

PROOF OF LEMMA 7

The vector $\bar{\mathbf{w}}^{t+1}$ can be rewritten as in the following.

$$\begin{aligned}\bar{\mathbf{w}}^{t+1} &= \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} \mathbf{w}^{L,m,u,t}, \\ &= \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} (\mathbf{w}^{L-1,m,u,t} - \eta_t \tilde{\nabla} f_{m,u}(\mathbf{w}^{L-1,m,u,t})),\end{aligned}\tag{37a}$$

$$= \mathbf{w}^t - \eta_t \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} \sum_{l=0}^{L-1} \tilde{\nabla} f_{m,u}(\mathbf{w}^{l,m,u,t}) \right],\tag{37b}$$

$$\begin{aligned}&= \mathbf{w}^t - \eta_t \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} \sum_{l=0}^{L-1} (\tilde{\nabla} f_{m,u}(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{b}^{l,t}) + \nabla f(\mathbf{b}^{l,t})) \right], \\ &= \mathbf{w}^t - \eta_t [\mathbf{e}^{0,t} + \nabla f(\mathbf{b}^{0,t}) + \dots + \mathbf{e}^{L-1,t} + \nabla f(\mathbf{b}^{L-1,t})], \\ &= \mathbf{w}^t - \eta_t (\mathbf{g}^t + \mathbf{e}^t).\end{aligned}\tag{37c}$$

where (37a) comes from Step 8 of Algorithm 1 and (37b) is due to the fact that

$$\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} = 1.$$

From (37c), the following holds

$$\begin{aligned}\mathbb{E}[\|\bar{\mathbf{w}}^{t+1} - \mathbf{w}^*\|_2^2] &= \mathbb{E}[\|\mathbf{w}^t - \eta_t \mathbf{g}^t - \mathbf{w}^* - \eta_t \mathbf{e}^t\|_2^2], \\ &= \mathbb{E}[\|\mathbf{w}^t - \eta_t \mathbf{g}^t - \mathbf{w}^*\|_2^2] - 2\eta_t \mathbb{E}[\langle \mathbf{w}^t - \eta_t \mathbf{g}^t - \mathbf{w}^*, \mathbf{e}^t \rangle] + \eta_t^2 \mathbb{E}[\|\mathbf{e}^t\|_2^2], \\ &\leq \mathbb{E}[\|\mathbf{w}^t - \eta_t \mathbf{g}^t - \mathbf{w}^*\|_2^2] + K\eta_t^2 \mathbb{E}[\|\mathbf{w}^t - \eta_t \mathbf{g}^t - \mathbf{w}^*\|_2^2] + \frac{1}{K} \mathbb{E}[\|\mathbf{e}^t\|_2^2] + \eta_t^2 \mathbb{E}[\|\mathbf{e}^t\|_2^2], \tag{38a} \\ &= (1 + K\eta_t^2) \mathbb{E}[\|\mathbf{w}^t - \eta_t \mathbf{g}^t - \mathbf{w}^*\|_2^2] + \left(\frac{1}{K} + \eta_t^2\right) \mathbb{E}[\|\mathbf{e}^t\|_2^2], \\ &= (1 + K\eta_t^2) \mathbb{E}[\|\mathbf{b}^{L,t} - \mathbf{w}^*\|_2^2] + \left(\frac{1}{K} + \eta_t^2\right) \mathbb{E}[\|\mathbf{e}^t\|_2^2],\end{aligned}\tag{38b}$$

where (38a) is due to the inequality $-2\langle \mathbf{x}, \mathbf{y} \rangle \leq \alpha \|\mathbf{x}\|_2^2 + \frac{1}{\alpha} \|\mathbf{y}\|_2^2$ for $\alpha > 0$ and (38b) follows from the definition of $\{\mathbf{b}^{l,t}\}_{l=0}^L$ in (16).

Next, applying the Cauchy-Schwarz inequality yields

$$\mathbb{E}[\|\mathbf{e}^t\|_2^2] \leq L \sum_{l=0}^{L-1} \mathbb{E}[\|\mathbf{e}^{l,t}\|_2^2] = L\mathbb{E}[\|\mathbf{e}^{0,t}\|_2^2] + L \sum_{l=1}^{L-1} \mathbb{E}[\|\mathbf{e}^{l,t}\|_2^2]. \quad (39)$$

Regarding the first term on the r.h.s. of (39), we have

$$\begin{aligned} \mathbb{E}[\|\mathbf{e}^{0,t}\|_2^2] &= \mathbb{E} \left[\left\| \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} [\tilde{\nabla} f_{m,u}(\mathbf{w}^{0,m,u,t}) - \nabla f(\mathbf{b}^{0,t})] \right\|_2^2 \right], \\ &= \mathbb{E} \left[\left\| \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} [\tilde{\nabla} f_{m,u}(\mathbf{w}^t) - \nabla f(\mathbf{w}^t)] \right\|_2^2 \right], \\ &= \mathbb{E} \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u}^2 \|\tilde{\nabla} f_{m,u}(\mathbf{w}^t) - \nabla f(\mathbf{w}^t)\|_2^2 \right], \end{aligned} \quad (40a)$$

$$\leq \zeta \mathbb{E}_{\{\mathcal{C}_m\}_{m=1}^M} \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u}^2 \right], \quad (40b)$$

