7 APPENDIX

In the anonymous appendix, we give the detailed convergence analyses of AsyREVEL algorithms (i.e., Theorems 2 and 3 in the submitted manuscript). First, we introduce some notations necessary for the proof. Let $m \in \{0, [q]\}$, we define

$$\begin{aligned}
V_m^t &:= \nabla_m f_{\mu_m}(w_0^t, \mathbf{w}^t), \quad v_m^t := \nabla_m f_{i,\mu_m}(w_0^t, \mathbf{w}^t) \\
\bar{v}_m^t &:= \nabla_m f_{i,\mu_m}(w_0^t, \mathbf{w}^{t-\tau_t}), \quad \widehat{v}_0^t := \hat{\nabla}_{w_0} f_{i_t}(w_0, \mathbf{w}^{t-\tau_t}; y_{i_t}) \\
\widehat{v}_{m_t}^t &:= \begin{cases} \hat{\nabla}_{w_m} f_{i_t}(w_0^t, \mathbf{w}^{t-\tau_t}; y_{i_t}), & \text{if } m = m_t \\ 0, & \text{else} \end{cases} ; \quad m \in [q] \\
\widehat{\mathbf{w}} &:= [w_1^{t-\tau_t^1}; \cdots ; w_q^{t-\tau_t^1}]
\end{aligned} \tag{13}$$

We define $f_{\mu_m} = \mathbb{E}_{u_m \in \mathcal{D}}[f(w_m + \mu_m u_m)]$, where $\mathcal{D}$ is a multiple variants Gaussian distribution with unit variance and zero mean for AsyREVEL-Gau or a uniform distribution over a unit ball for AsyREVEL-Uni. We define

$$f_\mu = \mathbb{E}_{u_0, \cdots, u_q \in \mathcal{D}}[f(w_0 + \mu_0 u_0, \cdots, w_q + \mu_q u_q)]$$

LEMMA 1. *Suppose that Assumption 2 holds, then we have*
1) f_{μ_m} is L_m-smooth and f_μ is L-smooth

$$\nabla_m f_{\mu_m} = \mathbb{E}_{u_m}[\hat{\nabla}_m f(w_0, \mathbf{w})], \nabla f_\mu = \mathbb{E}_u[\hat{\nabla} f(w_0, \mathbf{w})] \tag{14}$$

where $\hat{\nabla}_m f(w_0, \mathbf{w})$ is given by Eq. (6).
2) For any $w_m \in \mathbb{R}^{d_m}$,

$$|f_{\mu_m}(w_m) - f(w_m)| \leq \frac{L_m d_m {\mu_m}^2}{2} \tag{15}$$

$$\|\nabla_m f_{\mu_m}(w_0, \mathbf{w}) - \nabla_m f(w_0, \mathbf{w})\|_2^2 \leq \frac{{\mu_m}^2 L_m^2 (d_m + 3)^3}{4}, \tag{16}$$

$$\mathbb{E}_u \left[\|\hat{\nabla}_m f(w_0, \mathbf{w})\|_2^2 \right] \leq 2(d_m + 4) \|\nabla_m f(w_0, \mathbf{w})\|_2^2 + \frac{{\mu_m}^2 L_m^2 (d_m + 6)^3}{2}. \tag{17}$$

3) For any $w_m \in \mathbb{R}^{d_m}$,

$$\mathbb{E}_{\mathbf{u}} \left[\|\hat{\nabla}_m f(w_0, \mathbf{w}) - \nabla_m f_{\mu_m}(w_0, \mathbf{w})\|_2^2 \right] \leq 2(2d_m + 9) \|\nabla_m f(w_0, \mathbf{w})\|_2^2 + {\mu_m}^2 L_m^2 (d_m + 6)^3. \tag{18}$$

LEMMA 2. *Under Assumptions 1 to 4, for $m = 0, 1, \cdots, q$ there is*

$$\begin{aligned}
\mathbb{E}\|\widehat{v}_{m_t}^t - v_{m_t}^t\|^2 &= \mathbb{E}\|\widehat{v}_{m_t}^t - \bar{v}_{m_t}^t + \bar{v}_{m_t}^t - v_{m_t}^t\|^2 \\
&\leq 2\mathbb{E}\|\widehat{v}_{m_t}^t - \bar{v}_{m_t}^t\| + 2\mathbb{E}\|\bar{v}_{m_t}^t - v_{m_t}^t\|^2 \\
&\leq \frac{{\mu_m}^2 L_{m_t}^2 (d_{m_t} + 3)^3}{2} + 2L^2 \|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
\mathbb{E}\|\widehat{v}_{m_t}^t\|^2 &= \mathbb{E}\|\widehat{v}_{m_t}^t - v_{m_t}^t + v_{m_t}^t\|^2 \\
&\leq 2\mathbb{E}\|\widehat{v}_{m_t}^t - v_{m_t}^t\| + 2\mathbb{E}\|v_{m_t}^t\|^2 \\
&\leq {\mu_m}^2 L_{m_t}^2 d_{m_t}^2 + 4L^2 \|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 + 2\mathbb{E}\|v_{m_t}^t\|^2 \\
&\leq 3{\mu_m}^2 L_{m_t}^2 (d_{m_t} + 3)^3 + 4L^2 \|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 + 4\sigma_{m_t}^2
\end{aligned} \tag{20}$$

