New Proof

Anonymous Author(s)
Affiliation
Address
email

Contents

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

1 Notations.

In this situation, assume that for each i, $f_i(x)$ is L-smooth.

$\mathbf{x}^{(k)} = [(x_1^{(k)})^\top; (x_2^{(k)})^\top; \cdots; (x_n^{(k)})^\top] \in \mathbb{R}^{n\times d}$ is the matrix composed of the variables from all the nodes.

$\nabla \mathbf{f}^{(k)} = \nabla F(\mathbf{x}^{(k)}) = [\nabla F_1(x_1^{(k)})^\top; \cdots; \nabla F_n(x_n^{(k)})^\top] \in \mathbb{R}^{n\times d}$ is the matrix composed of the true gradients from all the nodes.

$\mathbf{g}^{(k)} = \nabla F(\mathbf{x}^{(k)}; \boldsymbol{\xi}^{(k)}) = [\nabla F_1(x_1^{(k)}; \xi_1^{(k)})^\top; \cdots; \nabla F_n(x_n^{(k)}; \xi_n^{(k)})^\top] \in \mathbb{R}^{n\times d}$ is the matrix composed of the stochastic gradients from all the nodes.

$\nabla \mathbf{f}(w^{(k)}) = \nabla F(\mathbf{w}^{(k)}) = [\nabla F_1(w^{(k)})^\top; \cdots; \nabla F_n(w^{(k)})^\top] \in \mathbb{R}^{n\times d}$ is the matrix composed of the values of the true gradient functions of all nodes evaluated at $w^{(k)}$.

$w^{(k)} = \pi_A^T \mathbf{x}^{(k)},\ \mathbf{w}^{(k)} = A_\infty \mathbf{x}^{(k)}$

$\bar{x} = \frac{1}{n}\mathbb{1}_n^T \mathbf{x},\ \bar{\mathbf{x}} = \frac{1}{n}\mathbb{1}_n \mathbb{1}_n^T \mathbf{x}$

$\Delta_x^{(k)} = \mathbf{x}^{(k)} - \mathbf{w}^{(k)}$

$\Delta_y^{(k)} = \mathbf{y}^{(k)} - B_\infty \mathbf{y}^{(k)} = (I - B_\infty)\mathbf{y}^{(k)}$

$\Delta_g^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$

$\bar{y} = \frac{1}{n}\mathbb{1}_n^T \mathbf{y},\ \bar{\mathbf{y}} = \frac{1}{n}\mathbb{1}_n \mathbb{1}_n^T \mathbf{y}$

$\overline{\nabla f}^{(k)} = \frac{1}{n}\mathbb{1}_n^T \nabla \mathbf{f}^{(k)}$

2 Analysis: Basic

2.1 Rolling Sum Lemma

Lemma 1 (Rolling Sum Lemma). *For a rolling sum using primitive and row-stochastic matrix $A \in \mathbb{R}^{n\times n}$, we have the following estimation:*

$$\sum_{k=0}^{T} \| \sum_{i=0}^{k} (A^{k-i} - A_\infty)\Delta^{(i)} \|_F^2 \le s_A^2 \sum_{i=0}^{T} \|\Delta^{(i)}\|_F^2, \tag{1}$$

where $\Delta^{(i)} \in \mathbb{R}^{n\times d}$ are arbitrary matrices, and s_A is defined by:

$$s_A := \max_{k\ge 0} \|A^k - A_\infty\|_2 \cdot \frac{1 + \frac{1}{2}\ln(\kappa(\pi_A))}{1 - \beta_A} \le \sqrt{n} \cdot \frac{1 + \frac{1}{2}\ln(\kappa(\pi_A))}{1 - \beta_A}. \tag{2}$$

Inequality (1) *also holds when we replace every A with column-stochastic B, where s_B is defined by:*

$$s_B := \max_{k\ge 0} \|B^k - B_\infty\|_2 \cdot \frac{1 + \frac{1}{2}\ln(\kappa(\pi_B))}{1 - \beta_B} \le \sqrt{n} \cdot \frac{2 + \ln(\kappa(\pi_B))}{1 - \beta_B}. \tag{3}$$

Proof. First, we prove that

$$\|A^i - A_\infty\|_2 \le \sqrt{\kappa(\pi_A)}\beta_A^i, \forall i \ge 0. \tag{4}$$

Notice that $\beta_A := \|A - A_\infty\|_{\pi_A}$ and

$$\|A^i - A_\infty\|_{\pi_A} = \|(A - A_\infty)^i\|_{\pi_A} \le \|A - A_\infty\|_{\pi_A}^i = \beta_A^i,$$

we have

$$\|(A^{k-i} - A_\infty)v\| = \|\Pi_A^{-1/2}(A^{k-i} - A_\infty)v\|_{\pi_A} \le \sqrt{\underline{\pi_A}}\beta_A^{k-i}\|v\|_{\pi_A} \le \sqrt{\kappa(\pi_A)}\beta_A^{k-i}\|v\|,$$

which proves (4).

