SUPPLEMENTARY MATERIAL FOR *Variance-Reduced Stochastic Quasi-Newton Methods for Decentralized Learning*

I. PROOF OF LEMMA 5

Proof. According to the update for $\mathbf{x}^{k+1}$ in (10), we have

$$\begin{aligned}&\|\mathbf{x}^{k+1}-\mathbf{W}_\infty\mathbf{x}^{k+1}\|^2 \qquad (53)\\ =&\|\mathbf{W}(\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k)-\alpha(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|^2\\ \le&(1+\vartheta)\|\mathbf{W}(\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k)\|^2+\Big(1+\frac{1}{\vartheta}\Big)\alpha^2\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|^2\\ \le&\frac{1+\sigma^2}{2}\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2+\frac{2\alpha^2}{1-\sigma^2}\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|^2,\end{aligned}$$

where the equality is due to $\mathbf{W}_\infty=\mathbf{W}\mathbf{W}_\infty$, the first inequality holds for any $\vartheta>0$ due to Young's inequality, and the second inequality holds by setting $\vartheta=\frac{1-\sigma^2}{2\sigma^2}$ and noting that $\|\mathbf{W}(\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k)\|^2=\|\mathbf{W}\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2\le\sigma^2\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2$ by (30) and $\sigma<1$ by Assumption 3.

Next, we bound the term $\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|$ in (53) via

$$\begin{aligned}&\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|\\ =&\|(I_{nd}-\mathbf{W}_\infty)(\mathbf{H}^k-\bar{M}I_{nd})\mathbf{g}^k+\bar{M}(I_{nd}-\mathbf{W}_\infty)\mathbf{g}^k\|\\ \le&\frac{M_2-M_1}{2}\|\mathbf{g}^k\|+\bar{M}\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|,\end{aligned}$$

where $\bar{M}:=\frac{M_1+M_2}{2}$ and the inequality holds because of the triangle inequality, $\|I_{nd}-\mathbf{W}_\infty\|_2\le1$, and $M_1I_{nd}\preceq\mathbf{H}^k\preceq M_2I_{nd}$. Further, we have

$$\begin{aligned}\|\mathbf{g}^k\|\le&\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|+\sqrt{n}\|\overline{\mathbf{g}}^k\|\\ \le&\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|+\sqrt{n}(\|\overline{\mathbf{g}}^k-\nabla F(\overline{\mathbf{x}}^k)\|+\|\nabla F(\overline{\mathbf{x}}^k)\|),\end{aligned}$$

which implies that

$$\begin{aligned}\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|\le&M_2\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|\\ &+\frac{M_2\gamma}{2}\sqrt{n}(\|\overline{\mathbf{g}}^k-\nabla F(\overline{\mathbf{x}}^k)\|+\|\nabla F(\overline{\mathbf{x}}^k)\|)\end{aligned}$$

with $\gamma=1-M_1/M_2$. By squaring both sides of the above inequality and using the fact that $(a+b)^2\le2(a^2+b^2)$, which holds for any $a,b\in\mathbb{R}$, we get

$$\begin{aligned}\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|^2\le&2M_2^2\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|^2 \qquad (54)\\ &+M_2^2\gamma^2n(\|\overline{\mathbf{g}}^k-\nabla F(\overline{\mathbf{x}}^k)\|^2+\|\nabla F(\overline{\mathbf{x}}^k)\|^2).\end{aligned}$$

To handle the second term in (54), we expand $\mathbb{E}\left[\|\overline{\mathbf{g}}^k-\nabla F(\overline{\mathbf{x}}^k)\|^2|\mathcal{F}^k\right]$ as

$$\begin{aligned}&\mathbb{E}\left[\|\overline{\mathbf{g}}^k-\nabla F(\overline{\mathbf{x}}^k)\|^2|\mathcal{F}^k\right] \qquad (55)\\ =&\mathbb{E}\left[\|\overline{\mathbf{v}}^k-\overline{\nabla f}(\mathbf{x}^k)+\overline{\nabla f}(\mathbf{x}^k)-\nabla F(\overline{\mathbf{x}}^k)\|^2|\mathcal{F}^k\right]\\ =&\mathbb{E}\left[\|\overline{\mathbf{v}}^k-\overline{\nabla f}(\mathbf{x}^k)\|^2+\|\overline{\nabla f}(\mathbf{x}^k)-\nabla F(\overline{\mathbf{x}}^k)\|^2|\mathcal{F}^k\right]\\ \le&\mathbb{E}\left[\frac{1}{n^2}\|\mathbf{v}^k-\nabla f(\mathbf{x}^k)\|^2+\frac{L^2}{n}\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2|\mathcal{F}^k\right],\end{aligned}$$

where we use $\overline{\mathbf{g}}^k=\overline{\mathbf{v}}^k$ from (31) to get the first equality and $\mathbb{E}\left[\left\langle\overline{\nabla f}\left(\mathbf{x}^k\right)-\nabla F\left(\overline{\mathbf{x}}^k\right),\overline{\mathbf{v}}^k-\overline{\nabla f}(\mathbf{x}^k)\right\rangle|\mathcal{F}^k\right]=0$ from (33) to get the second equality. To get the inequality, in addition to (34), we use the fact that $\{v_i^k-\nabla f_i(x_i^k)\}$ are independent given the event $\mathcal{F}^k$ and thus

