SUPPLEMENTARY MATERIAL

Proof to Theorem 1

Proof. We start from decomposing the expectation of $\|x_{k+1}-x^*\|^2$:

$$\begin{aligned}
&\mathbb{E}\|x_{k+1}-x^*\|^2\\
=\;&\mathbb{E}\|x_k-x^*\|^2+\mathbb{E}\|x_{k+1}-x_k\|^2+2\mathbb{E}\langle x_{k+1}-x_k, x_k-x^*\rangle\\
=\;&\mathbb{E}\|x_k-x^*\|^2+\gamma^2\mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k)-(\partial G_{j_k}(\tilde{x}))^\top\nabla F_{i_k}(\tilde{G})+\nabla f(\tilde{x})\right\|^2\\
&-2\gamma\mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k)-(\partial G_{j_k}(\tilde{x}))^\top\nabla F_{i_k}(\tilde{G})+\nabla f(\tilde{x}), x_k-x^*\right\rangle. \qquad (13)
\end{aligned}$$

Given the observation

$$\begin{aligned}
&\mathbb{E}\left\langle-(\partial G_{j_k}(\tilde{x}))^\top\nabla F_{i_k}(\tilde{G})+\nabla f(\tilde{x}), x_k-x^*\right\rangle\\
=\;&\mathbb{E}\left\langle-\mathbb{E}_{j_k,i_k}J^\top_{G_{j_k}}(\tilde{x})\nabla F_{i_k}(\tilde{G})+\nabla f(\tilde{x}), x_k-x^*\right\rangle\\
=\;&\mathbb{E}\langle-\nabla f(\tilde{x})+\nabla f(\tilde{x}), x_k-x^*\rangle\\
=\;&0,
\end{aligned}$$

It follows from (13) that

$$\begin{aligned}
\mathbb{E}\|x_{k+1}-x^*\|^2 =\;& \mathbb{E}\|x_k-x^*\|^2-2\gamma\underbrace{\mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k), x_k-x^*\right\rangle}_{=:T_1}\\
&+\gamma^2\underbrace{\mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k)-(\partial G_{j_k}(\tilde{x}))^\top\nabla F_{i_k}(\tilde{G})+\nabla f(\tilde{x})\right\|^2}_{=:T_2}. \qquad (14)
\end{aligned}$$

We then bound T_1. From the strong convexity of $f(x)$ we have the following inequality:

$$\langle\nabla f(x), x-x^*\rangle \geqslant \mu_f\|x_k-x^*\|^2. \qquad (15)$$

It follows that

$$\begin{aligned}
T_1 =\;& \mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k), x_k-x^*\right\rangle\\
=\;& \mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k)-\nabla f(x_k), x_k-x^*\right\rangle+\mathbb{E}\langle\nabla f(x_k), x_k-x^*\rangle\\
\overset{(15)}{\geqslant}\;& \underbrace{\mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k)-\nabla f(x_k), x_k-x^*\right\rangle}_{=:T_3}+\mathbb{E}\mu_f\|x_k-x^*\|^2. \qquad (16)
\end{aligned}$$

We then bound T_3. Recall that for any $\alpha>0$ we have

$$\frac{1}{\alpha}x^2+\alpha y^2 \geqslant 2|\langle x,y\rangle| \geqslant |\langle x,y\rangle|. \qquad (17)$$

It follows that

$$T_3 = \mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top\nabla F_{i_k}(\hat{G}_k)-\nabla f(x_k), x_k-x^*\right\rangle$$

$$
\begin{aligned}
&= \mathbb{E}\left\langle (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k)), x_k - x^* \right\rangle \\
&\overset{(17)}{\geqslant} -\frac{1}{\alpha} \underbrace{\mathbb{E}\left\| (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k)) \right\|^2}_{=:T_4} - \alpha \mathbb{E}\|x_k - x^*\|^2, \forall \alpha > 0 \quad (18)
\end{aligned}
$$

For T_4, from the definition of $\hat{G}_k$,

$$
\begin{aligned}
T_4 &= \mathbb{E}\left\| (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k)) \right\|^2 \\
&= \mathbb{E}\left\| (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}\left(\tilde{G} - \frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k))\right) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k)) \right\|^2 \\
&\leqslant \mathbb{E}\left\| (\partial G_{j_k}(x_k))^\top \right\|^2 \left\| \nabla F_{i_k}\left(\tilde{G} - \frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k))\right) - \nabla F_{i_k}(G(x_k)) \right\|^2 \\
&\overset{(7)}{\leqslant} B_G^2 \mathbb{E}\left\| \nabla F_{i_k}\left(\tilde{G} - \frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k))\right) - \nabla F_{i_k}(G(x_k)) \right\|^2 \\
&\overset{(8)}{\leqslant} B_G^2 L_F^2 \underbrace{\mathbb{E}\left\| \tilde{G} - \frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - G(x_k) \right\|^2}_{=:T_0}. \qquad (19)
\end{aligned}
$$

