I. Proof of Theorem 1

To prove Theorem 1, we first rewrite $F\left(\tilde{\mathbf{w}}_{n+1}\right)-F\left(\tilde{\mathbf{w}}_{n}\right)$ based on Assumption 2, which is expressed as

$$\begin{aligned}&F\left(\tilde{\mathbf{w}}_{n+1}\right)-F\left(\mathbf{w}_{n}\right)\\&\leq \nabla F\left(\mathbf{w}_{n}\right)^{T}\left(\tilde{\mathbf{w}}_{n+1}-\mathbf{w}_{n}\right)+\frac{L}{2}\left\|\tilde{\mathbf{w}}_{n+1}-\mathbf{w}_{n}\right\|^{2}\\&=\underbrace{-\eta\mathbf{g}_{n}^{T}\hat{\mathbf{g}}_{n}}_{A_1}+\underbrace{\frac{L\eta^{2}}{2}\|\hat{\mathbf{g}}_{n}\|^{2}}_{A_2}.\end{aligned}\tag{1}$$

For A_1 in (1), We rewrite it as

$$\begin{aligned}A_1=&-\frac{\eta}{K}\sum_{k=1}^{K}\mathbf{g}_{n}^{T}\left(\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}\right)\\=&\frac{\eta}{2K}\sum_{k=1}^{K}\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}-\mathbf{g}_{n}\right\|^{2}\\&-\frac{\eta}{2}\|\mathbf{g}_{n}\|^{2}-\frac{\eta}{2K}\sum_{k=1}^{K}\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}\right\|^{2}.\end{aligned}\tag{2}$$

Then, by exploiting the Jensen's Inequality, the term in A_2 in (1) is bounded by

$$\begin{aligned}A_2=&\frac{L\eta^{2}}{2}\left\|\sum_{k=1}^{K}\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{Kq_{k,n}}\right\|^{2}\\\leq&\frac{L\eta^{2}}{2K}\sum_{k=1}^{K}\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}\right\|^{2}.\end{aligned}\tag{3}$$

Combining the results from (2) and (3), the expectation of (1) is expressed as

$$\begin{aligned}&\mathbb{E}(F\left(\tilde{\mathbf{w}}_{n+1}\right))-F\left(\mathbf{w}_{n}\right)\leq-\frac{\eta}{2}\|\mathbf{g}_{n}\|^{2}\\&-\frac{\eta(1-L\eta)}{2K}\underbrace{\sum_{k=1}^{K}\mathbb{E}\left[\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}\right\|^{2}\right]}_{B_1}\\&+\frac{\eta}{2K}\underbrace{\sum_{k=1}^{K}\mathbb{E}\left[\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n}))\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}-\mathbf{g}_{n}\right\|^{2}\right]}_{B_2}.\end{aligned}\tag{4}$$

For the term B_1 in (4), we have

$$\begin{aligned}B_1=&\sum_{k=1}^{K}\mathbb{E}\left[\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}\right\|^{2}\right]\\=&\sum_{k=1}^{K}\frac{1}{q_{k,n}}\mathbb{E}\left[\left\|s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})\right\|^{2}\right]\\=&\sum_{k=1}^{K}\left(\frac{p_{k,n}}{q_{k,n}}\mathbb{E}\left[\|Q(\mathbf{g}_{k,n})\|^{2}\right]+\frac{1-p_{k,n}}{q_{k,n}}\|\bar{\mathbf{g}}\|^{2}\right)\\\overset{(a)}{=}&\sum_{k=1}^{K}\left(\frac{p_{k,n}}{q_{k,n}}\mathbb{E}\left[\|Q(\mathbf{g}_{k,n})-\mathbf{g}_{k,n}\|^{2}\right]+\frac{p_{k,n}}{q_{k,n}}\|\mathbf{g}_{k,n}\|^{2}\right.\\&\left.+\frac{1-p_{k,n}}{q_{k,n}}\|\bar{\mathbf{g}}\|^{2}\right)\\\overset{(b)}{\leq}&\sum_{k=1}^{K}\left(\frac{p_{k,n}}{q_{k,n}}\delta_{k,n}^{2}+\frac{p_{k,n}}{q_{k,n}}\|\mathbf{g}_{k,n}\|^{2}+\frac{1-p_{k,n}}{q_{k,n}}\|\bar{\mathbf{g}}\|^{2}\right),\end{aligned}\tag{5}$$

where (a) and (b) is due to Lemma 1. Next, we express B_2 as as (7) at the top of the next page, where (a) uses

$$\begin{aligned}&\mathbb{E}\left[\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}}\right]\\&=p_{k,n}\mathbf{g}_{k,n}+(1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}.\end{aligned}\tag{6}$$