Second, we want to prove that for all $k \geq 0$, we have

$$\sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 \leq M_A \cdot \frac{1 + \frac{1}{2}\ln(\kappa(\pi_A))}{1 - \beta_A} =: s_A. \tag{5}$$

Towards this end, we define $M_A := \max_{k \geq 0} \|A^k - A_\infty\|_2$. M_A is well-defined because of (4). We also define $p = \max\left\{\frac{\ln(\sqrt{\kappa(\pi_A)}) - \ln(M_A)}{-\ln(\beta_A)}, 0\right\}$, then we can verify that $\|A^i - A_\infty\|_2 \leq \min\{M_A, M_A\beta_A^{i-p}\}$. With this inequality, we can bound $\sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2$ as follows:

$$\begin{aligned}
\sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 = & \sum_{i=0}^{\min\{\lfloor p \rfloor, k\}} \|A^i - A_\infty\|_2 + \sum_{i=\min\{\lfloor p \rfloor, k\}+1}^{k} \|A^i - A_\infty\|_2 \\
\leq & \sum_{i=0}^{\min\{\lfloor p \rfloor, k\}} M_A + \sum_{i=\min\{\lfloor p \rfloor, k\}+1}^{k} M_A \beta_A^{i-p} \\
\leq & M_A \cdot (1 + \min\{\lfloor p \rfloor, k\}) + M_A \cdot \frac{1}{1-\beta_A} \beta_A^{\min\{\lfloor p \rfloor, k\}+1-p}.
\end{aligned} \tag{6}$$

If $p = 0$, (6) is simplified to $\sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 \leq M_A \cdot \frac{1}{1-\beta_A}$ and (5) is naturally satisfied. If $p > 0$, let $x = \min\{\lfloor p \rfloor, k\} + 1 - p \in [0, 1)$, (5) is simplified to

$$\sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 \leq M_A (x + p + \frac{\beta_A^x}{1-\beta_A}) \leq M_A (p + \frac{1}{1-\beta_A}).$$

Noting that $p \leq \frac{\frac{1}{2}\ln(\kappa(\pi_A))}{1-\beta_A}$, we finish the proof of (5).

Finally, to obtain (1), we use Jensen's inequality. For positive numbers $a_i, i \in [k]$ satisfying $\sum_{i=0}^{k} a_i = 1$, we have

$$\begin{aligned}
& \|\sum_{i=0}^{k} (A^{k-i} - A_\infty)\Delta^{(i)}\|_F^2 = \|\sum_{i=0}^{k} a_{k-i} \cdot a_{k-i}^{-1} (A^{k-i} - A_\infty)\Delta^{(i)}\|_F^2 \\
\leq & \sum_{i=0}^{k} a_{k-i} \|a_{k-i}^{-1} (A^{k-i} - A_\infty)\Delta^{(i)}\|_F^2 \leq \sum_{i=0}^{k} a_{k-i}^{-1} \|A^{k-i} - A_\infty\|_2^2 \|\Delta^{(i)}\|_F^2.
\end{aligned} \tag{7}$$

By choosing $a_{k-i} = (\sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2)^{-1} \|A^{k-i} - A_\infty\|_2$ in (7), we obtain that

$$\|\sum_{i=0}^{k} (A^{k-i} - A_\infty)\Delta^{(i)}\|_F^2 \leq \sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 \cdot \sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 \|\Delta^{(i)}\|_F^2. \tag{8}$$

By summing up (8) from $k = 0$ to T, we obtain that

$$\begin{aligned}
& \sum_{k=0}^{T} \|\sum_{i=0}^{k} (A^{k-i} - A_\infty)\Delta^{(i)}\|_F^2 \leq s_A \sum_{k=0}^{T} \sum_{i=0}^{k} \|A^{k-i} - A_\infty\|_2 \|\Delta^{(i)}\|_F^2 \\
\leq & s_A \sum_{i=0}^{T} (\sum_{k=i}^{T} \|A^{k-i} - A_\infty\|_2) \|\Delta^{(i)}\|_F^2 \leq s_A^2 \sum_{i=0}^{T} \|\Delta^{(i)}\|_F^2,
\end{aligned}$$

which finishes the proof of this lemma. The proof can be applied in the same way when B is column-stochastic.

□

2.2 Basic Transformation

The following statement holds for all $k \geq 0$.

1. $\bar{y}^{(k)} = \bar{g}^{(k)}, \forall k \geq 0$.

2. $\Delta_x^{(k+1)} = \mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)} = -\alpha \sum_{i=0}^{k} (A^{k-i} - A_\infty)\mathbf{y}^{(i)}$.

3. $\sum_{i=0}^{m-1} \mathbf{y}^{(k+i)} = \sum_{i=0}^{m-1} B^i \mathbf{y}^{(k)} + \sum_{i=0}^{m-1} B^{m-1-i}(\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)})$.

4. $\Delta_y^{(k+1)} = (B - B_\infty)\Delta_y^{(k)} + (B - B_\infty)(I - B_\infty)\Delta_g^{(k)}$.