Let $\alpha = \frac{\mu_f}{8}$ in (18) and put the bound of T_4 in it, we obtain

$$
\begin{aligned}
T_3 &\overset{(18)}{\geqslant} -\frac{1}{\alpha} T_4 - \alpha \mathbb{E}\|x_k - x^*\|^2 \\
&\overset{(19)}{\geqslant} -\frac{8B_G^2 L_F^2}{\mu_f} T_0 - \frac{\mu_f}{8} \mathbb{E}\|x_k - x^*\|^2. \qquad (20)
\end{aligned}
$$

Then put this bound on T_3 to (16).

$$
\begin{aligned}
T_1 &\overset{(16)}{\geqslant} T_3 + \mathbb{E}\mu_f \|x_k - x^*\|^2 \\
&\overset{(20)}{\geqslant} -\frac{8B_G^2 L_F^2}{\mu_f} T_0 + \frac{7\mu_f}{8} \mathbb{E}\|x_k - x^*\|^2. \qquad (21)
\end{aligned}
$$

Now we have T_1 bounded. We use this bound to bound the T_1 in the equality (14) at the beginning.

$$
\begin{aligned}
\mathbb{E}\|x_{k+1} - x^*\|^2 &\overset{(14)}{=} \mathbb{E}\|x_k - x^*\|^2 - 2\gamma T_1 + \gamma^2 T_2 \\
&\overset{(21)}{\leqslant} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f \gamma}{4} \mathbb{E}\|x_k - x^*\|^2 + \frac{16\gamma B_G^2 L_F^2}{\mu_f} T_0 + \gamma^2 T_2. \qquad (22)
\end{aligned}
$$

We then bound T_2. Recall for any β we have

$$
\|\beta_1 + \beta_2 + \cdots + \beta_t\|^2 \leqslant t\left(\|\beta_1\|^2 + \cdots + \|\beta_t\|^2\right), \forall t \in \mathbb{N}_+. \qquad (23)
$$

From the definition of T_2 in (14) we have the following bound on T_2:

$$\begin{aligned}
T_2 &= \mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(\tilde{x}))^\top \nabla F_{i_k}(\tilde{G}) + \nabla f(\tilde{x})\right\|^2 \\
&\overset{(23)}{\leqslant} 2\mathbb{E}\|\nabla f(\tilde{x})\|^2 + 2\mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(\tilde{x}))^\top \nabla F_{i_k}(\tilde{G})\right\|^2 \\
&\overset{(23)}{\leqslant} 2\mathbb{E}\|\nabla f(\tilde{x})\|^2 + 4\underbrace{\mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k))\right\|^2}_{\text{the same as } T_4} \\
&\quad +4\underbrace{\mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k)) - (\partial G_{j_k}(\tilde{x}))^\top \nabla F_{i_k}(\tilde{G})\right\|^2}_{=:T_5} \\
&\overset{(19)}{\leqslant} 2\mathbb{E}\|\nabla f(\tilde{x})\|^2 + 4B_G^2 L_F^2 T_0 + 4T_5. \qquad (24)
\end{aligned}$$

To bound T_5, we simply use the Lipschitzian condition (10)

$$\begin{aligned}
T_5 &= \mathbb{E}\left\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(g(x_k)) - (\partial G_{j_k}(\tilde{x}))^\top \nabla F_{i_k}(\tilde{G})\right\|^2 \\
&\overset{(10)}{\leqslant} L_f^2 \mathbb{E}\|x_k - \tilde{x}\|^2,
\end{aligned}$$

Put this bound back to (24) we obtain

$$T_2 \overset{(24)}{\leqslant} 2\mathbb{E}\|\nabla f(\tilde{x})\|^2 + 4B_G^2 L_F^2 T_0 + 4L_f^2 \mathbb{E}\|x_k - \tilde{x}\|^2. \qquad (25)$$

Now we have T_2 bounded, and we put this bound back to (22).

$$\begin{aligned}
\mathbb{E}\|x_{k+1} - x^*\|^2 &\overset{(22)}{\leqslant} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f\gamma}{4}\mathbb{E}\|x_k - x^*\|^2 + \frac{16\gamma B_G^2 L_F^2}{\mu_f} T_0 + \gamma^2 T_2 \\
&\overset{(25)}{\leqslant} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f\gamma}{4}\mathbb{E}\|x_k - x^*\|^2 + \frac{16\gamma B_G^2 L_F^2}{\mu_f} T_0 \\
&\quad +2\gamma^2 \mathbb{E}\|\nabla f(\tilde{x})\|^2 + 4\gamma^2 B_G^2 L_F^2 T_0 + 4\gamma^2 L_f^2 \mathbb{E}\|x_k - \tilde{x}\|^2 \\
&\overset{(11)}{=} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f\gamma}{4}\mathbb{E}\|x_k - x^*\|^2 + 2\gamma^2 L_f^2 \mathbb{E}\|\tilde{x} - x^*\|^2 \\
&\quad + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right) T_0 + 4\gamma^2 L_f^2 \mathbb{E}\|x_k - \tilde{x}\|^2, \qquad (26)
\end{aligned}$$

where the last step comes from (11) by letting $x = x_k$ and $y = x^*$.