PROOF. Under Assumption 2, and taking expectation w.r.t. the sample index i_t, we have

$$\begin{aligned}
&\mathbb{E} f_{\mu_m}(w_0^{t+1}, \mathbf{w}^{t+1}) \\
&= \mathbb{E}\left(f_{\mu_m}(w_0^t - \eta_0 \widehat{v}_0^t, \cdots, w_{m_t}^t - \eta_{m_t} \widehat{v}_{m_t}^t, \cdots) \right) \\
&\leq \mathbb{E}\left(f_{\mu_m}(w_0^t, \mathbf{w}^t) - \eta_0 \left\langle V_0^t, \widehat{v}_0^t \right\rangle - \eta_{m_t} \left\langle V_{m_t}^t, \widehat{v}_{m_t}^t \right\rangle + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2 \right) \\
&= \mathbb{E}\left(f_{\mu_m}(w_0^t, \mathbf{w}^t) - \eta_0 \left\langle V_0^t, \widehat{v}_0^t - v_0^t + v_0^t \right\rangle - \eta_{m_t} \left\langle V_m^t, \widehat{v}_{m_t}^t - v_{m_t}^t + v_{m_t}^t \right\rangle + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2 \right) \\
&\leq \mathbb{E}\left(f_{\mu_m}(w_0^t, \mathbf{w}^t) - \eta_0\|V_0^t\|^2 - \eta_0 \left\langle V_0^t, \widehat{v}_0^t - v_0^t \right\rangle - \eta_{m_t}\|V_{m_t}^t\|^2 - \eta_{m_t} \left\langle V_{m_t}^t, \widehat{v}_{m_t}^t - v_{m_t}^t \right\rangle + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2 \right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \frac{\eta_0}{2}\|V_0^t\|^2 + \frac{\eta_0}{2}\|\widehat{v}_0^t - v_0^t\|^2 - \frac{\eta_{m_t}}{2}\|V_{m_t}^t\|^2 + \frac{\eta_{m_t}}{2}\|\widehat{v}_{m_t}^t - v_{m_t}^t\| + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2\right)\\
&\leq \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \frac{\eta_0}{2}\|V_0^t\|^2 - \frac{\eta_{m_t}}{2}\|V_{m_t}^t\|^2 + (\eta_0 + \eta_{m_t} + 2L\eta_0^2 + 2L\eta_{m_t}^2)L^2\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2\right)\\
&+ (\frac{\eta_{m_t}}{4} + \frac{3L\eta_{m_t}^2}{2}){\mu_m}^2 L_{m_t}^2(d_{m_t}+3)^3 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2(d_0+3)^3 + 2L\eta_{m_t}^2\sigma_{m_t}^2 + 2L\eta_0^2\sigma_0^2
\end{aligned} \tag{21}
$$

Taking expectation w.r.t. m_t, and using Assumption 3, there is

$$
\begin{aligned}
&\mathbb{E}f_{\mu_m}(w_0^{t+1},\mathbf{w}^{t+1})\\
&\leq \mathbb{E}f_{\mu_m}(w_0^t,\mathbf{w}^t) - \frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^q p_m\frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \underbrace{(\eta_0 + 2L\eta_0^2 + \max_m(2L\eta_m^2 + \eta_m))L^2}_{\beta^t}\mathbb{E}\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2\\
&+ \sum_{m=1}^q p_m(\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2}){\mu_m}^2 L_m^2(d_m+3)^3 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2(d_0+3)^3 + \sum_{m=1}^q p_m 2L\eta_m^2\sigma_m^2 + 2L\eta_0^2\sigma_0^2
\end{aligned} \tag{22}
$$

According to Assumption 4, there is

$$
\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 = \|\sum_{i\in D(t)}\mathbf{w}^{i+1} - \mathbf{w}^i\|^2 \leq \tau\sum_{i=1}^{\tau}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 \tag{23}
$$

We than bound the term $\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2$. First, for $\mathbb{E}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2$

$$
\mathbb{E}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 = \mathbb{E}\eta_{m_t}^2\|\widehat{v}_{m_t}^2\| \leq \sum_{m=1}^q p_m\eta_m^2(3{\mu_m}^2 L_m^2(d_m+3)^3 + 4\sigma_m^2) + \max_m \eta_m^2 4L^2\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 \tag{24}
$$

Define a Lyapunov function as

$$
M^t = f_{\mu_m}(w_0^t,\mathbf{w}^t) + \sum_{i=1}^{\tau}\theta_i\|\mathbf{w}^{i+1} - \mathbf{w}^i\|^2 \tag{25}
$$