$$\begin{aligned}&\mathbb{E}\left[\|\overline{\mathbf{v}}^k-\overline{\nabla f}(\mathbf{x}^k)\|^2|\mathcal{F}^k\right]\\ =&\mathbb{E}\Big[\Big\|\frac{1}{n}\sum_{i=1}^n\left(v_i^k-\nabla f_i(x_i^k)\right)\Big\|^2|\mathcal{F}^k\Big]\\ =&\frac{1}{n^2}\mathbb{E}\left[\|\mathbf{v}^k-\nabla f(\mathbf{x}^k)\|^2|\mathcal{F}^k\right]. \qquad (56)\end{aligned}$$

For the third term in (54), we apply Fact 1 to get

$$\|\nabla F(\overline{\mathbf{x}}^k)\|^2\le2L(F(\overline{\mathbf{x}}^k)-F(x^*)), \qquad (57)$$

where we use $\nabla F(x^*)=0$. Taking conditional expectation on both sides of (54) and using (55) and (57), we get

$$\begin{aligned}&\mathbb{E}\left[\|(I_{nd}-\mathbf{W}_\infty)\mathbf{d}^k\|^2|\mathcal{F}^k\right] \qquad (58)\\ \le&M_2^2\cdot\mathbb{E}\Big[2\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|^2+\gamma^2L^2\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2\\ &+\frac{\gamma^2}{n}\|\mathbf{v}^k-\nabla f(\mathbf{x}^k)\|^2+2\gamma^2Ln\left(F(\overline{\mathbf{x}}^k)-F(x^*)\right)|\mathcal{F}^k\Big].\end{aligned}$$

Finally, upon taking total expectation in (53) and (58) and combining the results, we get (35) and complete the proof. □

II. PROOF OF LEMMA 6

Proof. Defining $v_{i,l}^k:=\left(\nabla f_{i,l}(x_i^k)-\nabla f_{i,l}(\tau_i^k)\right)+\nabla f_i(\tau_i^k)$ and using (8), we have

$$v_i^k=\frac{1}{b_i}\sum_{l\in S_i^k}v_{i,l}^k.$$

Conditioning on $\mathcal{F}^k$, since v_i^k is a random sampling of $\{v_{i,l}^k\}_{l=1}^{m_i}$ with size b_i and without replacement, we have

$$\mathrm{Var}(v_i^k)=\frac{m_i-b_i}{(m_i-1)b_i}\mathrm{Var}(v_{i,l}^k)\le B\cdot\mathrm{Var}(v_{i,l}^k),$$

where we use the definition of B in (23). This gives

$$\mathbb{E}[\|\mathbf{v}^k-\nabla f(\mathbf{x}^k)\|^2|\mathcal{F}^k]=\sum_{i=1}^n\mathrm{Var}(v_i^k)\le B\sum_{i=1}^n\mathrm{Var}(v_{i,l}^k). \qquad (59)$$

By the definition of $v_{i,l}^k$, we have

$$\begin{aligned}&\mathrm{Var}(v_{i,l}^k)\le\frac{1}{m_i}\sum_{l=1}^{m_i}\left\|\nabla f_{i,l}(x_i^k)-\nabla f_{i,l}(\tau_i^k)\right\|^2 \qquad (60)\\ \le&\frac{4}{m_i}\sum_{l=1}^{m_i}\left\|\nabla f_{i,l}(x_i^k)-\nabla f_{i,l}(\overline{\mathbf{x}}^k)\right\|^2+\left\|\nabla f_{i,l}(\overline{\mathbf{x}}^k)-\nabla f_{i,l}(x^*)\right\|^2\\ &+\left\|\nabla f_{i,l}(x^*)-\nabla f_{i,l}(\overline{\tau}^k)\right\|^2+\left\|\nabla f_{i,l}(\overline{\tau}^k)-\nabla f_{i,l}(\tau_i^k)\right\|^2\\ \le&\frac{4}{m_i}\sum_{l=1}^{m_i}L^2\left\|x_i^k-\overline{\mathbf{x}}^k\right\|^2+\left\|\nabla f_{i,l}(\overline{\mathbf{x}}^k)-\nabla f_{i,l}(x^*)\right\|^2\\ &+\left\|\nabla f_{i,l}(x^*)-\nabla f_{i,l}(\overline{\tau}^k)\right\|^2+L^2\left\|\overline{\tau}^k-\tau_i^k\right\|^2\\ =&4L^2\left(\left\|x_i^k-\overline{\mathbf{x}}^k\right\|^2+\left\|\overline{\tau}^k-\tau_i^k\right\|^2\right)+\frac{4}{m_i}\sum_{l=1}^{m_i}\\ &\left(\left\|\nabla f_{i,l}(\overline{\mathbf{x}}^k)-\nabla f_{i,l}(x^*)\right\|^2+\left\|\nabla f_{i,l}(x^*)-\nabla f_{i,l}(\overline{\tau}^k)\right\|^2\right).\end{aligned}$$