There is still one term, T_0, not bounded. We now start to bound it. From the definition of T_0 in (19):

$$\begin{aligned}
T_0 &= \mathbb{E}\left\|\tilde{G} - \frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - G(x_k)\right\|^2 \\
&= \mathbb{E}\left\|\frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - (\tilde{G} - G(x_k))\right\|^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{A^2}\mathbb{E}\left\|\sum_{1\le j\le A}((G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - (\tilde{G} - G(x_k)))\right\|^2 \\
&= \frac{1}{A^2}\sum_{1\le j\le A}\mathbb{E}\|(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - (\tilde{G} - G(x_k))\|^2,
\end{aligned}
$$

where the last step comes from the fact that the indices in $\mathcal{A}_k$ are independent. Specifically,

$$
\begin{aligned}
&\mathbb{E}\left\|\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k))\right\|^2 \\
=&\mathbb{E}\sum_{1\le j\le A}\|(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k))\|^2 \\
&+ 2\mathbb{E}\sum_{1\le j'<j\le A}\langle(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k)), (G_{\mathcal{A}_k[j']}(\tilde{x}) - G_{\mathcal{A}_k[j']}(x_k) - \tilde{G} + G(x_k))\rangle \\
=&\mathbb{E}\sum_{1\le j\le A}\|(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k))\|^2 \\
&+ 2\mathbb{E}\sum_{1\le j'<j\le A}\langle\mathbb{E}_{\mathcal{A}_k[j]}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k)), (G_{\mathcal{A}_k[j']}(\tilde{x}) - G_{\mathcal{A}_k[j']}(x_k) - \tilde{G} + G(x_k))\rangle \\
=&\mathbb{E}\sum_{1\le j\le A}\|(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k))\|^2 \\
&+ 2\mathbb{E}\sum_{1\le j'<j\le A}\langle 0, (G_{\mathcal{A}_k[j']}(\tilde{x}) - G_{\mathcal{A}_k[j']}(x_k) - \tilde{G} + G(x_k))\rangle \\
=&\mathbb{E}\sum_{1\le j\le A}\|(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k) - \tilde{G} + G(x_k))\|^2. \qquad (27)
\end{aligned}
$$

Finally T_0 can be bounded by

$$
\begin{aligned}
T_0 &= \frac{1}{A^2}\sum_{1\le j\le A}\mathbb{E}\|(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - (\tilde{G} - G(x_k))\|^2 \\
&\overset{(23)}{\leqslant} \frac{4}{A^2}\sum_{1\le j\le A}\mathbb{E}\Big(\|G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x^*)\|^2 + \left\|G_{\mathcal{A}_k[j]}(x_k) - G_{\mathcal{A}_k[j]}(x^*)\right\|^2 \\
&\quad + \|\tilde{G} - G(x^*)\|^2 + \|G(x_k) - G(x^*)\|^2\Big) \\
&\overset{(7)}{\leqslant} \frac{8B_G^2}{A^2}\sum_{1\le j\le A}\mathbb{E}(\|\tilde{x} - x^*\|^2 + \|x_k - x^*\|^2) \\
&= \frac{8B_G^2}{A}\mathbb{E}(\|\tilde{x} - x^*\|^2 + \|x_k - x^*\|^2). \qquad (28)
\end{aligned}
$$

By passing this bound to (26) we finally get all T terms bounded:

$$\mathbb{E}\|x_{k+1} - x^*\|^2$$

$$
\begin{aligned}
&\overset{(26)}{\leqslant} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f\gamma}{4}\mathbb{E}\|x_k - x^*\|^2 + 2\gamma^2 L_f^2 \mathbb{E}\|\tilde{x} - x^*\|^2 \\
&\quad + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right) T_0 + 4\gamma^2 L_f^2 \mathbb{E}\|x_k - \tilde{x}\|^2 \\
&\overset{(28)}{\leqslant} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f\gamma}{4}\mathbb{E}\|x_k - x^*\|^2 + 2\gamma^2 L_f^2 \mathbb{E}\|\tilde{x} - x^*\|^2 \\
&\quad + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right) \frac{8B_G^2}{A}\mathbb{E}(\|\tilde{x} - x^*\|^2 + \|x_k - x^*\|^2) + 4\gamma^2 L_f^2 \mathbb{E}\|x_k - x^* + x^* - \tilde{x}\|^2 \\
&\overset{(23)}{\leqslant} \mathbb{E}\|x_k - x^*\|^2 - \frac{7\mu_f\gamma}{4}\mathbb{E}\|x_k - x^*\|^2 + 2\gamma^2 L_f^2 \mathbb{E}\|\tilde{x} - x^*\|^2 \\
&\quad + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right) \frac{8B_G^2}{A}\mathbb{E}(\|\tilde{x} - x^*\|^2 + \|x_k - x^*\|^2) \\
&\quad + 8\gamma^2 L_f^2 \mathbb{E}(\|x_k - x^*\|^2 + \|\tilde{x} - x^*\|^2) \\
&= \mathbb{E}\|x_k - x^*\|^2 \\
&\quad - \left(\frac{7\mu_f\gamma}{4} - \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} - 8\gamma^2 L_f^2\right)\mathbb{E}\|x_k - x^*\|^2 \\
&\quad + \left(\left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} + 10\gamma^2 L_f^2\right)\mathbb{E}\|\tilde{x} - x^*\|^2.
\end{aligned}
$$