We further bound C_1 in (7) by

$$\begin{aligned}C_1=&\sum_{k=1}^{K}\|(1-p_{k,n})(s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}-\mathbf{g}_{k,n})+\mathbf{g}_{k,n}-\mathbf{g}_{n}\|^{2}\\\leq&2\sum_{k=1}^{K}(1-p_{k,n})^{2}\left\|s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}-\mathbf{g}_{k,n}\right\|^{2}\\&+2\sum_{k=1}^{K}\|\mathbf{g}_{k,n}-\mathbf{g}_{n}\|^{2}\\\overset{(a)}{\leq}&2\sum_{k=1}^{K}\left(\|\bar{\mathbf{g}}\|^{2}+\|\mathbf{g}_{k,n}\|^{2}-2\upsilon_{k,n}+\epsilon_{k,n}^{2}\right)\\&-4\sum_{k=1}^{K}p_{k,n}\left(\|\bar{\mathbf{g}}\|^{2}+\|\mathbf{g}_{k,n}\|^{2}-2\upsilon_{k,n}\right)\\&+2\sum_{k=1}^{K}p_{k,n}^{2}\left(\|\bar{\mathbf{g}}\|^{2}+\|\mathbf{g}_{k,n}\|^{2}-2\upsilon_{k,n}\right).\end{aligned}\tag{8}$$

where (a) is due to Assumption 3, and $\upsilon_{k,n}$ is defined as $\upsilon_{k,n}\triangleq\langle\mathbf{g}_{k,n},s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\rangle\geq 0$. Then, the second term in the right hand side (RHS) of (7), C_2, is rewritten as

$$\begin{aligned}C_2=&\sum_{k=1}^{K}(1-q_{k,n})\left\|p_{k,n}\mathbf{g}_{k,n}+(1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^{2}\\=&\sum_{k=1}^{K}\|\bar{\mathbf{g}}\|^{2}+2\sum_{k=1}^{K}p_{k,n}\left(\upsilon_{k,n}-\|\bar{\mathbf{g}}\|^{2}\right)\\&-\sum_{k=1}^{K}q_{k,n}\|\bar{\mathbf{g}}\|^{2}+2\sum_{k=1}^{K}p_{k,n}q_{k,n}\left(\|\bar{\mathbf{g}}\|^{2}-\upsilon_{k,n}\right)\\&+\sum_{k=1}^{K}p_{k,n}^{2}\left(\|\mathbf{g}_{k.n}\|^{2}+\|\bar{\mathbf{g}}\|^{2}-2\upsilon_{k,n}\right)\\&+\sum_{k=1}^{K}p_{k,n}^{2}q_{k,n}\left(2\upsilon_{k,n}-\|\mathbf{g}_{k,n}\|^{2}-\|\bar{\mathbf{g}}\|^{2}\right).\end{aligned}\tag{9}$$

For the third term in the RHS of (7), C_3, it is bounded by (10) at the next page, where (a) is due to Lemma 1.

We further rewrite the term E_1 in the RHS of (10) as

$$\begin{aligned}E_1=&\sum_{k=1}^{K}\frac{p_{k,n}}{q_{k,n}}\|\mathbf{g}_{k,n}\|^{2}-2\sum_{k=1}^{K}p_{k,n}\upsilon_{k,n}+\sum_{k=1}^{K}p_{k,n}q_{k,n}\|\bar{\mathbf{g}}\|^{2}\\&+2\sum_{k=1}^{K}p_{k,n}^{2}\left(\upsilon_{k,n}-\|\mathbf{g}_{k,n}\|^{2}\right)\end{aligned}$$

$$
\begin{aligned}
B_2 =&\sum_{k=1}^{K} \|p_{k,n}\mathbf{g}_{k,n} + (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}} - \mathbf{g}_n\|^2 \\
&+\sum_{k=1}^{K}\mathbb{E}\left[\left\|\frac{C(\mathbf{g}_{k,n})\cdot s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}} - p_{k,n}\mathbf{g}_{k,n} - (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^2\right] \\
=&\underbrace{\sum_{k=1}^{K}\|(1-p_{k,n})(s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}} - \mathbf{g}_{k,n}) + \mathbf{g}_{k,n} - \mathbf{g}_n\|^2}_{C_1} + \underbrace{\sum_{k=1}^{K}(1-q_{k,n})\|p_{k,n}\mathbf{g}_{k,n} + (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\|^2}_{C_2} \\
&+\underbrace{\sum_{k=1}^{K}q_{k,n}\mathbb{E}\left[\left\|\frac{s(\mathbf{g}_{k,n})\odot\hat{Q}_v(\mathbf{g}_{k,n})}{q_{k,n}} - p_{k,n}\mathbf{g}_{k,n} - (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^2\right]}_{C_3}, \qquad (7)
\end{aligned}
$$