Using Fact 1, we have

$$\begin{aligned}&\left\|\nabla f_{i,l}(\overline{\mathbf{x}}^k)-\nabla f_{i,l}(x^*)\right\|^2\\ \le&2L\left(f_{i,l}(\overline{\mathbf{x}}^k)-f_{i,l}(x^*)-\nabla f_{i,l}(x^*)^T(\overline{\mathbf{x}}^k-x^*)\right).\end{aligned}$$

This, together with the fact that $\frac{1}{nm_i}\sum_{i=1}^n\sum_{l=1}^{m_i}\nabla f_{i,l}(x^*)=\nabla F(x^*)=0$, yields

$$\sum_{i=1}^n\frac{1}{m_i}\sum_{l=1}^{m_i}\left\|\nabla f_{i,l}(\overline{\mathbf{x}}^k)-\nabla f_{i,l}(x^*)\right\|^2\le 2nL\left(F(\overline{\mathbf{x}}^k)-F(x^*)\right). \tag{61}$$

In a similar manner, we can bound $\left\|\nabla f_{i,l}(x^*)-\nabla f_{i,l}(\overline{\tau}^k)\right\|^2$. By combining (59)–(61), we have

$$\begin{aligned}&\mathbb{E}[\|\mathbf{v}^k-\nabla f(\mathbf{x}^k)\|^2]\\ \le&\mathbb{E}\Big[4L^2B\left(\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2+\|\tau^k-\mathbf{W}_\infty\tau^k\|^2\right)\\ &+8nLB\left(F(\overline{\mathbf{x}}^k)-F(x^*)\right)+8nLB\left(F(\overline{\tau}^k)-F(x^*)\right)\Big],\end{aligned}$$

which gives (36) and completes the proof. ☐

III. Proof of Corollary 1

Proof. By substituting (36) into (35), we have

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{x}^{k+1}-\mathbf{W}_\infty\mathbf{x}^{k+1}\|^2\right] \qquad (62)\\ \le&\left(\frac{1+\sigma^2}{2}+\frac{2\gamma^2\alpha^2M_2^2L^2}{1-\sigma^2}\left(1+\frac{4B}{n}\right)\right)\mathbb{E}\left[\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2\right]\\ &+\frac{2\alpha^2M_2^2}{1-\sigma^2}\cdot\Bigg(2\mathbb{E}\left[\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|^2\right]\\ &+2L\gamma^2\left(1+\frac{4B}{n}\right)\cdot n\mathbb{E}\left[F(\overline{\mathbf{x}}^k)-F(x^*)\right]\\ &+\frac{4L^2B\gamma^2}{n}\left(\mathbb{E}\left[\|\tau^k-\mathbf{W}_\infty\tau^k\|^2\right]+\frac{2}{L}n\mathbb{E}\left[F(\overline{\tau}^k)-F(x^*)\right]\right)\Bigg).\end{aligned}$$

By our choice of parameters in (24), we have

$$\begin{aligned}\frac{1+\sigma^2}{2}+\frac{2\gamma^2\alpha^2M_2^2L^2}{1-\sigma^2}(1+\frac{4B}{n})&\le 1-\frac{0.99(1-\sigma^2)}{2},\\ \frac{2\alpha^2M_2^2}{1-\sigma^2}&\le\frac{0.01(1-\sigma^2)\alpha\mu M_1}{L^2},\\ 2L\gamma^2(1+\frac{4B}{n})&\le 2.05L\gamma^2.\end{aligned}$$

After substituting the above inequalities into (62), we get

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{x}^{k+1}-\mathbf{W}_\infty\mathbf{x}^{k+1}\|^2\right] \qquad (63)\\ \le&\left(1-\frac{0.99(1-\sigma^2)}{2}\right)\mathbb{E}\left[\|\mathbf{x}^k-\mathbf{W}_\infty\mathbf{x}^k\|^2\right]\\ &+0.011(1-\sigma^2)\gamma^2\alpha\mu M_1\cdot\frac{2n}{L}\mathbb{E}\left[F(\overline{\mathbf{x}}^k)-F(x^*)\right]\\ &+0.02\alpha\mu M_1\cdot\frac{1-\sigma^2}{L^2}\mathbb{E}\left[\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|^2|\mathcal{F}^k\right]\\ &+\frac{0.04(1-\sigma^2)\gamma^2\alpha\mu M_1B}{n}\mathbb{E}\left[\|\tau^k-\mathbf{W}_\infty\tau^k\|^2\right]\\ &+\frac{0.04(1-\sigma^2)\gamma^2\alpha\mu M_1B}{n}\cdot\frac{2n}{L}\mathbb{E}\left[F(\overline{\tau}^k)-F(x^*)\right].\end{aligned}$$

By substituting the parameters in (24) and doing algebraic manipulations, we get (37) and complete the proof. Please see Part I [54] of the full version of this paper for details. ☐