Summing this inequality from $k = 0$ to $k = K - 1$, we obtain

$$
\begin{aligned}
&\mathbb{E}\|x_{k+K} - x^*\|^2 \\
\leqslant\ & \mathbb{E}\|\tilde{x} - x^*\|^2 \\
&- \left(\frac{7\mu_f\gamma}{4} - \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} - 8\gamma^2 L_f^2\right)\sum_{k=0}^{K-1}\mathbb{E}\|x_k - x^*\|^2 \\
&+ K\left(\left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} + 10\gamma^2 L_f^2\right)\mathbb{E}\|\tilde{x} - x^*\|^2.
\end{aligned}
$$

Discarding the left hand side, we complete the proof by

$$
\begin{aligned}
&\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\|x_k - x^*\|^2 \\
\leqslant\ & \frac{\frac{1}{K} + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} + 10\gamma^2 L_f^2}{\frac{7\mu_f\gamma}{4} - \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} - 8\gamma^2 L_f^2}\mathbb{E}\|\tilde{x} - x^*\|^2.
\end{aligned}
$$

□

Proof to Corollary 1

Proof. To appropriately choose γ, K and A in Algorithm 2, the key is to ensure the coefficient $\frac{\beta_1}{\beta_2} < 1$ in Theorem 1:

$$
\frac{\beta_1}{\beta_2} = \frac{\frac{1}{K} + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} + 10\gamma^2 L_f^2}{\frac{7\mu_f\gamma}{4} - \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right)\frac{8B_G^2}{A} - 8\gamma^2 L_f^2}.
$$

We choose A satisfying both

$$\begin{aligned}
4\gamma^2 B_G^2 L_F^2 \frac{8B_G^2}{A} &\leqslant \frac{\mu_f \gamma}{4}, \\
\frac{16\gamma B_G^2 L_F^2}{\mu_f} \frac{8B_G^2}{A} &\leqslant \frac{\mu_f \gamma}{4},
\end{aligned}$$

which is equivalent to

$$A \geqslant \max\left\{ \frac{128\gamma B_G^4 L_F^2}{\mu_f}, \frac{512 B_G^4 L_F^2}{\mu_f^2} \right\}.$$

We choose γ satisfying

$$8\gamma^2 L_f^2 \leqslant \frac{\mu_f \gamma}{4},$$

which is equivalent to

$$\gamma \leqslant \frac{\mu_f}{32 L_f^2}.$$

It follows that

$$\begin{aligned}
&\frac{\frac{1}{K} + \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right) \frac{8B_G^2}{A} + 10\gamma^2 L_f^2}{\frac{7\mu_f \gamma}{4} - \left(\frac{16\gamma B_G^2 L_F^2}{\mu_f} + 4\gamma^2 B_G^2 L_F^2\right) \frac{8B_G^2}{A} - 8\gamma^2 L_f^2} \\
\leqslant\ & \frac{\frac{1}{K} + \frac{13\mu_f \gamma}{16}}{\mu_f \gamma} \\
=\ & \frac{13}{16} + \frac{1}{K\mu_f \gamma}.
\end{aligned}$$

We then choose K satisfying

$$\frac{1}{K\mu_f \gamma} \leqslant \frac{1}{16},$$

which is equivalent to

$$K \geqslant \frac{16}{\mu_f \gamma}.$$

Thus choosing γ, A, and K appropriately in the following to satisfy all conditions derived above

$$\begin{aligned}
\gamma &= \frac{\mu_f}{32 L_f^2}, \\
A &= \frac{512 B_G^4 L_F^2}{\mu_f^2}, \\
K &= \frac{512 L_f^2}{\mu_f^2},
\end{aligned}$$

we obtain a linear convergence rate of coefficient $\frac{\beta_1}{\beta_2} = \frac{7}{8}$ from Theorem 1. □

Lemma 1. *Under the assumption in* (10), *we have*

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{1}{m}\sum_{j=1}^{m}(\partial G_j(x))^\top \nabla F_i(G(x)) - \frac{1}{m}\sum_{j=1}^{m}(\partial G_j(x^*))^\top \nabla F_i(G(x^*))\right\|^2 \leqslant 2L_f(f(x)-f^*).$$

Proof. Recall that at the optimal point we always have

$$f'(x^*) = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}(\partial G_j(x^*))^\top \nabla F_i(G(x^*)) = 0. \tag{29}$$

We can derive the Lipschitz constant of $F_i(G(x))$ from (10)

$$\begin{aligned}
&\|\nabla F_i(G(x)) - \nabla F_i(G(y))\| \\
=\ & \frac{1}{m}\left\|\sum_j (\partial G_j(x))^\top \nabla F_i(G(x)) - \sum_j (\partial G_j(y))^\top \nabla F_i(G(y))\right\| \\
\leqslant\ & \frac{1}{m}\sum_j \|(\partial G_j(x))^\top \nabla F_i(G(x)) - (\partial G_j(y))^\top \nabla F_i(G(y))\| \\
\leqslant\ & L_f\|x-y\|, \forall i.
\end{aligned} \tag{30}$$