Following Lemma 1 and Eq. 25, there is

$$
\begin{aligned}
&\mathbb{E}(M^{t+1} - M^t)\\
&= \mathbb{E}\left(f_{\mu_m}(w_0^{t+1},\mathbf{w}^{t+1}) + \sum_{i=1}^{\tau}\theta_i\|\mathbf{w}^{t+1+1-i} - \mathbf{w}^{t+1-i}\|^2 - f_{\mu_m}(w_0^t,\mathbf{w}^t) - \sum_{i=1}^{\tau}\theta_i\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2\right)\\
&= -\frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^q p_m\frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \beta^t\mathbb{E}\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 + \sum_{m=1}^q p_m 2L\eta_m^2\sigma_m^2 + 2L\eta_0^2\sigma_0^2\\
&+ \sum_{m=1}^q p_m(\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2}){\mu_m}^2 L_m^2(d_m+3)^3 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2(d_0+3)^3\\
&+ \theta_1\mathbb{E}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 + \sum_{i=1}^{\tau-1}(\theta_{i+1} - \theta_i)\mathbb{E}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 - \theta_\tau\mathbb{E}\|\mathbf{w}^{t+1-\tau} - \mathbf{w}^{t-\tau}\|^2\\
&\leq -\frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^q p_m\frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \beta^t\tau\sum_{i=1}^{\tau}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 + \sum_{m=1}^q p_m 2L\eta_m^2\sigma_m^2 + 2L\eta_0^2\sigma_0^2\\
&+ \sum_{m=1}^q p_m(\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2}){\mu_m}^2 L_m^2(d_m+3)^3 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2(d_0+3)^3\\
&+ \sum_{i=1}^{\tau-1}(\theta_{i+1} - \theta_i)\mathbb{E}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 - \theta_\tau\mathbb{E}\|\mathbf{w}^{t+1-\tau} - \mathbf{w}^{t-\tau}\|^2\\
&+ \theta_1(\sum_{m=1}^q p_m\eta_m^2(3{\mu_m}^2 L_m^2 d_m^2 + 4\sigma_m^2) + \max_m \eta_m^2 4L^2\tau\sum_{i=1}^{\tau}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2)\\
&\leq -\frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^q p_m\frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \sum_{m=1}^q p_m\eta_m^2(L + 4\theta_1)\sigma_m^2 + 2L\eta_0^2\sigma_0^2
\end{aligned}
$$

$$+\sum_{m=1}^{q} p_m(\frac{\eta_m}{4}+\frac{3L\eta_m^2}{2}+3\theta_1\eta_m^2){\mu_m}^2L_m^2(d_m+3)^3+(\frac{\eta_0}{4}+\frac{3L\eta_0^2}{2}){\mu_m}^2L_0^2(d_0+3)^3$$
$$+\sum_{i=1}^{\tau-1}(\beta_t\tau+\tau\theta_1\max_m\eta_m^2 4L^2+\theta_{i+1}-\theta_i)\mathbb{E}\|\mathbf{w}^{t+1-i}-\mathbf{w}^{t-i}\|^2+(\beta_t\tau+\tau\theta_1\max_m\eta_m^2 4L^2-\theta_\tau)\mathbb{E}\|\mathbf{w}^{t+1-\tau}-\mathbf{w}^{t-\tau}\|^2 \quad (26)$$

If we choose $\eta_0, \eta_m \leq \bar{\eta} \leq \frac{1}{4(L+2\theta_1)}$, then there is $\beta^t \leq \frac{3\bar{\eta}L^2}{2}$. Then for Eq. 26 there is

$$\mathbb{E}(M^{t+1}-M^t)$$
$$\leq -\frac{1}{2}\min\{\eta_0, p_m\eta_m\}\mathbb{E}\|\nabla f_{\mu_m}(w_0,\mathbf{w})\|^2+\sum_{m=1}^{q}p_m\eta_m^2(L+4\theta_1)\sigma_m^2+2L\eta_0^2\sigma_0^2$$
$$+\sum_{m=1}^{q} p_m(\frac{\eta_m}{4}+\frac{3L\eta_m^2}{2}+3\theta_1\eta_m^2){\mu_m}^2L_m^2(d_m+3)^3+(\frac{\eta_0}{4}+\frac{3L\eta_0^2}{2}){\mu_m}^2L_0^2(d_0+3)^3$$
$$-\sum_{i=1}^{\tau-1}(\theta_i-\theta_{i+1}-\frac{3}{2}\bar{\eta}L^2\tau-4\tau\theta_1L^2\bar{\eta}^2)\mathbb{E}\|\mathbf{w}^{t+1-i}-\mathbf{w}^{t-i}\|^2-(\theta_\tau-\frac{3}{2}\bar{\eta}L^2\tau-4\tau\theta_1L^2\bar{\eta}^2)\mathbb{E}\|\mathbf{w}^{t+1-\tau}-\mathbf{w}^{t-\tau}\|^2 \quad (27)$$