IV. Proof of Lemma 7

Proof. According to Assumption 4, we have

$$\begin{aligned}&\mathbb{E}\left[\|\overline{\mathbf{d}}^k-\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)\|^2\right] \qquad (64)\\ \le&\mathbb{E}\left[2\|\overline{\mathbf{d}}^k-\overline{\mathbf{H}}^k\overline{\mathbf{g}}^k\|^2+2M_2^2\|\overline{\mathbf{g}}^k-\nabla F(\overline{\mathbf{x}}^k)\|^2\right].\end{aligned}$$

For the first term in (64), we compute

$$\begin{aligned}\overline{\mathbf{d}}^k=\frac{1}{n}\sum_{i=1}^nH_i^kg_i^k=&\frac{1}{n}\sum_{i=1}^nH_i^k(g_i^k-\overline{\mathbf{g}}^k)+\overline{\mathbf{H}}^k\overline{\mathbf{g}}^k \qquad (65)\\ =&\frac{1}{n}\sum_{i=1}^n\left(H_i^k-\bar{M}I_d\right)(g_i^k-\overline{\mathbf{g}}^k)+\overline{\mathbf{H}}^k\overline{\mathbf{g}}^k,\end{aligned}$$

where $\bar{M}=\frac{M_1+M_2}{2}$. Thus, we have

$$\begin{aligned}\|\overline{\mathbf{d}}^k-\overline{\mathbf{H}}^k\overline{\mathbf{g}}^k\|^2\le&\frac{1}{n}\sum_{i=1}^n\left\|\left(H_i^k-\bar{M}I_d\right)(g_i^k-\overline{\mathbf{g}}^k)\right\|^2 \qquad (66)\\ \le&\frac{M_2^2\gamma^2}{4n}\|\mathbf{g}^k-\mathbf{W}_\infty\mathbf{g}^k\|^2.\end{aligned}$$

Taking total expectation on both sides of (55) and (66) and substituting the results into the two terms on the right-hand side of (64), we get (38) and complete the proof. ☐

V. Proof of Corollary 2

Proof. Taking the average of the update for $\mathbf{x}^{k+1}$ in (10) over all the nodes, we have

$$\overline{\mathbf{x}}^{k+1}=\overline{\mathbf{x}}^k-\alpha\overline{\mathbf{d}}^k.$$

Then, we can bound the global cost at the average $\overline{\mathbf{x}}^{k+1}$ as

$$\begin{aligned}F(\overline{\mathbf{x}}^{k+1})\le&F(\overline{\mathbf{x}}^k)-\alpha\left\langle\nabla F(\overline{\mathbf{x}}^k),\overline{\mathbf{d}}^k\right\rangle+\frac{L\alpha^2}{2}\|\overline{\mathbf{d}}^k\|^2 \qquad (67)\\ \le&F(\overline{\mathbf{x}}^k)-\alpha\left\langle\nabla F(\overline{\mathbf{x}}^k),\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)\right\rangle\\ &+\alpha\left\langle\nabla F(\overline{\mathbf{x}}^k),\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)-\overline{\mathbf{d}}^k\right\rangle\\ &+\alpha^2L\left(\|\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)\|^2+\|\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)-\overline{\mathbf{d}}^k\|^2\right),\end{aligned}$$

where we use Assumption 1 in the first inequality. For the second term in the last inequality of (67), we have

$$\left\langle\nabla F(\overline{\mathbf{x}}^k),\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)\right\rangle\ge M_1\|\nabla F(\overline{\mathbf{x}}^k)\|^2. \tag{68}$$

In order to bound the third term in the last inequality of (67), we derive

$$\begin{aligned}&\mathbb{E}\left[\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k)-\overline{\mathbf{d}}^k\mid\mathcal{F}^k\right]\\ =&\mathbb{E}\Big[\overline{\mathbf{H}}^k(\nabla F(\overline{\mathbf{x}}^k)-\overline{\mathbf{g}}^k)-\frac{1}{n}\sum_{i=1}^n\left(H_i^k-\bar{M}I_d\right)(g_i^k-\overline{\mathbf{g}}^k)|\mathcal{F}^k\Big]\\ =&\mathbb{E}\Big[\overline{\mathbf{H}}^k\left(\nabla F(\overline{\mathbf{x}}^k)-\overline{\nabla f}(\mathbf{x}^k)\right)+\left(\overline{\mathbf{H}}^k-\bar{M}I_d\right)\left(\overline{\nabla f}(\mathbf{x}^k)-\overline{\mathbf{v}}^k\right)\\ &-\frac{1}{n}\sum_{i=1}^n\left(H_i^k-\bar{M}I_d\right)(g_i^k-\overline{\mathbf{g}}^k)\mid\mathcal{F}^k\Big],\end{aligned}$$

where the first equality holds because of (65) and the second equality holds because of the fact that $\mathbb{E}[\overline{\mathbf{g}}^k|\mathcal{F}^k]=\mathbb{E}[\overline{\mathbf{v}}^k|\mathcal{F}^k]=\overline{\nabla f}(\mathbf{x}^k)$. Then, using the triangle inequality and the fact that $\|H_i^k - \bar{M} I_d\|_2 \le \frac{M_2\gamma}{2}, \forall i$ and $\|\overline{\mathbf{H}}^k - \bar{M} I_d\|_2 \le \frac{M_2\gamma}{2}$, we have