2.3 Technical Lemmas

Lemma 2. *The gradient consensus error can be written as the following rolling sum:*

$$\begin{aligned}\|\Delta_y^{(k+1)}\|_F^2 = &\sum_{i=0}^{k} \|(B - B_\infty)^{k-i}(I - B_\infty)\Delta_g^{(i)}\|_F^2 \\ &+ 2\sum_{i=0}^{k} \left\langle (B - B_\infty)^{k-i+1}\Delta_y^{(i)}, (B - B_\infty)^{k-i}(I - B_\infty)\Delta_g^{(i)} \right\rangle.\end{aligned}$$

Proof. Taking norm on both sides of $\Delta_y^{(k+1)} = (B - B_\infty)\Delta_y^{(k)} + (B - B_\infty)(I - B_\infty)\Delta_g^{(k)}$, we obtain that

$$\begin{aligned}\|\Delta_y^{(k+1)}\|_F^2 = \|(B - B_\infty)\Delta_y^{(k)}\|_F^2 + 2\left\langle (B - B_\infty)\Delta_y^{(k)}, (B - B_\infty)(I - B_\infty)\mathbf{g}^{(k)} \right\rangle \\ + \|(B - B_\infty)(I - B_\infty)\mathbf{g}^{(k)}\|_F^2.\end{aligned}$$

We can unfold the term $\|(B - B_\infty)\Delta_y^{(k)}\|_F^2$ in the same manner. By repeating the unfolding process from k to 0, we obtain the lemma. ☐

Lemma 3.

$$\begin{aligned}&\sum_{k=0}^{T}\sum_{i=0}^{k} \mathbb{E}\left[\|(B^{k-i} - B_\infty)\Delta_g^{(i)}\|_F^2\right] \\ \leq &6n\sigma^2(T+1)s_B M_B + 18 s_B^2 L^2 \sum_{k=0}^{T} \mathbb{E}\left[\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|_F^2\right] + 9\alpha^2 s_B^2 L^2 \sum_{k=0}^{T} \mathbb{E}\left[\|A_\infty \mathbf{y}^{(k)}\|_F^2\right]\end{aligned}$$

Proof. Consider $\mathbb{E}\left[\|(B^{k-i} - B_\infty)\Delta_g^{(i)}\|^2\right]$, we have that

$$\begin{aligned}&\mathbb{E}\left[\|(B^{k-i} - B_\infty)\Delta_g^{(i)}\|^2\right] \\ \leq &3\mathbb{E}\left[\|(B^{k-i} - B_\infty)(\mathbf{g}^{(i+1)} - \nabla f(\mathbf{x}^{(i+1)}))\|^2\right] + 3\mathbb{E}\left[\|(B^{k-i} - B_\infty)(\nabla f(\mathbf{x}^{(i+1)}) - \nabla f(\mathbf{x}^{(i)}))\|^2\right] \\ &+3\mathbb{E}\left[\|(B^{k-i} - B_\infty)(\mathbf{g}^{(i)} - \nabla f(\mathbf{x}^{(i)}))\|^2\right] \\ \leq &6n\sigma^2\|B^{k-i} - B_\infty\|^2 + 3\mathbb{E}\left[\|(B^{k-i} - B_\infty)(\nabla f(\mathbf{x}^{(i+1)}) - \nabla f(\mathbf{x}^{(i)}))\|^2\right]\end{aligned}$$

For the first part, we have that

$$\sum_{k=0}^{T}\sum_{i=0}^{k} 6n\sigma^2\|B^{k-i} - B_\infty\|^2 \leq 6n\sigma^2 \sum_{k=0}^{T} M_B \sum_{i=0}^{k} \|B^{k-i} - B_\infty\| \leq 6n\sigma^2 \sum_{k=0}^{T} M_B s_B = 6n\sigma^2(T+1)s_B M_B$$

For the second part, by applying Lemma 1 on $\sum_{k=0}^{T}\sum_{i=0}^{k} 3\mathbb{E}\left[\|(B^{k-i} - B_\infty)(\nabla f(\mathbf{x}^{(i+1)}) - \nabla f(\mathbf{x}^{(i)}))\|^2\right]$, we obtain that

$$\sum_{k=0}^{T}\sum_{i=0}^{k} 3\mathbb{E}\left[\|(B^{k-i} - B_\infty)(\nabla f(\mathbf{x}^{(i+1)}) - \nabla f(\mathbf{x}^{(i)}))\|^2\right] \leq 3s_B^2 \sum_{k=0}^{T} \mathbb{E}\left[\|\nabla f(\mathbf{x}^{(i+1)}) - \nabla f(\mathbf{x}^{(i)})\|_F^2\right]$$

Note that

$$\nabla f(\mathbf{x}^{(i+1)}) - \nabla f(\mathbf{x}^{(i)}) = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{w}^{(k+1)}) + \nabla f(\mathbf{w}^{(k+1)}) - \nabla f(\mathbf{w}^{(k)}) + \nabla f(\mathbf{w}^{(k)}) - \nabla f(\mathbf{x}^{(k)})$$

we can apply Cauchy's inequality and obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\|\nabla f(\mathbf{x}^{(i+1)})-\nabla f(\mathbf{x}^{(i)})\|_F^2\right]\\
\leq&3\mathbb{E}\left[\|\nabla f(\mathbf{x}^{(k+1)})-\nabla f(\mathbf{w}^{(k+1)})\|_F^2\right]+3\mathbb{E}\left[\|\nabla f(\mathbf{w}^{(k+1)})-\nabla f(\mathbf{w}^{(k)})\|_F^2\right]+3\mathbb{E}\left[\|\nabla f(\mathbf{w}^{(k)})-\nabla f(\mathbf{x}^{(k)})\|_F^2\right]\\
\leq&3L^2\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|_F^2+3L^2\|\mathbf{x}^{(k)}-\mathbf{w}^{(k)}\|_F^2+3\alpha^2L^2\mathbb{E}\left[\|A_\infty\mathbf{y}^{(k)}\|_F^2\right]
\end{aligned}
$$