$$
\begin{aligned}
C_3 =&\sum_{k=1}^{K}p_{k,n}q_{k,n}\mathbb{E}\left[\left\|\frac{Q(\mathbf{g}_{k,n})}{q_{k,n}} - p_{k,n}\mathbf{g}_{k,n} - (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^2\right] \\
&+\sum_{k=1}^{K}(1-p_{k,n})q_{k,n}\left\|\frac{s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}}{q_{k,n}} - p_{k,n}\mathbf{g}_{k,n} - (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^2 \\
=&\sum_{k=1}^{K}\frac{p_{k,n}}{q_{k,n}}\mathbb{E}\left[\|Q(\mathbf{g}_{k,n}) - \mathbf{g}_{k,n}\|^2\right] + \sum_{k=1}^{K}p_{k,n}q_{k,n}\left\|\frac{1-p_{k,n}q_{k,n}}{q_{k,n}}\mathbf{g}_{k,n} - (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^2 \\
&+\sum_{k=1}^{K}(1-p_{k,n})q_{k,n}\left\|\frac{1-(1-p_{k,n})q_{k,n}}{q_{k,n}}s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}} - p_{k,n}\mathbf{g}_{k,n}\right\|^2 \\
\overset{\text{(a)}}{\leq}&\underbrace{\sum_{k=1}^{K}p_{k,n}q_{k,n}\left\|\frac{1-p_{k,n}q_{k,n}}{q_{k,n}}\mathbf{g}_{k,n} - (1-p_{k,n})s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right\|^2}_{E_1} \\
&+\underbrace{\sum_{k=1}^{K}(1-p_{k,n})q_{k,n}\left\|\frac{1-(1-p_{k,n})q_{k,n}}{q_{k,n}}s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}} - p_{k,n}\mathbf{g}_{k,n}\right\|^2}_{E_2} + \sum_{k=1}^{K}\frac{p_{k,n}}{q_{k,n}}\delta_{k,n}^2. \qquad (10)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k=1}^{K}p_{k,n}^2q_{k,n}\left(2\upsilon_{k,n} - 2\|\bar{\mathbf{g}}\|^2\right) \\
&+\sum_{k=1}^{K}p_{k,n}^3q_{k,n}\left(\|\mathbf{g}_{k,n}\|^2 + \|\bar{\mathbf{g}}\|^2 - 2\upsilon_{k,n}\right). \qquad (11)
\end{aligned}
$$

Similarly, the term E_2 is expressed as

$$
\begin{aligned}
E_2 =&\sum_{k=1}^{K}(1-p_{k,n})q_{k,n}\left\|\frac{1-(1-p_{k,n})q_{k,n}}{q_{k,n}}s(\mathbf{g}_{k,n})\odot\bar{\mathbf{g}}\right. \\
&\left.-p_{k,n}\mathbf{g}_{k,n}\right\|^2 \\
=&-2K\|\bar{\mathbf{g}}\|^2 + \sum_{k=1}^{K}\frac{1-p_{k,n}}{q_{k,n}}\|\bar{\mathbf{g}}\|^2 + \sum_{k=1}^{K}q_{k,n}\|\bar{\mathbf{g}}\|^2 \\
&+2\sum_{k=1}^{K}p_{k,n}\left(2\|\bar{\mathbf{g}}\|^2 - \upsilon_{k,n}\right) + 2\sum_{k=1}^{K}p_{k,n}^2\left(\upsilon_{k,n} - \|\bar{\mathbf{g}}\|^2\right) \\
&+\sum_{k=1}^{K}p_{k,n}q_{k,n}\left(2\upsilon_{k,n} - 3\|\bar{\mathbf{g}}\|^2\right) \\
&+\sum_{k=1}^{K}p_{k,n}^2q_{k,n}\left(\|\mathbf{g}_{k,n}\|^2 - 4\upsilon_{k,n} + 3\|\bar{\mathbf{g}}\|^2\right) \\
&+\sum_{k=1}^{K}p_{k,n}^3q_{k,n}\left(2\upsilon_{k,n} - \|\mathbf{g}_{k,n}\|^2 - \|\bar{\mathbf{g}}\|^2\right) \qquad (12)
\end{aligned}
$$