$$\begin{aligned}
&\left\|\mathbb{E}\left[\overline{\mathbf{H}}^k \nabla F(\overline{\mathbf{x}}^k) - \overline{\mathbf{d}}^k \mid \mathcal{F}^k\right]\right\|^2 \\
\le&\mathbb{E}\Big[4M_2^2\left\|\nabla F(\overline{\mathbf{x}}^k) - \overline{\nabla f}(\mathbf{x}^k)\right\|^2 + \frac{M_2^2\gamma^2}{2n}\|\mathbf{g}^k - \mathbf{W}_\infty \mathbf{g}^k\|^2 \\
&+ M_2^2\gamma^2\|\overline{\nabla f}(\mathbf{x}^k) - \overline{\mathbf{v}}^k\|^2 | \mathcal{F}^k\Big] \\
\le&\frac{M_2^2}{n}\cdot\mathbb{E}\Big[4L^2\|\mathbf{x}^k - \mathbf{W}_\infty\mathbf{x}^k\|^2 + \frac{\gamma^2}{2}\|\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k\|^2 \\
&+ \frac{\gamma^2}{n}\|\mathbf{v}^k - \nabla f(\mathbf{x}^k)\|^2 | \mathcal{F}^k\Big],
\end{aligned}$$

where the last inequality holds because of (34) and (56). It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\mathbb{E}\left[\left\langle \nabla F(\overline{\mathbf{x}}^k), \left(\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k) - \overline{\mathbf{d}}^k\right)\right\rangle \mid \mathcal{F}^k\right] \qquad (69)\\
\le&\frac{M_1}{2}\|\nabla F(\overline{\mathbf{x}}^k)\|^2 + \frac{1}{2M_1}\left\|\mathbb{E}\left[\overline{\mathbf{H}}^k\nabla F(\overline{\mathbf{x}}^k) - \overline{\mathbf{d}}^k \mid \mathcal{F}^k\right]\right\|^2 \\
\le&\frac{M_1}{2}\|\nabla F(\overline{\mathbf{x}}^k)\|^2 + \frac{M_2^2}{2nM_1}\cdot\mathbb{E}\Big[4L^2\|\mathbf{x}^k - \mathbf{W}_\infty\mathbf{x}^k\|^2 \\
&+ \frac{\gamma^2}{2}\|\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k\|^2 + \frac{\gamma^2}{n}\|\mathbf{v}^k - \nabla f(\mathbf{x}^k)\|^2 | \mathcal{F}^k\Big].
\end{aligned}$$

By taking total expectation on both sides of (67) and using (38) (68), and (69), we have

$$\begin{aligned}
&\mathbb{E}\left[F(\overline{\mathbf{x}}^{k+1})\right] \qquad (70)\\
\le&\mathbb{E}\left[F(\overline{\mathbf{x}}^k)\right] - \left(\frac{\alpha M_1}{2} - \alpha^2 L M_2^2\right)\mathbb{E}\left[\|\nabla F(\overline{\mathbf{x}}^k)\|^2\right] \\
&+ \left(\frac{2\alpha M_2^2}{nM_1} + \frac{2M_2^2\alpha^2 L}{n}\right)L^2\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{W}_\infty\mathbf{x}^k\|^2\right] \\
&+ \left(\frac{\alpha M_2^2\gamma^2}{2nM_1} + \frac{2M_2^2\alpha^2 L}{n}\right)\frac{1}{n}\mathbb{E}\left[\|\mathbf{v}^k - \nabla f(\mathbf{x}^k)\|^2\right] \\
&+ \left(\frac{\alpha M_2^2\gamma^2}{4nM_1} + \frac{M_2^2\alpha^2\gamma^2 L}{2n}\right)\mathbb{E}\left[\|\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k\|^2\right].
\end{aligned}$$

Next, we bound the four coefficients on the right-hand side of (70). By our choice of parameters in (24), we have

$$\begin{aligned}
\frac{\alpha M_1}{2} - \alpha^2 L M_2^2 &= \frac{\alpha M_1}{2}\left(1 - \frac{2\alpha L M_2^2}{M_1}\right) \ge 0.495\alpha M_1,\\
\frac{2\alpha M_2^2}{nM_1} + \frac{2M_2^2\alpha^2 L}{n} &= \frac{\alpha M_2^2}{nM_1}(2 + 2\alpha M_1 L) \le \frac{2.01\alpha M_2^2}{nM_1},\\
\frac{\alpha M_2^2\gamma^2}{2nM_1} + \frac{2M_2^2\alpha^2 L}{n} &\le \frac{\alpha M_2^2\eta}{2nM_1},\\
\frac{\alpha M_2^2\gamma^2}{4nM_1} + \frac{M_2^2\alpha^2\gamma^2 L}{2n} &\le \frac{1.01\alpha M_2^2\gamma^2}{4nM_1}, \qquad (71)
\end{aligned}$$

where we use $\alpha M_2 L \le \frac{M_1}{200M_2}$, $\alpha M_1 L \le \frac{1}{200}$, and $\eta := \gamma^2 + 4\alpha L M_1 \le 1.02$. Moreover, by Assumption 2, we have

$$\|\nabla F(\overline{\mathbf{x}}^k)\|^2 \ge 2\mu\left[F(\overline{\mathbf{x}}^k) - F(x^*)\right]. \qquad (72)$$