Let $\theta_1=\frac{3/2\eta\tau^2L^2}{1-4\tau^2\eta^2L^2}\leq\frac{1}{2}\tau L$ and $\eta_0=\eta_m=\eta\leq\frac{1}{4(\tau+1)L}$ and choose $\theta_2,\cdots,\theta_\tau$ as

$$\theta_{i+1}=\theta_i-\frac{3}{2}\eta L^2\tau-4\tau\theta_1L^2\eta^2, \quad \text{for } i=1,\cdots,\tau-1 \quad (28)$$

Following form Eq. 28 and the definition of θ_1, there is $\theta_\tau=\theta_1-(\tau-1)\frac{3\eta L^2}{2}\tau-4(\tau-1)\tau\theta_1L^2\eta^2\geq 0$. Then Eq. 27 reduces to

$$\mathbb{E}(M^{t+1}-M^t)\leq -\frac{1}{2}\min_m p_m\eta\mathbb{E}\|\nabla f_{\mu_m}(w_0,\mathbf{w})\|^2+2L\eta^2\sigma_0^2$$
$$+\sum_{m=1}^{q}p_m\eta^2(L+2\tau L)\sigma_m^2++\sum_{m=1}^{q}p_m(\frac{\eta}{4}+\frac{3L\eta^2}{2}+\frac{3}{2}\tau L\eta^2){\mu_m}^2L_m^2(d_m+3)^3+(\frac{\eta}{4}+\frac{3L\eta^2}{2}){\mu_m}^2L_0^2(d_0+3)^3 \quad (29)$$

Summing Eq. 29 over $t=0,\cdots,T-1$, there is

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_{\mu_m}(w_0,\mathbf{w})\|^2\leq\frac{f_{\mu_m}^0-f_{\mu_m}^*}{\frac{1}{2}\min_m p_mT\eta}+\frac{\sum_{m=1}^q p_m\eta(L+2\tau L)\sigma_m^2+2L\eta\sigma_0^2}{\frac{1}{2}\min_m p_m}$$
$$+\frac{\sum_{m=1}^q p_m(\frac{1}{4}+\frac{3L\eta+}{2}\frac{3}{2}\tau L\eta){\mu_m}^2L_m^2(d_m+3)^3+(\frac{1}{4}+\frac{3L\eta}{2}){\mu_m}^2L_0^2(d_0+3)^3}{\frac{1}{2}\min_m p_m} \quad (30)$$

According to Lemma 1, there is

$$\mathbb{E}\|\nabla_m f(w_0,\mathbf{w})\|^2\leq 2\mathbb{E}\|\nabla_m f_{\mu_m}(w_0,\mathbf{w})\|^2+\frac{{\mu_m}^2L_m^2(d_m+3)^3}{2}. \quad (31)$$

Thus, there is

$$\mathbb{E}\|\nabla f(w_0,\mathbf{w})\|^2\leq 2\sum_{m=0}^{q}\mathbb{E}\|\nabla_m f_{\mu_m}(w_0,\mathbf{w})\|^2+\sum_{m=0}^{q}\frac{{\mu_m}^2L_m^2(d_m+3)^3}{2}$$
$$\leq 2\mathbb{E}\|\nabla f_{\mu_m}(w_0,\mathbf{w})\|^2+\sum_{m=0}^{q}\frac{{\mu_m}^2L_m^2(d_m+3)^3}{2}. \quad (32)$$

Similarly, according to Lemma 1, there is

$$f(w_0^0,\mathbf{w}^0)-f^*\leq f_{\mu_m}(w_0^0,\mathbf{w}^0)-f_{\mu_m}^*+\sum_{m=0}^{q}\frac{L_md_m{\mu_m}^2}{2} \quad (33)$$

Applying Eqs. 32 and 33 to Eq. 30, there is

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f(w_0,\mathbf{w})\|^2\leq\frac{f^0-f^*}{\frac{1}{4}\min_m p_mT\eta}+\frac{\sum_{m=1}^q p_m\eta(L+2\tau L)\sigma_m^2+2L\eta\sigma_0^2}{\frac{1}{4}\min_m p_m}+\sum_{m=0}^{q}\frac{L_md_m{\mu_m}^2}{2T}+\sum_{m=0}^{q}\frac{{\mu_m}^2L_m^2(d_m+3)^3}{2}$$
$$+\frac{\sum_{m=0}^q p_m(\frac{1}{4}+\frac{3L\eta}{2}+\frac{3}{2}\tau L\eta){\mu_m}^2L_m^2(d_m+3)^3}{\frac{1}{4}\min_m p_m} \quad (34)$$

Let $L_* = \max\{\{L_m\}_{m=0}^q, L\}$, $d_* = \max\{d_m + 3\}_{m=0}^q$, $\sigma_*^2 = \max_m \sigma_m^2$, $\frac{1}{p_*} = \min_m p_m$, then Eq. 34 reduces to

$$\begin{aligned} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f(w_0,\mathbf{w})\|^2 \leq & \frac{4p_*(f^0-f^*)}{T\eta} + 8p_*(L+\tau L)\eta\sigma_*^2 + \frac{(q+1)L_* d_* {\mu_m}^2}{2T} + \frac{(q+1){\mu_m}^2 L_*^2 d_*^3}{2} \\ & + p_*(2+3L_*\eta + \frac{3}{2}\tau L_*\eta){\mu_m}^2 L_*^2 d_*^3 \end{aligned} \tag{35}$$