From this Lipschitz condition, we obtain

$$\begin{aligned}
F_i(G(x)) \overset{(30)}{\geqslant}\ & F_i(G(x^*)) + \frac{1}{m}\left\langle \sum_{j=1}^{m}(\partial G_j(x^*))^\top \nabla F_i(G(x^*)), x-x^*\right\rangle \\
& + \frac{1}{2L_f}\left\|\frac{1}{m}\sum_{j=1}^{m}(\partial G_j(x))^\top \nabla F_i(x) - \frac{1}{m}\sum_{j=1}^{m}(\partial G_j(x^*))^\top \nabla F_i(x^*)\right\|^2.
\end{aligned}$$

Summing from $i=1$ to $i=n$, using (29) and noting that $\frac{1}{n}\sum_{i=1}^{n} F_i(G(x)) = f(x)$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{1}{m}\sum_{j=1}^{m}(\partial G_j(x))^\top \nabla F_i(x) - \frac{1}{m}\sum_{j=1}^{m}(\partial G_j(x^*))^\top \nabla F_i(x^*)\right\|^2 \leqslant 2L_f(f(x)-f^*),$$

completing the proof. □

Proof to Theorem 2

Proof. Note that in this proof we redefine the terms $T_1, T_2, \ldots$, and they may not refer to the same expressions in the proof of Theorem 1. From

$$x_{k+1} - x_k = -\gamma((\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}').$$

we immediately obtain

$$\mathbb{E}\|x_{k+1}-x^*\|^2 = \mathbb{E}\|x_k - x^*\|^2 + \mathbb{E}\|x_{k+1}-x_k\|^2 + 2\mathbb{E}\langle x_{k+1}-x_k, x_k - x^*\rangle$$

$$
\begin{aligned}
&= \mathbb{E}\|x_k - x^*\|^2 + \gamma^2\mathbb{E}\|(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}'\|^2 \\
&\quad -2\gamma\mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}', x_k - x^*\rangle.
\end{aligned}
$$

Note that the last term can be simplified:

$$
\begin{aligned}
& \mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}', x_k - x^*\rangle \\
= & \mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - \mathbb{E}_{i_k}(\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}', x_k - x^*\rangle \\
= & \mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - \tilde{f}' + \tilde{f}', x_k - x^*\rangle \\
= & \mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k), x_k - x^*\rangle.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\mathbb{E}\|x_{k+1} - x^*\|^2 &= \mathbb{E}\|x_k - x^*\|^2 - 2\gamma \underbrace{\mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k), x_k - x^*\rangle}_{=:T_1} \\
&\quad +\gamma^2 \underbrace{\mathbb{E}\|(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}'\|^2}_{=:T_2}. \qquad (31)
\end{aligned}
$$

First we estimate the lower bound for T_1.:

$$
\begin{aligned}
T_1 &= \mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k), x_k - x^*\rangle \\
&= \underbrace{\mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - \nabla f(x_k), x_k - x^*\rangle}_{=:T_3} + \mathbb{E}\langle\nabla f(x_k), x_k - x^*\rangle \\
&\geqslant T_3 + \mathbb{E}(f(x_k) - f^*). \qquad (32)
\end{aligned}
$$

Then we estimate the lower bound for T_3

$$
\begin{aligned}
T_3 &= \mathbb{E}\langle(\hat{G}'_k)^\top \nabla F_{i_k}(\hat{G}_k) - \nabla f(x_k), x_k - x^*\rangle \\
&= \mathbb{E}\left\langle(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k)), x_k - x^*\right\rangle,
\end{aligned}
$$

where j_k is a new (imaginary) random variable that is chosen uniformly randomly from $\{1, \cdots, m\}$ and is independent of other random variables. $\mathbb{E}$ also takes expectation on j_k. Thus using the same technique as we use in (18) while proving Theorem 1, we obtain

$$
T_3 \geqslant -\frac{1}{\alpha}\underbrace{\mathbb{E}\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k))\|^2}_{=:T_4} - \alpha\mathbb{E}\|x_k - x^*\|^2, \forall \alpha > 0.
$$

and

$$
\begin{aligned}
T_4 &= \mathbb{E}\|(\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(\hat{G}_k) - (\partial G_{j_k}(x_k))^\top \nabla F_{i_k}(G(x_k))\|^2 \\
&\overset{(19)}{\leqslant} B_G^2 L_F^2 T_0,
\end{aligned}
$$

where

$$
T_0 := \mathbb{E}\left\|\tilde{G} - \frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x}) - G_{\mathcal{A}_k[j]}(x_k)) - G(x_k)\right\|^2.
$$

Let $\alpha = \frac{\mu_f}{8}$, we obtain

$$T_3 \geqslant -\frac{8B_G^2L_F^2}{\mu_f}T_0 - \frac{\mu_f}{8}\mathbb{E}\|x_k - x^*\|^2.$$

Put the bound of T_3 into (32) and note that

$$\mu_f\|x_k - x^*\|^2 \leqslant 2(f(x_k) - f^*). \tag{33}$$

We obtain

$$\begin{aligned} T_1 &\geqslant -\frac{8B_G^2L_F^2}{\mu_f}T_0 - \frac{\mu_f}{8}\mathbb{E}\|x_k - x^*\|^2 + \mathbb{E}(f(x_k) - f^*) \\ &\overset{(33)}{\geqslant} -\frac{8B_G^2L_F^2}{\mu_f}T_0 + \frac{3}{4}\mathbb{E}(f(x_k) - f^*). \end{aligned} \tag{34}$$