So we obtain the lemma

$$
\begin{aligned}
&\sum_{k=0}^{T}\sum_{i=0}^{k}\mathbb{E}\left[\|(B^{k-i}-B_\infty)\Delta_g^{(i)}\|_F^2\right]\\
\leq&6n\sigma^2(T+1)s_BM_B+18s_B^2L^2\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|_F^2\right]+9\alpha^2s_B^2L^2\sum_{k=0}^{T}\mathbb{E}\left[\|A_\infty\mathbf{y}^{(k)}\|_F^2\right]
\end{aligned}
$$

□

Lemma 4.

$$
\begin{aligned}
&\sum_{i=0}^{k}\mathbb{E}\left[\left\langle(B^{k-i+1}-B_\infty)\Delta_y^{(i)},(B^{k-i}-B_\infty)\Delta_g^{(i)}\right\rangle\right]\\
\leq&(0.5\alpha\eta_1^{-1}+\eta_2^{-1})L\sum_{i=0}^{k}b_{k-i}\mathbb{E}\left[\|\Delta_y^{(i)}\|\right]+0.5\eta_1\alpha L\sum_{i=0}^{k}b_{k-i}\mathbb{E}\left[\|A_\infty\mathbf{y}^{(i)}\|\right]\\
&+0.5\eta_2L\sum_{i=0}^{k}b_{k-i}\mathbb{E}\left[\|\mathbf{x}^{(i+1)}-\mathbf{w}^{(i+1)}\|\right]+0.5\eta_2L\sum_{i=0}^{k}b_{k-i}\mathbb{E}\left[\|\mathbf{x}^{(i)}-\mathbf{w}^{(i)}\|_F\right]+n\sigma^2\sum_{i=0}^{k}b_{k-i}
\end{aligned}
$$

Proof. Notice that

$$
\mathbb{E}\left[\Delta_g^{(i)}|\mathcal{F}^{(i)}\right]=\mathbb{E}\left[(\nabla f^{(i+1)}-\nabla f^{(i)})+(\nabla f^{(i)}-\mathbf{g}^{(i)})|\mathcal{F}^{(i)}\right]
$$

and the basic transformation $(B-B_\infty)^{k-i}(I-B_\infty)=(B^{k-i}-B_\infty)(I-B_\infty)=B^{k-i}-B_\infty$, the term $\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}\Delta_g^{(i)}\right\rangle\right]$ can be decomposed to two terms of inner product.

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}(I-B_\infty)\Delta_g^{(i)}\right\rangle\right]\\
=&\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}(\nabla f^{(i+1)}-\nabla f^{(i)})\right\rangle\right]\\
&+\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}(\nabla f^{(i)}-\mathbf{g}^{(i)})\right\rangle\right]
\end{aligned}
$$

The first term is $\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}(\nabla f^{(i+1)}-\nabla f^{(i)})\right\rangle\right]$, which can be bounded by the Cauchy-Schwarz inequality as follows

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}(\nabla f^{(i+1)}-\nabla f^{(i)})\right\rangle\right]\\
\leq&L\|(B-B_\infty)^{k-i+1}\|_2\|(B-B_\infty)^{k-i}\|\mathbb{E}\left[\|\Delta_y^{(i)}\|\|\mathbf{x}^{(i+1)}-\mathbf{x}^{(i)}\|\right]
\end{aligned}
\tag{9}
$$

Let $b_{k-i}=\|(B-B_\infty)^{k-i+1}\|_2\|(B-B_\infty)^{k-i}\|_2$. By further using triangle inequality on the relation $\mathbf{x}^{(i+1)}-\mathbf{x}^{(i)}=\mathbf{x}^{(i+1)}-\mathbf{w}^{(i+1)}+\mathbf{w}^{(i+1)}-\mathbf{w}^{(i)}+\mathbf{w}^{(i)}-\mathbf{x}^{(i)}$, we can bound $\|\mathbf{x}^{(i+1)}-\mathbf{x}^{(i)}\|$ in 9 as:

$$
\|\mathbf{x}^{(i+1)}-\mathbf{x}^{(i)}\|\leq\|\mathbf{x}^{(i+1)}-\mathbf{w}^{(i+1)}\|+\alpha\|A_\infty\mathbf{y}^{(i)}\|+\|\mathbf{x}^{(i)}-\mathbf{w}^{(i)}\|
$$

so we obtain that

$$
\mathbb{E}\left[\left\langle(B-B_\infty)^{k-i+1}\Delta_y^{(i)},(B-B_\infty)^{k-i}(\nabla f^{(i+1)}-\nabla f^{(i)})\right\rangle\right]
\tag{10}
$$

$$
\begin{aligned}
&\leq \alpha L b_{k-i} \mathbb{E}\left[\|A_\infty \mathbf{y}^{(i)}\| \|\Delta_y^{(i)}\|\right] + L b_{k-i} \mathbb{E}\left[\|\mathbf{x}^{(i+1)} - \mathbf{w}^{(i+1)}\| \|\Delta_y^{(i)}\|\right] \\
&\quad + L b_{k-i} \mathbb{E}\left[\|\mathbf{x}^{(i)} - \mathbf{w}^{(i)}\| \|\Delta_y^{(i)}\|\right]
\end{aligned}
$$