Substituting (71) and (72) into (70) and subtracting $F(x^*)$ on both sides, and then multiplying n on both sides and using the definition $\tilde{\mu} := 0.99M_1\mu$, we have

$$\begin{aligned}
&n\mathbb{E}\left[F(\overline{\mathbf{x}}^{k+1}) - F(x^*)\right] \qquad (73)\\
\le&(1 - \tilde{\mu}\alpha)\cdot n\mathbb{E}\left[F(\overline{\mathbf{x}}^k) - F(x^*)\right] + \frac{\alpha M_2^2\eta}{2nM_1}\mathbb{E}\left[\|\mathbf{v}^k - \nabla f(\mathbf{x}^k)\|^2\right] \\
&+ \frac{1.01\alpha M_2^2}{M_1}\Big(2L^2\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{W}_\infty\mathbf{x}^k\|^2\right] + \frac{\gamma^2}{4}\mathbb{E}\left[\|\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k\|^2\right]\Big).
\end{aligned}$$

After substituting (36) into (73), we have

$$\begin{aligned}
&\frac{2n}{L}\mathbb{E}\left[F(\overline{\mathbf{x}}^{k+1}) - F(x^*)\right] \\
\le&\frac{2}{L}\left(\frac{\alpha M_2^2\eta}{2nM_1}\cdot 4L^2 B + \frac{2.02\alpha M_2^2}{M_1}L^2\right)\mathbb{E}\left[\|\mathbf{x}^k - \mathbf{W}_\infty\mathbf{x}^k\|^2\right] \\
&+ \left(\frac{\alpha M_2^2\eta}{2nM_1}8LB + (1 - \tilde{\mu}\alpha)\right)\cdot\frac{2n}{L}\mathbb{E}\left[F(\overline{\mathbf{x}}^k) - F(x^*)\right] \\
&+ \frac{0.51\alpha\gamma^2 M_2^2 L}{M_1(1 - \sigma^2)}\cdot\frac{1 - \sigma^2}{L^2}\mathbb{E}\left[\|\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k\|^2\right] \\
&+ \frac{\alpha M_2^2\eta}{LnM_1}4L^2 B\cdot\mathbb{E}\left[\|\tau^k - \mathbf{W}_\infty\tau^k\|^2\right] \\
&+ \frac{\alpha M_2^2\eta}{nLM_1}4L^2 B\cdot\frac{2n}{L}\mathbb{E}\left[F(\overline{\tau}^k) - F(\tau^*)\right].
\end{aligned}$$

By substituting the parameters in (24) and doing algebraic manipulations, we get (39) and complete the proof. Please see Part I [54] of the full version of this paper for details. □

VI. Proof of Lemma 8

Proof. By the update for $\mathbf{g}^{k+1}$ in (10), we have

$$\begin{aligned}
&\mathbf{g}^{k+1} - \mathbf{W}_\infty\mathbf{g}^{k+1} \qquad (74)\\
=&\mathbf{W}(\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k) + (I_{nd} - \mathbf{W}_\infty)(\mathbf{v}^{k+1} - \mathbf{v}^k).
\end{aligned}$$

By squaring both sides of (74) and then taking total expectation, we get

$$\begin{aligned}
&\mathbb{E}\left[\|\mathbf{g}^{k+1} - \mathbf{W}_\infty\mathbf{g}^{k+1}\|^2\right] \qquad (75)\\
\le&\mathbb{E}\Big[(1 + \vartheta_1)\|\mathbf{W}(\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k)\|^2 \\
&+ (1 + \vartheta_1^{-1})\|(I_{nd} - \mathbf{W}_\infty)(\mathbf{v}^{k+1} - \mathbf{v}^k)\|^2\Big] \\
\le&\mathbb{E}\Big[\frac{1 + \sigma^2}{2}\|\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k\|^2 + \frac{2}{1 - \sigma^2}\|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2\Big],
\end{aligned}$$

where the first inequality holds for any $\vartheta_1 > 0$ due to Young's inequality and the second inequality follows by setting $\vartheta_1 = \frac{1-\sigma^2}{2\sigma^2}$ and noting that $\|\mathbf{W}(\mathbf{g}^k - \mathbf{W}_\infty\mathbf{g}^k)\|^2 \le \sigma^2\|\mathbf{g}^k - \overline{\mathbf{g}}^k\|^2$ and $\|I_{nd} - \mathbf{W}_\infty\|_2 = 1$.