Now we have T_1 bounded. We then start to bound T_2. From the definition of T_2 we have

$$\begin{aligned} T_2 &= \mathbb{E}\|(\hat{G}_k')^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) + \tilde{f}'\|^2 \\ &\overset{(23)}{\leqslant} 2\mathbb{E}\|\tilde{f}'\|^2 + 2\mathbb{E}\|(\hat{G}_k')^\top \nabla F_{i_k}(\hat{G}_k) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G})\|^2 \\ &= 2\mathbb{E}\|\tilde{f}'\|^2 + 2\mathbb{E}\Bigg\|(\hat{G}_k')^\top \nabla F_{i_k}(\hat{G}_k) - \frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k)) \\ &\quad + \frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k)) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G})\Bigg\|^2 \\ &\overset{(23)}{\leqslant} 2\mathbb{E}\|\tilde{f}'\|^2 + 4\mathbb{E}\left\|(\hat{G}_k')^\top \nabla F_{i_k}(\hat{G}_k) - \frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k))\right\|^2 \\ &\quad + 4\mathbb{E}\left\|\frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k)) - (\tilde{G}')^\top \nabla F_{i_k}(\tilde{G})\right\|^2 \\ &\overset{(23)}{\leqslant} 2\mathbb{E}\|\tilde{f}'\|^2 + 4\mathbb{E}\left\|(\hat{G}_k')^\top \nabla F_{i_k}(\hat{G}_k) - \frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k))\right\|^2 \\ &\quad + 8\mathbb{E}\left\|\frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k)) - \frac{1}{m}\sum_{j=1}^m (\partial G_j(x^*))^\top \nabla F_{i_k}(G(x^*))\right\|^2 \\ &\quad + 8\mathbb{E}\left\|(\tilde{G}')^\top \nabla F_{i_k}(\tilde{G}) - \frac{1}{m}\sum_{j=1}^m (\partial G_j(x^*))^\top \nabla F_{i_k}(G(x^*))\right\|^2 \\ &\leqslant 4L_f\mathbb{E}(f(\tilde{x}) - f^*) + 4\,\underbrace{\mathbb{E}\left\|(\hat{G}_k')^\top \nabla F_{i_k}(\hat{G}_k) - \frac{1}{m}\sum_{j=1}^m (\partial G_j(x_k))^\top \nabla F_{i_k}(G(x_k))\right\|^2}_{=:T_5} \end{aligned} \tag{35}$$

$$+16(L_f\mathbb{E}(f(\tilde{x})-f^*)+L_f\mathbb{E}(f(x_k)-f^*)), \tag{36}$$

where the last step comes from Lemma 1 and $\frac{\|\nabla f(\tilde{x})\|^2}{2L_f}\leqslant f(\tilde{x})-f^*$.

Note that T_5 can be bounded by

$$\begin{aligned}
T_5 &= \mathbb{E}\left\|(\hat{G}_k')^\top\nabla F_{i_k}(\hat{G}_k)-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\nabla F_{i_k}(G(x_k))\right\|^2\\
&\overset{(23)}{\leqslant} 2\mathbb{E}\left\|(\hat{G}_k')^\top\nabla F_{i_k}(\hat{G}_k)-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\nabla F_{i_k}(\hat{G}_k)\right\|^2\\
&\quad+2\mathbb{E}\left\|\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\nabla F_{i_k}(G(x_k))-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\nabla F_{i_k}(\hat{G}_k)\right\|^2\\
&\overset{(12),\,(7)}{\leqslant} 2B_F^2\mathbb{E}\left\|(\hat{G}_k')^\top-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\right\|^2+2B_G^2\mathbb{E}\|\nabla F_{i_k}(G(x_k))-\nabla F_{i_k}(\hat{G}_k)\|^2\\
&\leqslant 2B_F^2\mathbb{E}\left\|(\tilde{G}')^\top-\frac{1}{B}\left(\sum_{1\le j\le B}((\partial G_{\mathcal{B}_j[j]}(\tilde{x}))^\top-(\partial G_{\mathcal{B}_k[j]}(x_k))^\top)\right)-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\right\|^2\\
&\quad+2B_G^2L_F^2\mathbb{E}\left\|\tilde{G}-\frac{1}{A}\sum_{1\le j\le A}(G_{\mathcal{A}_k[j]}(\tilde{x})-G_{\mathcal{A}_k[j]}(x_k))-G(x_k)\right\|^2\\
&= \frac{2B_F^2}{B^2}\mathbb{E}\left\|-\sum_{1\le j\le B}\left(((\partial G_{\mathcal{B}_k[j]}(\tilde{x}))^\top-(\partial G_{\mathcal{B}_k[j]}(x_k))^\top)-\left((\tilde{G}')^\top-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\right)\right)\right\|^2\\
&\quad+\frac{2B_G^2L_F^2}{A^2}\mathbb{E}\left\|-\sum_{1\le j\le A}((G_{\mathcal{A}_k[j]}(\tilde{x})-G_{\mathcal{A}_k[j]}(x_k))-(G(x_k)-\tilde{G}))\right\|^2
\end{aligned}$$