Combining all the results in (4)-(12), it yields

$$
\begin{aligned}
&\mathbb{E}(F(\tilde{\mathbf{w}}_{n+1})) - F(\mathbf{w}_n) \\
&\leq -\frac{\eta}{2}\|\mathbf{g}_n\|^2 + \frac{\eta}{2}\|\bar{\mathbf{g}}\|^2 + \frac{\eta}{K}\sum_{k=1}^{K}(\|\mathbf{g}_{k,n}\|^2 + \epsilon_{k,n}^2 - 2\upsilon_{k,n}) \\
&+\frac{\eta}{K}\sum_{k=1}^{K}p_{k,n}\left(-\|\bar{\mathbf{g}}\|^2 - 2\|\mathbf{g}_{k,n}\|^2 + 3\upsilon_{k,n}\right)
\end{aligned}
$$

$$G'(\alpha_{k,n}) = A_{k,n} \exp\left(\frac{H_v(\beta_{k,n})}{1-\alpha_{k,n}}\right) \frac{H_v(\beta_{k,n})}{(1-\alpha_{k,n})^2} + B_{k,n} \exp\left(\frac{2H_v(\beta_{k,n})}{1-\alpha_{k,n}}\right) \frac{2H_v(\beta_{k,n})}{(1-\alpha_{k,n})^2}$$
$$+ C_{k,n} \exp\left(\frac{H_v(\beta_{k,n})}{1-\alpha_{k,n}} - \frac{H_s(\beta_{k,n})}{\alpha_{k,n}}\right) \left(\frac{H_v(\beta_{k,n})}{(1-\alpha_{k,n})^2} + \frac{H_s(\beta_{k,n})}{\alpha_{k,n}^2}\right) + D_{k,n} \exp\left(-\frac{H_s(\beta_{k,n})}{\alpha_{k,n}}\right) \frac{H_s(\beta_{k,n})}{\alpha_{k,n}^2}. \quad (15)$$

$$+ \frac{\eta}{2K} \sum_{k=1}^{K} p_{k,n}^2 \left(\|\bar{\mathbf{g}}\|^2 + \|\mathbf{g}_{k,n}\|^2 - 2\upsilon_{k,n}\right)$$
$$+ \frac{L\eta^2}{2K} \sum_{k=1}^{K} \frac{p_{k,n}}{q_{k,n}} (\delta_{k,n}^2 + \|\mathbf{g}_{k,n}\|^2 - \|\bar{\mathbf{g}}\|^2)$$
$$+ \frac{L\eta^2}{2K} \sum_{k=1}^{K} \frac{1}{q_{k,n}} \|\bar{\mathbf{g}}\|^2 . \quad (13)$$

The proof is complete.

II. Proof of Lemma 2

For $\alpha_{k,n} \in (0,1)$, due to $H_s(\beta_{k,n}) < 0$ and $H_v(\beta_{k,n}) < 0$, we can derive that

$$\begin{aligned}
&\lim_{\alpha_{k,n}\to 0^+} G'(\alpha_{k,n}) \\
&= \lim_{\alpha_{k,n}\to 0^+} \Bigg(C_{k,n} \exp\left(-\frac{H_s(\beta_{k,n})}{\alpha_{k,n}}\right) \frac{H_s(\beta_{k,n})}{\alpha_{k,n}^2} \\
&\quad + D_{k,n} \exp\left(-\frac{H_s(\beta_{k,n})}{\alpha_{k,n}}\right) \frac{H_s(\beta_{k,n})}{\alpha_{k,n}^2} \Bigg) \\
&< 0,
\end{aligned} \quad (14)$$

where $G'(\alpha_{k,n})$ is the first-order derivative of $G(\alpha_{k,n})$ with respect to $\alpha_{k,n}$ given by (15) at the top of this page.

Therefore, the optimal solution $\alpha_{k,n}^*$ can be discussed in the following cases:

(1) If there exists $0 < x_1 < \cdots < x_i < 1$, which satisfies $\forall 1 \le j \le i$, $G'(x_j) = 0$, $G(\alpha_{k,n}, \beta_{k,n})$ takes a minimal value at $\alpha_{k,n} = \alpha_{k,n}^*$, where

$$\alpha_{k,n}^* = \underset{\alpha_{k,n} \in \{x_1,\ldots,x_i,1\}}{\arg\min} G(\alpha_{k,n}, \beta_{k,n}) \quad (16)$$

(2) If $G'(\alpha_{k,n}) = 0$ contains no solution x between (0,1), we can conclude that $G'(\alpha_{k,n}) < 0$ for $\forall \alpha_{k,n} \in (0,1)$. Then $G(\alpha_{k,n}, \beta_{k,n})$ takes a minimal value at $\alpha_{k,n}^* = 1$.

By combining (1) and (2), we complete the proof.