For the second term on the right-hand side of (75), we expand $\mathbb{E}\left[\|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2\right]$ as

$$\begin{aligned}
&\mathbb{E}\left[\|\mathbf{v}^{k+1} - \mathbf{v}^k\|^2\right] \qquad (76)\\
\le&2\mathbb{E}\left[\|\mathbf{v}^{k+1} - \mathbf{v}^k - \left(\nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k)\right)\|^2\right] \\
&+ 2\mathbb{E}\left[\|\nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k)\|^2\right] \\
\le&2\mathbb{E}\left[\|\mathbf{v}^{k+1} - \nabla f(\mathbf{x}^{k+1})\|^2\right] + 2\mathbb{E}\left[\|\mathbf{v}^k - \nabla f(\mathbf{x}^k)\|^2\right] \\
&+ 2L^2\mathbb{E}\left[\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2\right],
\end{aligned}$$

where in the last inequality, we use the fact that

$$\begin{aligned}&\mathbb{E}\left[\left\langle \mathbf{v}^{k+1}-\nabla f(\mathbf{x}^{k+1}),\mathbf{v}^{k}-\nabla f(\mathbf{x}^{k})\right\rangle\right]\\=&\mathbb{E}\left[\mathbb{E}\left[\left\langle \mathbf{v}^{k+1}-\nabla f(\mathbf{x}^{k+1}),\mathbf{v}^{k}-\nabla f(\mathbf{x}^{k})\right\rangle|\mathcal{F}^{k+1}\right]\right]=0.\end{aligned}$$

Substituting (76) into (75) yields (40). □

VII. Proof of Lemma 9

Proof. By the update rule for $\mathbf{x}^{k+1}$ in (10), we have

$$\mathbf{x}^{k+1}-\mathbf{x}^{k}=(\mathbf{W}-I_{nd})(\mathbf{x}^{k}-\mathbf{W}_{\infty}\mathbf{x}^{k})-\alpha\mathbf{d}^{k},$$

where we use $(\mathbf{W}-I_{nd})\mathbf{W}_{\infty}\mathbf{x}^{k}=0$. Then, we have

$$\begin{aligned}\|\mathbf{x}^{k+1}-\mathbf{x}^{k}\|^2&\le 2\|(\mathbf{W}-I_{nd})(\mathbf{x}^{k}-\mathbf{W}_{\infty}\mathbf{x}^{k})\|^2+2\alpha^2\|\mathbf{d}^{k}\|^2\\&\le 8\|\mathbf{x}^{k}-\mathbf{W}_{\infty}\mathbf{x}^{k}\|^2+2\alpha^2M_2^2\|\mathbf{g}^{k}\|^2,\end{aligned}\tag{77}$$

where we use $\|\mathbf{W}-I_{nd}\|_2\le 2$ and $\|\mathbf{d}^k\|^2=\sum_{i=1}^{n}\|H_i^k g_i\|^2\le M_2^2\|\mathbf{g}^k\|^2$ in the last inequality. For the term $\|\mathbf{g}^k\|^2$, we have

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{g}^{k}\|^2\right]&\le\mathbb{E}\left[2\|\mathbf{g}^{k}-\mathbf{W}_{\infty}\mathbf{g}^{k}\|^2+2\|\mathbf{W}_{\infty}\mathbf{g}^{k}\|^2\right]\\&=\mathbb{E}\left[2\|\mathbf{g}^{k}-\mathbf{W}_{\infty}\mathbf{g}^{k}\|^2+2n\|\overline{\mathbf{g}}^{k}\|^2\right].\end{aligned}\tag{78}$$

For the term $\|\overline{\mathbf{g}}^k\|^2$, we have

$$\begin{aligned}\mathbb{E}\|\overline{\mathbf{g}}^{k}\|^2=&\mathbb{E}\|\overline{\mathbf{v}}^{k}\|^2\\=&\mathbb{E}\left[\|\overline{\mathbf{v}}^{k}-\overline{\nabla f}(\mathbf{x}^{k})\|^2+\|\overline{\nabla f}(\mathbf{x}^{k})\|^2\right]\\=&\mathbb{E}\left[\frac{1}{n^2}\|\mathbf{v}^{k}-\nabla f(\mathbf{x}^{k})\|^2+\|\overline{\nabla f}(\mathbf{x}^{k})\|^2\right],\end{aligned}\tag{79}$$

where we use $\mathbb{E}[\langle\overline{\mathbf{v}}^k-\overline{\nabla f}(\mathbf{x}^k),\overline{\nabla f}(\mathbf{x}^k)\rangle\mid\mathcal{F}^k]=0$ from (33) to get the second equality and $\mathbb{E}\left[\mathbb{E}\left[\left\langle v_i^k-\nabla f_i(x_i^k),v_j^k-\nabla f_j(x_j^k)\right\rangle|\mathcal{F}^k\right]\right]=0,\ \forall i\neq j$ to get the last equality. For the term $\|\overline{\nabla f}(\mathbf{x}^k)\|^2$, we have

$$\begin{aligned}\|\overline{\nabla f}(\mathbf{x}^{k})\|^2\le&2\|\overline{\nabla f}(\mathbf{x}^{k})-\nabla F(\overline{\mathbf{x}}^{k})\|^2+2\|\nabla F(\overline{\mathbf{x}}^{k})\|^2\\\le&\frac{2L^2}{n}\|\mathbf{x}^{k}-\mathbf{W}_{\infty}\mathbf{x}^{k}\|^2+4L\left(F(\overline{\mathbf{x}}^{k})-F(x^*)\right),\end{aligned}\tag{80}$$

where we use (34) and (57) to derive the last inequality.