Choosing $\eta = \min\{\frac{1}{4(\tau+1)L}, \frac{m_0}{\sqrt{T}}\}$ with constant $m_0 > 0$ and $\mu_m = \mathcal{O}(\frac{1}{\sqrt{T}})$ such as $\mu_m = \frac{1}{\sqrt{T}L_* d_*^{3/2}}$, there is

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f(w_0,\mathbf{w})\|^2 \leq \frac{4p_*(f^0-f^*)}{\sqrt{T}m_0} + \frac{8p_* m_0(L+\tau L)\sigma_*^2}{\sqrt{T}} + \frac{(q+1)}{2T^2 L_* d_*^2} + \frac{(q+1)+3p_*}{2T} \tag{36}$$

Thus, since τ is a constant independent to T, we can obtain the final result. This completes the proof. $\square$

LEMMA 3. *Suppose that Assumption 2 holds, then we have*
1) $f_{\mu_m}(w_m)$ *is* L_m*-smooth and* $f_{\mu_m}(w_0,\mathbf{w})$ *is* L*-smooth*

$$\nabla_m f_{\mu_m}(w_0,\mathbf{w}) = \mathbb{E}_u\left[\hat{\nabla}_m f(w_0,\mathbf{w})\right] \tag{37}$$

$$\nabla f_{\mu_m}(w_0,\mathbf{w}) = \mathbb{E}_u\left[\hat{\nabla} f(w_0,\mathbf{w})\right] \tag{38}$$

where $\mathbf{u}$ *is drawn from the uniform distribution over the unit Euclidean sphere, and* $\hat{\nabla}_{w_m} f(w_0,\mathbf{w})$ *is given by Eq. (6).*
2) For any $w_m \in \mathbb{R}^{d_m}$,

$$|f_{\mu_m}(w_m) - f(w_m)| \le \frac{L_m d_m \mu_m^2}{2} \tag{39}$$

$$\|\nabla_m f_{\mu_m}(w_0,\mathbf{w}) - \nabla_m f(w_0,\mathbf{w})\|_2^2 \le \frac{{\mu_m}^2 L_m^2 d_m^2}{4}, \tag{40}$$

LEMMA 4. *Under Assumptions 1 to 4, for* $m = 0, 1, \cdots, q$ *there is*

$$\begin{aligned}\mathbb{E}\|\widehat{v}_{m_t}^t - v_{m_t}^t\|^2 &= \mathbb{E}\|\widehat{v}_{m_t}^t - \bar{v}_{m_t}^t + \bar{v}_{m_t}^t - v_{m_t}^t\|^2 \\ &\le 2\mathbb{E}\|\widehat{v}_{m_t}^t - \bar{v}_{m_t}^t\| + 2\mathbb{E}\|\bar{v}_{m_t}^t - v_{m_t}^t\|^2 \\ &\le \frac{{\mu_m}^2 L_{m_t}^2 d_{m_t}^2}{2} + 2L^2\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2\end{aligned} \tag{41}$$

and

$$\begin{aligned}\mathbb{E}\|\widehat{v}_{m_t}^t\|^2 &= \mathbb{E}\|\widehat{v}_{m_t}^t - v_{m_t}^t + v_{m_t}^t\|^2 \\ &\le 2\mathbb{E}\|\widehat{v}_{m_t}^t - v_{m_t}^t\| + 2\mathbb{E}\|v_{m_t}^t\|^2 \\ &\le {\mu_m}^2 L_{m_t}^2 d_{m_t}^2 + 4L^2\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 + 2\mathbb{E}\|v_{m_t}^t\|^2 \\ &\le 3{\mu_m}^2 L_{m_t}^2 d_{m_t}^2 + 4L^2\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 + 4\sigma_{m_t}^2\end{aligned} \tag{42}$$

PROOF. Under Assumption 2, and taking expectation w.r.t. the sample index i_t, we have