By Young inequality, we can further bound 10 as

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle (B - B_\infty)^{k-i+1} \Delta_y^{(i)}, (B - B_\infty)^{k-i} (\nabla f^{(i+1)} - \nabla f^{(i)}) \right\rangle\right] \\
\leq& 0.5 L b_{k-i} (\alpha \eta_1^{-1} + 2\eta_2^{-1}) \mathbb{E}\left[\|\Delta_y^{(i)}\|\right] + 0.5 \eta_1 \alpha L b_{k-i} \mathbb{E}\left[\|A_\infty \mathbf{y}^{(i)}\|\right] \\
&+ 0.5 \eta_2 L b_{k-i} \mathbb{E}\left[\|\mathbf{x}^{(i+1)} - \mathbf{w}^{(i+1)}\|\right] + 0.5 \eta_2 L b_{k-i} \mathbb{E}\left[\|\mathbf{x}^{(i)} - \mathbf{w}^{(i)}\|\right]
\end{aligned} \tag{11}
$$

For the second term decomposed from $\mathbb{E}\left[\left\langle (B - B_\infty)^{k-i+1} \Delta_y^{(i)}, (B - B_\infty)^{k-i} (I - B_\infty) \Delta_g^{(i)} \right\rangle\right]$, which is $\mathbb{E}\left[\left\langle (B - B_\infty)^{k-i+1} \Delta_y^{(i)}, (B - B_\infty)^{k-i} (\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right]$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle (B - B_\infty)^{k-i+1} \Delta_y^{(i)}, (B - B_\infty)^{k-i} (\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right] \\
=&\mathbb{E}\left[\left\langle (B^{k-i+1} - B_\infty)(I - B_\infty) \mathbf{y}^{(i)}, (B^{k-i} - B_\infty)(\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right] \\
=&\mathbb{E}\left[\left\langle (B^{k-i+1} - B_\infty)(B \mathbf{y}^{(i-1)} + \mathbf{g}^{(i)} - \mathbf{g}^{(i-1)}), (B^{k-i} - B_\infty)(\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right]
\end{aligned}
$$

Since $\mathbf{y}^{(i-1)}, \mathbf{g}^{(i-1)}$ and $\nabla f^{(i)}$ are $\mathcal{F}^{(i-1)}$-measurable, $\mathbb{E}\left[\nabla f^{(l)} - \mathbf{g}^{(l)} | \mathcal{F}^{(l-1)}\right] = 0$. Therefore, we can further obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle (B^{k-i+1} - B_\infty) \Delta_y^{(i)}, (B^{k-i} - B_\infty)(\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right] \\
=&\mathbb{E}\left[\left\langle (B^{k-i+1} - B_\infty)(\mathbf{g}^{(i)} - \nabla f^{(i)}), (B^{k-i} - B_\infty)(\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right]
\end{aligned}
$$

The above expression can be reduced to

$$
\begin{aligned}
&\mathbb{E}\left[\left\langle (B^{k-i+1} - B_\infty) \Delta_y^{(i)}, (B^{k-i} - B_\infty)(\nabla f^{(i)} - \mathbf{g}^{(i)}) \right\rangle\right] \\
=&\mathbb{E}\left[\operatorname{tr}\left((\mathbf{g}^{(i)} - \nabla f^{(i)})^\top \operatorname{diag}((B_\infty - B^{k-i+1})^\top (B^{k-i} - B_\infty))(\mathbf{g}^{(i)} - \nabla f^{(i)}) \right)\right] \\
\leq& \sigma^2 \sum_{p=1}^n \left| \sum_{q=1}^n (B_\infty - B^{k-i+1})_{qp} (B^{k-i} - B_\infty)_{qp} \right| \\
\leq& \sigma^2 \sum_{p=1}^n \sqrt{ \sum_{q=1}^n (B_\infty - B^{k-i+1})_{qp}^2 \sum_{q=1}^n (B^{k-i} - B_\infty)_{qp}^2 } \\
\leq& \sigma^2 \|B_\infty - B^{k-i+1}\| \cdot \|B^{k-i} - B_\infty\| \leq n \sigma^2 b_{k-i}
\end{aligned} \tag{12}
$$

Combine 11 and 12, we obtain the lemma. □

Since $\sum_{k=0}^T \sum_{l=0}^k c_{k-l} \|\Delta^{(l)}\|_F^2 = \sum_{l=0}^T \|\Delta^{(l)}\|_F^2 \sum_{k=l}^T c_{k-l}$, next we give a brief discussion of the size of $\sum_{k=l}^T c_{k-l}$.