Taking total expectation on both sides of (77), substituting (78)–(80) into the result, and using the fact $8+8\alpha^2M_2^2L^2<8.01$, we obtain (41) and complete the proof. □

VIII. Proof of Lemma 10

Proof. If $\bmod\,(k+1,\mathcal{T})=0$, then $\mathbf{v}^{k+1}=\nabla f(\mathbf{x}^{k+1})$ and the proof is trivial. We consider $\bmod\,(k+1,\mathcal{T})\neq 0$ and thus $\tau^{k+1}=\tau^k$. Then, we have

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{v}^{k+1}-\nabla f(\mathbf{x}^{k+1})\|^2\right]\\\le&4L^2B\Big(\mathbb{E}\left[\|\mathbf{x}^{k+1}-\mathbf{W}_{\infty}\mathbf{x}^{k+1}\|^2\right]+\frac{2n}{L}\mathbb{E}\left[F(\overline{\mathbf{x}}^{k+1})-F(x^*)\right]\\&+\mathbb{E}\left[\|\tau^{k}-\mathbf{W}_{\infty}\tau^{k}\|^2\right]+\frac{2n}{L}\mathbb{E}\left[F(\overline{\tau}^{k})-F(x^*)\right]\Big)\\\le&4L^2B\left(1-\frac{0.99(1-\sigma^2)}{2}+\frac{4.04\alpha M_2^2L}{M_1}+\frac{4.08\alpha M_2^2LB}{nM_1}\right)\\&\quad\times\mathbb{E}\left[\|\mathbf{x}^{k}-\mathbf{W}_{\infty}\mathbf{x}^{k}\|^2\right]\\&+4L^2B\left(\frac{0.02(1-\sigma^2)\alpha\mu M_1}{L^2}+\frac{2.02\alpha M_2^2}{LM_1}\right)\\&\quad\times\mathbb{E}\left[\|\mathbf{g}^{k}-\mathbf{W}_{\infty}\mathbf{g}^{k}\|^2\right]\\&+4L^2B\left(1-\widetilde{\mu}\alpha+\frac{4.08\alpha M_2^2LB}{nM_1}+0.011(1-\sigma^2)\alpha\mu M_1\right)\\&\quad\times\frac{2n}{L}\mathbb{E}\left[F(\overline{\mathbf{x}}^{k})-F(x^*)\right]\\&+4L^2B\left(1+\frac{4.08\alpha M_2^2LB}{nM_1}+\frac{0.01(1-\sigma^2)\alpha\mu M_1}{L^2}\cdot\frac{4L^2B}{n}\right)\\&\quad\times\left(\mathbb{E}\left[\|\tau^{k}-\mathbf{W}_{\infty}\tau^{k}\|^2\right]+\frac{2n}{L}\mathbb{E}\left[\|F(\overline{\tau}^{k})-F(x^*)\|^2\right]\right),\end{aligned}$$

where the first inequality follows from Lemma 6 and the fact that $\tau^{k+1}=\tau^k$; the second inequality follows from (63), (73), and Lemma 6. By substituting the parameters in (24) and doing algebraic manipulations, we get (42) and complete the proof. Please see Part I [54] of the full version of this paper for details. □

IX. Proof of Corollary 3

Proof. Substituting (36), (41), and (42) into (40), we obtain

$$\begin{aligned}&\frac{1-\sigma^2}{L^2}\mathbb{E}\left[\|\mathbf{g}^{k+1}-\mathbf{W}_{\infty}\mathbf{g}^{k+1}\|^2\right]\\\le&\Big(32.04+32B+64\alpha^2M_2^2L^2\frac{B}{n}\Big)\mathbb{E}\left[\|\mathbf{x}^{k}-\mathbf{W}_{\infty}\mathbf{x}^{k}\|^2\right]\\&+\left(32\alpha^2M_2^2L^2+32.16B+64\alpha^2M_2^2L^2\frac{B}{n}\right)\\&\quad\times\frac{2n}{L}\mathbb{E}\left[F(\overline{\mathbf{x}}^{k})-F(x^*)\right]\\&+\left(\frac{1+\sigma^2}{2}+\frac{16L^2\alpha^2M_2^2}{1-\delta^2}+\frac{4}{1-\sigma^2}4L^2B\frac{3\alpha M_2^2}{LM_1}\right)\\&\quad\times\frac{1-\sigma^2}{L^2}\mathbb{E}\left[\|\mathbf{g}^{k}-\mathbf{W}_{\infty}\mathbf{g}^{k}\|^2\right]\\&+\left(64\alpha^2M_2^2L^2\frac{B}{n}+32.16B\right)\mathbb{E}\left[\|\tau^{k}-\mathbf{W}_{\infty}\tau^{k}\|^2\right]\\&+\left(64\alpha^2M_2^2L^2\frac{B}{n}+32.16B\right)\frac{2n}{L}\mathbb{E}\left[F(\overline{\tau}^{k})-F(\tau^*)\right].\end{aligned}$$

By substituting the parameters in (24) and doing algebraic manipulations, we get (43) and complete the proof. Please see Part I [54] of the full version of this paper for details. □