_{t-2,k}^i)\|_2^2 \\
&\leq \sum_{t=1}^{N}\sum_{i=1}^{P}\sum_{k=0}^{K_t^i-1} \|\nabla F(\omega_{t-1,k}^i)\|_2^2[-\frac{\eta_t}{2P} + \frac{\eta_t^2 L K_t^i}{2P} + \frac{\eta_t^3 L^2 (K_t^i - 1)K_t^i}{2P} + \frac{\eta_t \eta_{t-1}^2 L^2 \bar{K} K_t^i}{P}]
\end{aligned} \tag{19}$$

Obviously, if

$$LK_{max}(\eta_t + \eta_t^2 LK_{max} + 2\eta_{t-1}^2 L\bar{K} - \eta_t^2 L) \leq 1, \tag{20}$$

then

$$\sum_{t=1}^{N} T_7 \leq 0. \tag{21}$$

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

$$\begin{aligned}
&\mathbb{E}F(\omega_{N+1}) - F(\omega_1) \\
&\leq -\frac{\bar{K}}{2}\sum_{t=1}^{N} \eta_t \|\nabla F(\omega_t)\|_2^2 + \sum_{t=1}^{N} T_6 + \sum_{t=1}^{N} T_7 \\
&\leq -\frac{\bar{K}}{2}\sum_{t=1}^{N} \eta_t \|\nabla F(\omega_t)\|_2^2 + \sum_{t=1}^{N}[\frac{\eta_t^3 L^2 \sigma^2 M}{2} + \frac{\eta_t^3 L^2 \sigma^2 \bar{K}^2}{2} \\
&- \frac{\eta_t^3 L^2 \sigma^2 \bar{K}}{2} + \frac{\eta_t \eta_{t-1}^2 L^2 \sigma^2 \bar{K}^2}{P} + \frac{\eta_t^2 L \sigma^2 \bar{K}}{2P}].
\end{aligned} \tag{22}$$

By defining $\omega_\star$ is the global optimal point that achieves the F_{inf} in Assumption 2, we have the fact

$$F(\omega_\star) - F(\omega_1) \leq \mathbb{E}F(\omega_{N+1}) - F(\omega_1) \tag{23}$$

And thus, moving sum of squared gradient to left, we have

$$\begin{aligned}
\sum_{t=1}^{N} \eta_t \mathbb{E}\|\nabla F(\omega_t)\|_2^2 &\leq \frac{1}{\bar{K}}[2(F(\omega_1) - F(\omega^\star)) + \sum_{t=1}^{N} \eta_t [\eta_t^2 L^2 \sigma^2 M \\
&+ \eta_t^2 L^2 \sigma^2 \bar{K}^2 - \eta_t^2 L^2 \sigma^2 \bar{K} + \frac{2\eta_{t-1}^2 L^2 \sigma^2 \bar{K}^2}{P} + \frac{2\eta_t L \sigma^2 \bar{K}}{P}]]
\end{aligned} \tag{24}$$

As the learning rate is fixed $\eta_t = \bar{\eta}$, the condition 20 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

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N} \mathbb{E}\|\nabla F(\omega_t)\|_2^2 \\
&\leq \frac{2[F(\omega_1) - F(\omega^\star)]}{N\bar{\eta}\bar{K}} + (\frac{\bar{\eta} LM}{\bar{K}} + \bar{\eta} L\bar{K} - \bar{\eta} L + \frac{2\bar{\eta} L\bar{K}}{P} + \frac{1}{P})\bar{\eta} L\sigma^2
\end{aligned} \tag{25}$$

which completes the proof. □

Corollary 1. *Under the condition of Theorem 1, take*

$$\bar{\eta} = \sqrt{\frac{(F(\omega_1) - F(\omega^\star))P}{\bar{K} L \sigma^2 N}} \tag{26}$$

Then for any iteration times

$$N \geq \frac{(F(\omega_1) - F(\omega^\star))(\frac{LM}{\bar{K}} + L\bar{K} - L + \frac{2L\bar{K}}{P})P^3}{\bar{K}L\sigma^2} \tag{27}$$

the output of Alg. 1 and Alg. 2 satisfies the following ergodic convergence rate

$$\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\|\nabla F(\omega_t)\|_2^2 \leq 4\sqrt{\frac{(F(\omega_1) - F(\omega^\star))\sigma^2}{\bar{K}P}} * \frac{1}{\sqrt{N}} \tag{28}$$

Proof. Assume $(\frac{LM}{\bar{K}} + L\bar{K} - L + \frac{2L\bar{K}}{P})\bar{\eta} \leq \frac{1}{P}$, and then the bound of Threorem 1 can be transformed into:

$$\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\|\nabla F(\omega_t)\|_2^2 \leq \frac{2[F(\omega_1) - F(\omega_\star)]}{N\bar{\eta}\bar{K}} + \frac{2\bar{\eta}L\sigma^2}{P} \tag{29}$$

Let $f(\bar{\eta}) = \frac{2[F(\omega_1)-F(\omega_\star)]}{N\bar{\eta}\bar{K}} + \frac{2\bar{\eta}L\sigma^2}{P}$, then

$$\bar{\eta} = \sqrt{\frac{2[F(\omega_1) - F(\omega_\star)]}{N\bar{K}L\sigma^2}} \tag{30}$$

when $f'(\bar{\eta}) = 0$. Accordingly,

$$f(\bar{\eta}) \leq 4\sqrt{\frac{(F(\omega_1) - F(\omega^\star))\sigma^2}{\bar{K}P}} * \frac{1}{\sqrt{N}} \tag{31}$$

Put (30) into condition $(\frac{LM}{\bar{K}} + L\bar{K} - L + \frac{2L\bar{K}}{P})\bar{\eta} \leq \frac{1}{P}$, and the bound (27) of N can be achieved. □

Theorem 2. *(Nonconvex objective, diminishing stepsize) Suppose algorithm is run with diminishing learning rate η_t satisfying*

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

where $K_{max} = \max\{K_t^i,\ t = 1, 2, \ldots, N \text{ and } i = 1, 2, \ldots, P\}$. Then the expected average squared gradient norms satisfy the following bounds for all $N \in \mathbb{N}$:

$$\begin{aligned}\frac{1}{\sum\limits_{t=1}^{N}\eta_t}\sum_{t=1}^{N}\eta_t\mathbb{E}\|\nabla F(\omega_t)\|_2^2 \leq{}& \frac{1}{\bar{K}\sum\limits_{t=1}^{N}\eta_t}[2(F(\omega_1) - F(\omega^\star))\\ &+ \sum_{t=1}^{N}\eta_t[\eta_t^2 L^2\sigma^2 M + \eta_t^2 L^2\sigma^2\bar{K}^2 - \eta_t^2 L^2\sigma^2\bar{K}\\ &+ \frac{2\eta_{t-1}^2 L^2\sigma^2\bar{K}^2}{P} + \frac{2\eta_t L\sigma^2\bar{K}}{P}]]\end{aligned} \tag{32}$$

Proof. Under condition (20) with diminishing step size η_t, we divide $\frac{1}{\sum_{t=1}^{N}\eta_t}$ from both sides of equation (24), and we could have achieve the equation (32) which completes proof. □