Lemma 5. *For $b_{k-l} := \|B^{k-l} - B_\infty\|_2 \|B^{k-l+1} - B_\infty\|_2$, we have the following inequality:*

$$
\sum_{k=l}^T b_{k-l} \leq M_B^2 \frac{1 + \ln(\kappa(\pi_B))}{1 - \beta_B^2} \leq 2 M_B s_B \tag{13}
$$

Proof. By definition of $M_B := \max_{i \geq 0} \{\|B^i - B_\infty\|_2\}$, we have $b_{k-l} \leq M_B^2$. Besides, alike to (4), we have $\|B^i - B_\infty\|_2 \leq \sqrt{\kappa(\pi_B)} \beta_B^i$. Thus, by defining $p = \max\left\{ \frac{\ln(\kappa(\pi_B)) - 2\ln(M_B)}{-\ln(\beta_B)}, 0 \right\}$, we can verify that $b_i \le \min M_B^2, M_B^2\beta_B^{2i+1-p}, \forall i \ge 0$. With this inequality, we can bound $\sum_{k=l}^T b_{k-l}$ as follows:

$$\begin{aligned}\sum_{k=l}^{T} b_{k-l} \le & \sum_{i=0}^{\min\{\lfloor\frac{p-1}{2}\rfloor, i\}} M_B^2 + \sum_{i=\min\{\lfloor\frac{p-1}{2}\rfloor, i\}+1}^{T-l} M_B^2\beta_B^{2i+1-p} \\ \le & M_B^2 \cdot (1+\min\{\lfloor\frac{p-1}{2}\rfloor, i\}) + M_B^2 \cdot \frac{1}{1-\beta_B^2}\beta_B^{2+2\lfloor\frac{p-1}{2}\rfloor-p} \end{aligned} \tag{14}$$

Then, we can repeat the discussion of (6) in Lemma 1 and obtain this lemma. □

Lemma 6.

$$\begin{aligned}&\sum_{k=0}^{T}\sum_{i=0}^{k}\mathbb{E}\left[\left\langle (B^{k-i+1}-B_\infty)\Delta_y^{(i)}, (B^{k-i}-B_\infty)\Delta_g^{(i)}\right\rangle\right] \\ \le & M_B s_B(\alpha\eta_1^{-1}+2\eta_2^{-1})L\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|\right] + M_B s_B\eta_1\alpha L\sum_{k=0}^{T}\mathbb{E}\left[\|A_\infty\mathbf{y}^{(k)}\|\right] \\ & + 2M_B s_B\eta_2 L\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|\right] + 2M_B s_B n\sigma^2(T+1)\end{aligned}$$

Proof. Notice that

$$\sum_{k=0}^{T}\sum_{i=0}^{k} b_{k-i}\mathbb{E}\left[\Delta^{(i)}\right] = \sum_{i=0}^{T}\mathbb{E}\left[\Delta^{(i)}\right]\sum_{k=i}^{T} b_{k-i} \le 2M_B s_B\sum_{i=0}^{T}\mathbb{E}\left[\Delta^{(i)}\right] = 2M_B s_B\sum_{k=0}^{T}\mathbb{E}\left[\Delta^{(k)}\right]$$

We substitute Lemma 5 in Lemma 4, and obtain that

$$\begin{aligned}&\sum_{k=0}^{T}\sum_{i=0}^{k}\mathbb{E}\left[\left\langle (B^{k-i+1}-B_\infty)\Delta_y^{(i)}, (B^{k-i}-B_\infty)\Delta_g^{(i)}\right\rangle\right] \\ \le & (0.5\alpha\eta_1^{-1}+\eta_2^{-1})L\sum_{k=0}^{T}\sum_{i=0}^{k} b_{k-i}\mathbb{E}\left[\|\Delta_y^{(i)}\|\right] + 0.5\eta_1\alpha L\sum_{k=0}^{T}\sum_{i=0}^{k} b_{k-i}\mathbb{E}\left[\|A_\infty\mathbf{y}^{(i)}\|\right] \\ & + 0.5\eta_2 L\sum_{k=0}^{T}\sum_{i=0}^{k} b_{k-i}\mathbb{E}\left[\|\mathbf{x}^{(i+1)}-\mathbf{w}^{(i+1)}\|\right] + 0.5\eta_2 L\sum_{k=0}^{T}\sum_{i=0}^{k} b_{k-i}\mathbb{E}\left[\|\mathbf{x}^{(i)}-\mathbf{w}^{(i)}\|\right] + n\sigma^2\sum_{k=0}^{T}\sum_{i=0}^{k} b_{k-i} \\ \le & M_B s_B(\alpha\eta_1^{-1}+2\eta_2^{-1})L\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|\right] + M_B s_B\eta_1\alpha L\sum_{k=0}^{T}\mathbb{E}\left[\|A_\infty\mathbf{y}^{(k)}\|\right] \\ & + M_B s_B\eta_2 L\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|\right] + M_B s_B\eta_2 L\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k)}-\mathbf{w}^{(k)}\|\right] + 2M_B s_B n\sigma^2(T+1)\end{aligned}$$

So we obtain the lemma

$$\begin{aligned}&\sum_{k=0}^{T}\sum_{i=0}^{k}\mathbb{E}\left[\left\langle (B^{k-i+1}-B_\infty)\Delta_y^{(i)}, (B^{k-i}-B_\infty)\Delta_g^{(i)}\right\rangle\right] \\ \le & M_B s_B(\alpha\eta_1^{-1}+2\eta_2^{-1})L\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|\right] + M_B s_B\eta_1\alpha L\sum_{k=0}^{T}\mathbb{E}\left[\|A_\infty\mathbf{y}^{(k)}\|\right] \\ & + 2M_B s_B\eta_2 L\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|\right] + 2M_B s_B n\sigma^2(T+1)\end{aligned}$$