$$\begin{aligned}&\mathbb{E}f_{\mu_m}(w_0^{t+1},\mathbf{w}^{t+1}) \\ &= \mathbb{E}\left(f_{\mu_m}(w_0^t - \eta_0\widehat{v}_0^t, \cdots, w_{m_t}^t - \eta_{m_t}\widehat{v}_{m_t}^t, \cdots)\right) \\ &\le \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \eta_0\left\langle V_0^t, \widehat{v}_0^t\right\rangle - \eta_{m_t}\left\langle V_{m_t}^t, \widehat{v}_{m_t}^t\right\rangle + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2\right) \\ &= \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \eta_0\left\langle V_0^t, \widehat{v}_0^t - v_0^t + v_0^t\right\rangle - \eta_{m_t}\left\langle V_m^t, \widehat{v}_{m_t}^t - v_{m_t}^t + v_{m_t}^t\right\rangle + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2\right) \\ &\le \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \eta_0\|V_0^t\|^2 - \eta_0\left\langle V_0^t, \widehat{v}_0^t - v_0^t\right\rangle - \eta_{m_t}\|V_{m_t}^t\|^2 - \eta_{m_t}\left\langle V_{m_t}^t, \widehat{v}_{m_t}^t - v_{m_t}^t\right\rangle + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2\right) \\ &\le \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \frac{\eta_0}{2}\|V_0^t\|^2 + \frac{\eta_0}{2}\|\widehat{v}_0^t - v_0^t\|^2 - \frac{\eta_{m_t}}{2}\|V_{m_t}^t\|^2 + \frac{\eta_{m_t}}{2}\|\widehat{v}_{m_t}^t - v_{m_t}^t\| + \frac{L\eta_0^2}{2}\|\widehat{v}_0^t\|^2 + \frac{L\eta_{m_t}^2}{2}\|\widehat{v}_{m_t}^t\|^2\right) \\ &\le \mathbb{E}\left(f_{\mu_m}(w_0^t,\mathbf{w}^t) - \frac{\eta_0}{2}\|V_0^t\|^2 - \frac{\eta_{m_t}}{2}\|V_{m_t}^t\|^2 + (\eta_0 + \eta_{m_t} + 2L\eta_0^2 + 2L\eta_{m_t}^2)L^2\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2\right) \\ &+ \left(\frac{\eta_{m_t}}{4} + \frac{3L\eta_{m_t}^2}{2}\right){\mu_m}^2 L_{m_t}^2 d_{m_t}^2 + \left(\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}\right){\mu_m}^2 L_0^2 d_0^2 + 2L\eta_{m_t}^2\sigma_{m_t}^2 + 2L\eta_0^2\sigma_0^2\end{aligned} \tag{43}$$

Taking expectation w.r.t. m_t, and using Assumption 3, there is

$$\begin{aligned}&\mathbb{E}f_{\mu_m}(w_0^{t+1},\mathbf{w}^{t+1}) \\ &\le \mathbb{E}f_{\mu_m}(w_0^t,\mathbf{w}^t) - \frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^{q} p_m\frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \underbrace{(\eta_0 + 2L\eta_0^2 + \max_m(2L\eta_m^2 + \eta_m))L^2}_{\beta^t}\mathbb{E}\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 \\ &+ \sum_{m=1}^{q} p_m\left(\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2}\right){\mu_m}^2 L_m^2 d_{m_t}^2 + \left(\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}\right){\mu_m}^2 L_0^2 d_0^2 + \sum_{m=1}^{q} p_m 2L\eta_m^2\sigma_m^2 + 2L\eta_0^2\sigma_0^2\end{aligned} \tag{44}$$

According to Assumption 4, there is

$$\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 = \|\sum_{i\in D(t)} \mathbf{w}^{i+1} - \mathbf{w}^i\|^2 \le \tau\sum_{i=1}^{\tau}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 \tag{45}$$

We than bound the term $\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2$. First, for $\mathbb{E}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2$

$$\mathbb{E}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 = \mathbb{E}\eta_{m_t}^2 \|\widehat{v}_{m_t}^2\| \leq \sum_{m=1}^{q} p_m \eta_m^2 (3{\mu_m}^2 L_m^2 d_{m_t}^2 + 4\sigma_m^2) + \max_m \eta_m^2 4L^2 \|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 \tag{46}$$

Define a Lyapunov function as

$$M^t = f_{\mu_m}(w_0^t, \mathbf{w}^t) + \sum_{i=1}^{\tau} \theta_i \|\mathbf{w}^{i+1} - \mathbf{w}^i\|^2 \tag{47}$$