Using the same technique as in (27), the above inequality continues as

$$\begin{aligned}
&= \frac{2B_F^2}{B^2}\mathbb{E}\sum_{1\le j\le B}\left\|(\partial G_{\mathcal{B}_k[j]}(x_k))^\top-(\partial G_{\mathcal{B}_k[j]}(\tilde{x}))^\top+\left((\tilde{G}')^\top-\frac{1}{m}\sum_{j=1}^m(\partial G_j(x_k))^\top\right)\right\|^2\\
&\quad+\frac{2B_G^2L_F^2}{A^2}\mathbb{E}\sum_{1\le j\le A}\|G_{\mathcal{A}_k[j]}(x_k)-G_{\mathcal{A}_k[j]}(\tilde{x})+(G(x_k)-\tilde{G})\|^2\\
&\overset{(7),(9),(23)}{\leqslant} \frac{2B_F^2}{B^2}\sum_{1\le j\le B}8L_G^2(\mathbb{E}\|\tilde{x}-x^*\|^2+\mathbb{E}\|x_k-x^*\|^2)\\
&\quad+\frac{2B_G^2L_F^2}{A^2}\sum_{1\le j\le A}8B_G^2(\mathbb{E}\|\tilde{x}-x^*\|^2+\mathbb{E}\|x_k-x^*\|^2)\\
&= 16\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)(\mathbb{E}\|\tilde{x}-x^*\|^2+\mathbb{E}\|x_k-x^*\|^2)
\end{aligned}$$

$$\leqslant \frac{32}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)(\mathbb{E}(f(\tilde{x})-f^*)+\mathbb{E}(f(x_k)-f^*)).$$

Now we continue to bound T_2 in (36) using the bound for T_5 above:

$$\begin{aligned}
T_2 &\overset{(36)}{\leqslant} 4L_f\mathbb{E}(f(\tilde{x})-f^*)+4T_5+16(L_f(f(\tilde{x})-f^*)+L_f\mathbb{E}(f(x_k)-f^*)) \\
&= 20L_f\mathbb{E}(f(\tilde{x})-f^*)+16L_f\mathbb{E}(f(x_k)-f^*)+4T_5 \\
&\leqslant 20L_f\mathbb{E}(f(\tilde{x})-f^*)+16L_f\mathbb{E}(f(x_k)-f^*) \\
&\quad +\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)(\mathbb{E}(f(\tilde{x})-f^*)+\mathbb{E}(f(x_k)-f^*)) \\
&= \left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+16L_f\right)\mathbb{E}(f(x_k)-f^*) \\
&\quad +\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+20L_f\right)\mathbb{E}(f(\tilde{x})-f^*). \qquad (37)
\end{aligned}$$

Now we have T_2 bounded. Finally we put the bounds of T_2, T_1 in (37) and (34) into (31) and note that using the same procedure in the proof of Theorem 1 (see (28)) we have

$$T_0 \leqslant \frac{8B_G^2}{A}\mathbb{E}(\|\tilde{x}-x^*\|^2+\|x_k-x^*\|^2). \qquad (38)$$

We obtain:

$$\begin{aligned}
&\mathbb{E}\|x_{k+1}-x^*\|^2 \\
\overset{(31)}{=}\ &\mathbb{E}\|x_k-x^*\|^2-2\gamma T_1+\gamma^2T_2 \\
\overset{(34),(37)}{\leqslant}\ &\mathbb{E}\|x_k-x^*\|^2-2\gamma\left(-\frac{8B_G^2L_F^2}{\mu_f}T_0+\frac{3}{4}\mathbb{E}(f(x_k)-f^*)\right) \\
&+\gamma^2\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+16L_f\right)\mathbb{E}(f(x_k)-f^*) \\
&+\gamma^2\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+20L_f\right)\mathbb{E}(f(\tilde{x})-f^*) \\
\overset{(38)}{\leqslant}\ &\mathbb{E}\|x_k-x^*\|^2-2\gamma\left(-\frac{8B_G^2L_F^2}{\mu_f}\left(\frac{8B_G^2}{A}\mathbb{E}(\|\tilde{x}-x^*\|^2+\|x_k-x^*\|^2)\right)+\frac{3}{4}\mathbb{E}(f(x_k)-f^*)\right) \\
&+\gamma^2\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+16L_f\right)\mathbb{E}(f(x_k)-f^*) \\
&+\gamma^2\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+20L_f\right)\mathbb{E}(f(\tilde{x})-f^*) \\
\overset{(33)}{\leqslant}\ &\mathbb{E}\|x_k-x^*\|^2-2\gamma\left(-\frac{128B_G^4L_F^2}{\mu_f^2A}\mathbb{E}(f(\tilde{x})-f^*+f(x_k)-f^*)+\frac{3}{4}\mathbb{E}(f(x_k)-f^*)\right) \\
&+\gamma^2\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+16L_f\right)\mathbb{E}(f(x_k)-f^*) \\
&+\gamma^2\left(\frac{128}{\mu_f}\left(\frac{B_F^2L_G^2}{B}+\frac{B_G^4L_F^2}{A}\right)+20L_f\right)\mathbb{E}(f(\tilde{x})-f^*)
\end{aligned}$$

$$
\begin{aligned}
= \quad & \mathbb{E}\|x_k - x^*\|^2 \\
& - \left(\frac{3\gamma}{2} - \frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} - \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 16 L_f \right) \right) \mathbb{E}(f(x_k) - f^*) \\
& + \left(\frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} + \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 20 L_f \right) \right) \mathbb{E}(f(\tilde{x}) - f^*).
\end{aligned}
$$