□

2.4 Gradient Consensus lemma

Lemma 7. *By setting* $\eta_1 = 10M_B s_B \alpha L, \eta_2 = 20M_b s_B L$*, and* $\alpha < \frac{1}{25\sqrt{n}\|\pi_A\| M_B s_B L}$*, we have*

$$\sum_{k=0}^{T} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] < 20M_B s_B n(T+1)\sigma^2 + 200 s_B^2 M_B^2 L^2 \sum_{k=0}^{T} \mathbb{E}\left[\|\Delta_x^{(k+1)}\|^2\right] + 120nc^2\alpha^2 s_B^2 M_B^2 L^2 \sum_{k=0}^{T} \mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right]$$

Proof. We substitute Lemma 3 and Lemma 6 in Lemma 2, and obtain that

$$\begin{aligned}&\left(1 - 2M_B s_B L(\alpha\eta_1^{-1} + 2\eta_2^{-1})\right) \sum_{k=0}^{T} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] \\ \leq & 10M_B s_B n(T+1)\sigma^2 + (18 s_B^2 L^2 + 4M_B s_B \eta_2 L) \sum_{k=0}^{T} \mathbb{E}\left[\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2\right] \\ &+ (9\alpha^2 s_B^2 L^2 + 2M_B s_B \eta_1 \alpha L) \sum_{k=0}^{T} \mathbb{E}\left[\|A_\infty \mathbf{y}^{(k)}\|^2\right]\end{aligned}$$

Noting that

$$\mathbb{E}\left[\|A_\infty \mathbf{y}^{(k)}\|^2\right] \leq 2\mathbb{E}\left[\|A_\infty B_\infty \mathbf{y}^{(k)}\|^2\right] + 2\mathbb{E}\left[\|A_\infty \Delta_y^{(k)}\|^2\right] \leq 2nc^2\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right] + 2\|A_\infty\|^2\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]$$

Where $c = n\pi_A^T \pi_B$. And we have

$$\begin{aligned}&\left(1 - 2M_B s_B L(\alpha\eta_1^{-1} + 2\eta_2^{-1}) - 2\|A_\infty\|_2^2(9\alpha^2 s_B^2 L^2 + 2M_B s_B \eta_1 \alpha L)\right) \sum_{k=0}^{T} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] \\ \leq & 10M_B s_B n(T+1)\sigma^2 + (18 s_B^2 L^2 + 4M_B s_B \eta_2 L) \sum_{k=0}^{T} \mathbb{E}\left[\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2\right] \\ &+ 2nc^2(9\alpha^2 s_B^2 L^2 + 2M_B s_B \eta_1 \alpha L) \sum_{k=0}^{T} \mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right]\end{aligned}$$

By setting $\eta_1 = \mathbf{p} \cdot M_B s_B \alpha L, \eta_2 = 2\mathbf{p} \cdot M_B s_B L$, we have

$$\begin{aligned}&\left(1 - 2M_B s_B L(\alpha\eta_1^{-1} + 2\eta_2^{-1}) - 2\|A_\infty\|_2^2(9\alpha^2 s_B^2 L^2 + 2M_B s_B \eta_1 \alpha L)\right) \\ = & 1 - \frac{4}{\mathbf{p}} - 2\alpha^2 s_B^2 L^2 \|A_\infty\|_2^2 (9 + 2M_B^2 \mathbf{p})\end{aligned}$$

Let $s_B L \|A_\infty\|_2$ be denoted as $\mathbf{D} = s_B L \|A_\infty\|_2$. We want $\frac{1}{2} \leq 1 - \frac{4}{\mathbf{p}} - 2\mathbf{D}^2\alpha^2(9 + 2M_B^2\mathbf{p})$; this is equivalent to the following inequality

$$2\mathbf{D}^2\alpha^2(9\mathbf{p} + 2M_B^2\mathbf{p}^2) \leq \frac{\mathbf{p}}{2} - 4$$

By setting $\mathbf{p} = 10$, solving the inequality yields an upper bound for α:

$$\alpha < \sqrt{\frac{1}{2\mathbf{D}^2(200M_B^2 + 90)}} = \sqrt{\frac{1}{2s_B^2 L^2 \|A_\infty\|_2^2 (200M_B^2 + 90)}}$$

Substituting $\eta_1 = 10 \cdot M_B s_B \alpha L, \eta_2 = 20 \cdot M_B s_B L$, we obtain that

$$\sum_{k=0}^{T} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] \leq 20M_B s_B n(T+1)\sigma^2 + 2s_B^2 L^2 (18 + 80M_B^2) \sum_{k=0}^{T} \mathbb{E}\left[\|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2\right]$$

$$+ 4nc^2\alpha^2 s_B^2 L^2(9+20M_B^2)\sum_{k=0}^{T}\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right]$$

Since M_B is typically larger than 1, we can simplify the upper bound

$$\alpha < \frac{1}{25\sqrt{n}\|\pi_A\|M_B s_B L} = \frac{1}{25M_B s_B L\|A_\infty\|} < \sqrt{\frac{1}{580M_B^2 s_B^2 L^2\|A_\infty\|_2^2}} < \sqrt{\frac{1}{2s_B^2 L^2\|A_\infty\|_2^2(200M_B^2+90)}}$$

and the inequality

$$\begin{aligned}\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] \leq & 20M_B s_B n(T+1)\sigma^2 + 2s_B^2 L^2(18+80M_B^2)\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|^2\right] \\ & + 4nc^2\alpha^2 s_B^2 L^2(9+20M_B^2)\sum_{k=0}^{T}\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right] \\ < & 20M_B s_B n(T+1)\sigma^2 + 200s_B^2 M_B^2 L^2\sum_{k=0}^{T}\mathbb{E}\left[\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|^2\right] \\ & + 120nc^2\alpha^2 s_B^2 M_B^2 L^2\sum_{k=0}^{T}\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right]\end{aligned}$$