Following Lemma 3 and Eq. 47, there is

$$\begin{aligned}
&\mathbb{E}(M^{t+1} - M^t) \\
&= \mathbb{E}\left(f_{\mu_m}(w_0^{t+1}, \mathbf{w}^{t+1}) + \sum_{i=1}^{\tau} \theta_i \|\mathbf{w}^{t+1+1-i} - \mathbf{w}^{t+1-i}\|^2 - f_{\mu_m}(w_0^t, \mathbf{w}^t) - \sum_{i=1}^{\tau} \theta_i \|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 \right) \\
&= -\frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^{q} p_m \frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \beta^t \mathbb{E}\|\widehat{\mathbf{w}}^t - \mathbf{w}^t\|^2 + \sum_{m=1}^{q} p_m 2L\eta_m^2\sigma_m^2 + 2L\eta_0^2\sigma_0^2 \\
&+ \sum_{m=1}^{q} p_m (\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2}){\mu_m}^2 L_m^2 d_m^2 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2 d_0^2 \\
&+ \theta_1 \mathbb{E}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 + \sum_{i=1}^{\tau-1} (\theta_{i+1} - \theta_i)\mathbb{E}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 - \theta_\tau \mathbb{E}\|\mathbf{w}^{t+1-\tau} - \mathbf{w}^{t-\tau}\|^2 \\
&\leq -\frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^{q} p_m \frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \beta^t \tau \sum_{i=1}^{\tau} \|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 + \sum_{m=1}^{q} p_m 2L\eta_m^2\sigma_m^2 + 2L\eta_0^2\sigma_0^2 \\
&+ \sum_{m=1}^{q} p_m (\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2}){\mu_m}^2 L_m^2 d_m^2 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2 d_0^2 \\
&+ \sum_{i=1}^{\tau-1} (\theta_{i+1} - \theta_i)\mathbb{E}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 - \theta_\tau \mathbb{E}\|\mathbf{w}^{t+1-\tau} - \mathbf{w}^{t-\tau}\|^2 \\
&+ \theta_1 (\sum_{m=1}^{q} p_m \eta_m^2 (3{\mu_m}^2 L_m^2 d_m^2 + 4\sigma_m^2) + \max_m \eta_m^2 4L^2 \tau \sum_{i=1}^{\tau} \|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2) \\
&\leq -\frac{\eta_0}{2}\mathbb{E}\|V_0^t\|^2 - \sum_{m=1}^{q} p_m \frac{\eta_m}{2}\mathbb{E}\|V_m^t\|^2 + \sum_{m=1}^{q} p_m \eta_m^2 (L + 4\theta_1)\sigma_m^2 + 2L\eta_0^2\sigma_0^2 \\
&+ \sum_{m=1}^{q} p_m (\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2} + 3\theta_1\eta_m^2){\mu_m}^2 L_m^2 d_m^2 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2 d_0^2 \\
&+ \sum_{i=1}^{\tau-1} (\beta_t\tau + \tau\theta_1 \max_m \eta_m^2 4L^2 + \theta_{i+1} - \theta_i)\mathbb{E}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 + (\beta_t\tau + \tau\theta_1 \max_m \eta_m^2 4L^2 - \theta_\tau)\mathbb{E}\|\mathbf{w}^{t+1-\tau} - \mathbf{w}^{t-\tau}\|^2
\end{aligned} \tag{48}$$

If we choose $\eta_0, \eta_m \leq \bar{\eta} \leq \frac{1}{4(L+2\theta_1)}$, then there is $\beta^t \leq \frac{3\bar{\eta}L^2}{2}$. Then for Eq. 48 there is

$$\begin{aligned}
&\mathbb{E}(M^{t+1} - M^t) \\
&\leq -\frac{1}{2}\min\{\eta_0, p_m\eta_m\}\mathbb{E}\|\nabla f_{\mu_m}(w_0, \mathbf{w})\|^2 + \sum_{m=1}^{q} p_m \eta_m^2 (L + 4\theta_1)\sigma_m^2 + 2L\eta_0^2\sigma_0^2 \\
&+ \sum_{m=1}^{q} p_m (\frac{\eta_m}{4} + \frac{3L\eta_m^2}{2} + 3\theta_1\eta_m^2){\mu_m}^2 L_m^2 d_m^2 + (\frac{\eta_0}{4} + \frac{3L\eta_0^2}{2}){\mu_m}^2 L_0^2 d_0^2 \\
&- \sum_{i=1}^{\tau-1} (\theta_i - \theta_{i+1} - \frac{3}{2}\bar{\eta}L^2\tau - 4\tau\theta_1 L^2\bar{\eta}^2)\mathbb{E}\|\mathbf{w}^{t+1-i} - \mathbf{w}^{t-i}\|^2 - (\theta_\tau - \frac{3}{2}\bar{\eta}L^2\tau - 4\tau\theta_1 L^2\bar{\eta}^2)\mathbb{E}\|\mathbf{w}^{t+1-\tau} - \mathbf{w}^{t-\tau}\|^2
\end{aligned} \tag{49}$$

Let $\theta_1 = \frac{3/2\eta\tau^2 L^2}{1-4\tau^2\eta^2 L^2} \leq \frac{1}{2}\tau L$ and $\eta_0 = \eta_m = \eta \leq \frac{1}{4(\tau+1)L}$ and choose $\theta_2, \cdots, \theta_\tau$ as

$$\theta_{i+1} = \theta_i - \frac{3}{2}\eta L^2\tau - 4\tau\theta_1 L^2\eta^2, \quad \text{for } i = 1, \cdots, \tau - 1 \tag{50}$$

Following form Eq. 50 and the definition of θ_1, there is $\theta_\tau = \theta_1 - (\tau-1)\frac{3\eta L^2}{2}\tau - 4(\tau-1)\tau\theta_1 L^2\eta^2 \geq 0$. Then Eq. 49 reduces to

$$\mathbb{E}(M^{t+1} - M^t) \leq -\frac{1}{2}\min_m p_m \eta \mathbb{E}\|\nabla f_{\mu_m}(w_0, \mathbf{w})\|^2 + 2L\eta^2\sigma_0^2$$
$$+ \sum_{m=1}^{q} p_m \eta^2 (L + 2\tau L)\sigma_m^2 + \sum_{m=1}^{q} p_m (\frac{\eta}{4} + \frac{3L\eta^2}{2} + \frac{3}{2}\tau L\eta^2){\mu_m}^2 L_m^2 d_m^2 + (\frac{\eta}{4} + \frac{3L\eta^2}{2}){\mu_m}^2 L_0^2 d_0^2 \tag{51}$$