$$\leq \zeta, \quad (40c)$$

where (40a) is due to the unbiasedness in Assumption 3 and the fact that $\{\tilde{\nabla} f_{m,u}(\mathbf{w}^t)\}$ are independent, (40b) follows from the boundedness Assumption 3, and (40c) is due to the fact that $\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u}^2 \leq 1$. Now, the term $\mathbb{E}[\|\mathbf{e}^{l,t}\|_2^2]$ in (39) is rewritten by

$$\begin{aligned} \mathbb{E}[\|\mathbf{e}^{l,t}\|_2^2] &= \mathbb{E} \left[\left\| \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} [\tilde{\nabla} f_{m,u}(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{b}^{l,t})] \right\|_2^2 \right], \\ &= \mathbb{E} \left[\left\| \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} [\tilde{\nabla} f_{m,u}(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{w}^{l,m,u,t}) + \nabla f(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{b}^{l,t})] \right\|_2^2 \right], \\ &= \mathbb{E} \left[\left\| \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} [\tilde{\nabla} f_{m,u}(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{w}^{l,m,u,t})] \right\|_2^2 \right] \\ &\quad + \mathbb{E} \left[\left\| \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} [\nabla f(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{b}^{l,t})] \right\|_2^2 \right], \end{aligned} \quad (41a)$$

where (41a) follows from Assumption 3 that $\mathbb{E}[\tilde{\nabla} f_{m,u}(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{w}^{l,m,u,t})] = \mathbf{0}$. Thus, the

following holds

$$\mathbb{E}[\|\mathbf{e}^{l,t}\|_2^2] \leq \zeta + \mathbb{E} \left[\left(\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u}^2 \right) \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \|\nabla f(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{b}^{l,t})\|_2^2 \right], \quad (42a)$$

$$\leq \zeta + \mathbb{E} \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \|\nabla f(\mathbf{w}^{l,m,u,t}) - \nabla f(\mathbf{b}^{l,t})\|_2^2 \right], \quad (42b)$$

$$\leq \zeta + \mathbb{E}_{\{\mathcal{C}_m\}_{m=1}^M} \left[\mathbb{E} \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \gamma^2 \|\mathbf{w}^{l,m,u,t} - \mathbf{b}^{l,t}\|_2^2 \middle| \{\mathcal{C}_m\}_{m=1}^M \right] \right], \quad (42c)$$

$$= \zeta + \mathbb{E}_{\{\mathcal{C}_m\}_{m=1}^M} \left[\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{G}_m} \mathbb{1}_{\mathcal{C}_m}(d_{m,u}) \gamma^2 \mathbb{E} \left[\|\mathbf{w}^{l,m,u,t} - \mathbf{b}^{l,t}\|_2^2 \right] \right], \quad (42d)$$

$$= \zeta + \sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{G}_m} \frac{c_m}{g_m} \gamma^2 \mathbb{E} \left[\|\mathbf{w}^{l,m,u,t} - \mathbf{b}^{l,t}\|_2^2 \right], \quad (42e)$$

$$= \zeta + K\gamma^2 a^{l,t}, \quad (42f)$$

where (42a) follows from $\mathbb{E}[\|\tilde{\nabla} f_{m,u}(\mathbf{x}) - \nabla f(\mathbf{x})\|_2^2] \leq \zeta$ and the Cauchy-Schwarz inequality, (42b) is due to the fact that $\sum_{m=1}^M \sum_{d_{m,u} \in \mathcal{C}_m} \omega_{m,u} = 1$, (42c) follows from Assumption 1, (42d) is due to the fact that $\Pr[d_{m,u} \in \mathcal{C}_m] = \frac{1}{g_m}$, (42e) follows from $\frac{c_m}{g_m} \leq 1$, and (42f) is due to the definition of $a^{l,t}$ in (17).

From (18b), plugging (40c) and (42f) into (39) leads to

$$\begin{aligned} \mathbb{E}[\|\mathbf{e}^t\|_2^2] &\leq L\zeta + \sum_{l=1}^{L-1} L \left(\zeta + K\gamma^2 a^{l,t} \right), \\ &= L^2\zeta + LK\gamma^2 \sum_{l=1}^{L-1} a^{l,t}, \\ &\leq L^2\zeta + LK\gamma^2 (L-1)\eta_t^2 L\zeta (1 + L\eta_t^2 \gamma^2)^L, \\ &\leq L^2\zeta + KL^3\gamma^2 \zeta \eta_t^2 e^{L^2\eta_t^2 \gamma^2}, \end{aligned} \quad (43a)$$

$$\leq L^2\zeta + KL^3\gamma^2 \zeta \eta_t^2 e, \quad (43b)$$

where (43a) is due to the fact that $(1+x) \leq e^x$ for $x \geq 0$, and (43b) follows from $\eta_t \leq \frac{1}{L\gamma}$.

Plugging (43b) into (38b) and using (18a) give

$$\mathbb{E}[\|\bar{\mathbf{w}}^{t+1} - \mathbf{w}^*\|_2^2] \leq (1 + K\eta_t^2)(1 - \mu\eta_t)^L \mathbb{E}[\|\mathbf{w}^t - \mathbf{w}^*\|_2^2] + \left(\frac{1}{K} + \eta_t^2 \right) \left(L^2\zeta + KL^3\gamma^2 \zeta \eta_t^2 e \right),$$

which completes the proof.