Summing from $k = 0$ to $k = K - 1$, we obtain

$$
\begin{aligned}
\mathbb{E}\|x_K - x^*\|^2 \quad \leqslant \quad & \mathbb{E}\|\tilde{x} - x^*\|^2 \\
& - \left(\frac{3\gamma}{2} - \frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} - \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 16 L_f \right) \right) \sum_{k=0}^{K-1} \mathbb{E}(f(x_k) - f^*) \\
& + \left(\frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} + \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 20 L_f \right) \right) K\mathbb{E}(f(\tilde{x}) - f^*).
\end{aligned}
$$

Discarding the LHS and note that $\|\tilde{x} - x^*\|^2 \leqslant \frac{2}{\mu_f}(f(\tilde{x}) - f^*)$, we obtain

$$
\frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}(f(x_k) - f^*) \quad \leqslant \quad \frac{\frac{2}{K\mu_f} + \frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} + \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 20 L_f \right)}{\frac{3\gamma}{2} - \frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} - \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 16 L_f \right)} \mathbb{E}(f(\tilde{x}) - f^*),
$$

completing the proof. □

Proof to Corollary 2

Proof. To appropriately choose parameters γ, K, A, and B, the key is to ensure the coefficient $\frac{\beta_3}{\beta_4} < 1$ in Therom 2:

$$
\frac{\beta_3}{\beta_4} = \frac{\frac{2}{K\mu_f} + \frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} + \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 20 L_f \right)}{\frac{3\gamma}{2} - \frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} - \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 16 L_f \right)}.
$$

We choose A, B, and γ satisfying (39), (40), (41), and (42):

$$
\begin{aligned}
\frac{256\gamma B_G^4 L_F^2}{\mu_f^2 A} \quad & \leqslant \quad \frac{\gamma}{4} && (39) \\
\Rightarrow A \quad & \geqslant \quad \frac{1024 B_G^4 L_F^2}{\mu_f^2} \\
\gamma^2 \frac{128}{\mu_f} \frac{B_F^2 L_G^2}{B} \quad & \leqslant \quad \frac{\gamma}{16} && (40) \\
\Rightarrow B \quad & \geqslant \quad \gamma \frac{2048}{\mu_f} B_F^2 L_G^2 \\
\gamma^2 \frac{128}{\mu_f} \frac{B_G^4 L_F^2}{A} \quad & \leqslant \quad \frac{\gamma}{16} && (41)
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow A &\geqslant \gamma \frac{2048}{\mu_f} B_G^4 L_F^2 \\
20\gamma^2 L_f &\leqslant \frac{\gamma}{16} \qquad (42)\\
\Rightarrow \gamma &\leqslant \frac{1}{320 L_f}.
\end{aligned}
$$

Then we have the following bound on the coefficient

$$
\begin{aligned}
\nu &= \frac{\frac{2}{K\mu_f} + \frac{256\gamma B_G^4 L_F^2}{\mu_f A} + \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 20 L_f \right)}{\frac{3\gamma}{2} - \frac{256\gamma B_G^4 L_F^2}{\mu_f A} - \gamma^2 \left(\frac{128}{\mu_f} \left(\frac{B_F^2 L_G^2}{B} + \frac{B_G^4 L_F^2}{A} \right) + 16 L_f \right)} \\
&\leqslant \frac{\frac{2}{K\mu_f} + \frac{\gamma}{4} + \frac{3\gamma}{16}}{\frac{3\gamma}{2} - \frac{\gamma}{4} - \frac{3\gamma}{16}} \\
&= \frac{\frac{2}{K\mu_f} + \frac{7\gamma}{16}}{\frac{17\gamma}{16}} \\
&= \frac{32}{17K\mu_f\gamma} + \frac{7}{17}.
\end{aligned}
$$

We then choose K satisfying

$$
\frac{32}{17K\mu_f\gamma} \leqslant \frac{2}{17},
$$

which is equivalent to

$$
K \geqslant \frac{16}{\mu_f \gamma}.
$$

Thus choosing γ, A, and K appropriately in the following to satisfy all conditions derived above

$$
\begin{aligned}
\gamma &= \frac{1}{320 L_f} \\
K &\geqslant \frac{16}{\mu_f \gamma} = \frac{5120 L_f}{\mu_f} \\
A &\geqslant \max\left\{ \frac{1024 B_G^4 L_F^2}{\mu_f^2}, \gamma \frac{2048}{\mu_f} B_G^4 L_F^2 \right\} = \max\left\{ \frac{1024 B_G^4 L_F^2}{\mu_f^2}, \frac{32 B_G^4 L_F^2}{5\mu_f L_f} \right\} \\
B &\geqslant \gamma \frac{2048}{\mu_f} B_F^2 L_G^2 = \frac{32 B_F^2 L_G^2}{5 \mu_f L_f},
\end{aligned}
$$

we will obtain a 9/17 linear convergence rate with $\frac{\beta_3}{\beta_4} = \frac{9}{17}$ from Theorem 2. □