Summing Eq. over $t = 0, \cdots, T-1$, there is

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_{\mu_m}(w_0, \mathbf{w})\|^2 \leq \frac{f_{\mu_m}^0 - f_{\mu_m}^*}{\frac{1}{2}\min_m p_m T\eta} + \frac{\sum_{m=1}^{q} p_m \eta (L + 2\tau L)\sigma_m^2 + 2L\eta\sigma_0^2}{\frac{1}{2}\min_m p_m}$$
$$+ \frac{\sum_{m=1}^{q} p_m (\frac{1}{4} + \frac{3L\eta +}{2}\frac{3}{2}\tau L\eta){\mu_m}^2 L_m^2 d_m^2 + (\frac{1}{4} + \frac{3L\eta}{2}){\mu_m}^2 L_0^2 d_0^2}{\frac{1}{2}\min_m p_m} \tag{52}$$

According to Lemma 3, there is

$$\mathbb{E}\|\nabla_m f(w_0, \mathbf{w})\|^2 \leq 2\mathbb{E}\|\nabla_m f_{\mu_m}(w_0, \mathbf{w})\|^2 + \frac{{\mu_m}^2 L_m^2 d_m^2}{2}. \tag{53}$$

Thus, there is

$$\mathbb{E}\|\nabla f(w_0, \mathbf{w})\|^2 \leq 2\sum_{m=0}^{q}\mathbb{E}\|\nabla_m f_{\mu_m}(w_0, \mathbf{w})\|^2 + \sum_{m=0}^{q}\frac{{\mu_m}^2 L_m^2 d_m^2}{2}$$
$$\leq 2\mathbb{E}\|\nabla f_{\mu_m}(w_0, \mathbf{w})\|^2 + \sum_{m=0}^{q}\frac{{\mu_m}^2 L_m^2 d_m^2}{2}. \tag{54}$$

Similarly, according to Lemma 3, there is

$$f(w_0^0, \mathbf{w}^0) - f^* \leq f_{\mu_m}(w_0^0, \mathbf{w}^0) - f_{\mu_m}^* + \sum_{m=0}^{q}\frac{L_m {\mu_m}^2}{2} \tag{55}$$

Applying Eqs. 54 and 55 to Eq. 52, there is

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f(w_0, \mathbf{w})\|^2 \leq \frac{f^0 - f^*}{\frac{1}{4}\min_m p_m T\eta} + \frac{\sum_{m=1}^{q} p_m \eta (L + 2\tau L)\sigma_m^2 + 2L\eta\sigma_0^2}{\frac{1}{4}\min_m p_m} + \sum_{m=0}^{q}\frac{L_m d_m {\mu_m}^2}{2T} + \sum_{m=0}^{q}\frac{{\mu_m}^2 L_m^2 d_m^2}{2}$$
$$+ \frac{\sum_{m=0}^{q} p_m (\frac{1}{4} + \frac{3L\eta}{2} + \frac{3}{2}\tau L\eta){\mu_m}^2 L_m^2 d_m^2}{\frac{1}{4}\min_m p_m} \tag{56}$$

Let $L_* = \max\{\{L_m\}_{m=0}^q, L\}$, $d_* = \max\{d_m\}_{m=0}^q$, $\sigma_*^2 = \max_m \sigma_m^2$, $\frac{1}{p_*} = \min_m p_m$, then Eq. 56 reduces to

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f(w_0, \mathbf{w})\|^2 \leq \frac{4p_*(f^0 - f^*)}{T\eta} + 8p_*(L + \tau L)\eta\sigma_*^2 + \frac{(q+1)L_*{\mu_m}^2}{2T} + \frac{(q+1){\mu_m}^2 L_*^2 d_*^2}{2}$$
$$+ p_*(2 + 3L_*\eta + \frac{3}{2}\tau L_*\eta){\mu_m}^2 L_*^2 d_*^2 \tag{57}$$

Choosing $\eta = \min\{\frac{1}{4(\tau+1)L}, \frac{m_0}{\sqrt{T}}\}$ with constant $m_0 > 0$ and $\mu_m = \mathcal{O}(\frac{1}{\sqrt{T}})$ such as $\mu_m = \frac{1}{\sqrt{T}L_* d_*}$, there is

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f(w_0, \mathbf{w})\|^2 \leq \frac{4p_*(f^0 - f^*)}{\sqrt{T}m_0} + \frac{8p_* m_0 (L + \tau L)\sigma_*^2}{\sqrt{T}} + \frac{(q+1)}{2T^2 L_* d_*^2} + \frac{(q+1) + 3p_*}{2T} \tag{58}$$

Thus, since τ is a constant independent to T, we can obtain the final result. This completes the proof. $\square$