We finish the proof of the lemma. □

2.5 Consensus Lemma 1

Lemma 8. *By setting $\alpha \leq \min\{\frac{1}{20s_A s_B M_B L}, \frac{1}{25\sqrt{n}\|\pi_A\|M_B s_B L}\}$, we have*

$$\begin{aligned}\sum_{k=0}^{T}\|\Delta_x^{(k+1)}\|^2 \leq & 2\left(2\alpha^2 s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\bar{g}^{(k)}\|^2 \\ & + 80n\alpha^2 s_A^2 M_B s_B(T+1)\sigma^2\end{aligned}$$

Proof. By definition of $\mathbf{w}^{(k)}$, we have $\Delta_x^{(k+1)} = \mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)} = -\alpha\sum_{i=0}^{k}(A^{k-i}-A_\infty)\mathbf{y}^{(i)}$. This implies that

$$\begin{aligned}&\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|^2 \\ =&\alpha^2\|\sum_{i=0}^{k}(A^{k-i}-A_\infty)(I-B_\infty)\mathbf{y}^{(i)} + \sum_{i=0}^{k}(A^{k-i}-A_\infty)B_\infty\mathbf{y}^{(i)}\|^2 \\ =&\alpha^2\|\sum_{i=0}^{k}(A^{k-i}-A_\infty)(I-B_\infty)\mathbf{y}^{(i)} + \sum_{i=0}^{k}(A^{k-i}-A_\infty)(n\pi_B^T-\mathbb{1}_n)\bar{y}^{(i)}\|^2 \\ \leq&2\alpha^2\|\sum_{i=0}^{k}(A^{k-i}-A_\infty)(I-B_\infty)\mathbf{y}^{(i)}\| + 2\alpha^2\|\sum_{i=0}^{k}(A^{k-i}-A_\infty)(n\pi_B^T-\mathbb{1}_n)\bar{y}^{(i)}\|\end{aligned}$$

By summing up $k=0$ to T, we have that

$$\begin{aligned}&\sum_{k=0}^{T}\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|^2 \\ \leq&2\alpha^2\sum_{k=0}^{T}\|\sum_{i=0}^{k}(A^{k-i}-A_\infty)(I-B_\infty)\mathbf{y}^{(i)}\| + 2\alpha^2\sum_{k=0}^{T}\|\sum_{i=0}^{k}(A^{k-i}-A_\infty)(n\pi_B^T-\mathbb{1}_n)\bar{y}^{(i)}\| \\ \leq&2\alpha^2 s_A^2\sum_{k=0}^{T}\|\Delta_y^{(k)}\|^2 + 2\alpha^2 s_A^2\|n\pi_B^T-\mathbb{1}_n\|^2\sum_{k=0}^{T}\|\bar{g}^{(k)}\|^2 \qquad (15)\end{aligned}$$

By further applying Lemma 7 in 15, we have

$$\left(1-400\alpha^2 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|^2$$
$$\leq\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\bar{g}^{(k)}\|^2$$
$$+40n\alpha^2 s_A^2 M_B s_B (T+1)\sigma^2 \tag{16}$$

By setting

$$\alpha\leq\min\{\frac{1}{20 s_A s_B M_B L},\ \frac{1}{25\sqrt{n}\|\pi_A\|M_B s_B L}\}$$

we have $1-400\alpha^2 s_A^2 s_B^2 M_B^2 L^2\geq 0.5$. Therefore, we can double both sides of 16 and complete the proof. ☐

2.6 Consensus Lemma 2

Lemma 9. *By setting* $\alpha\leq\min\{\frac{1}{cL},\ \frac{1}{25\sqrt{n}\|\pi_A\|M_B s_B L},\ \frac{1}{80 s_A s_B M_B L},\ \frac{\sqrt{n}}{8 s_A\|n\pi_B-\mathbb{1}_n\|L}\}$*, we have*

$$\sum_{k=0}^{T}\|\Delta_x^{(k)}\|^2 \leq 80n\alpha^2 s_A^2 M_B s_B (T+1)\sigma^2+\frac{4}{n}\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)(T+1)\sigma^2$$
$$+8\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\nabla f(w^{(k)})\|^2$$

Proof. Consider the inequality 16 in Lemma 8

$$\left(1-400\alpha^2 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\mathbf{x}^{(k+1)}-\mathbf{w}^{(k+1)}\|^2$$
$$\leq\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\bar{g}^{(k)}\|^2$$
$$+40n\alpha^2 s_A^2 M_B s_B (T+1)\sigma^2$$

Since

$$\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right]=\mathbb{E}\left[\|\frac{1}{n}\mathbb{1}_n^T\left(\mathbf{g}^{(k)}-\nabla\mathbf{f}^{(k)}+\nabla\mathbf{f}^{(k)}-\nabla\mathbf{f}(w^{(k)})+\nabla\mathbf{f}(w^{(k)})\right)\|^2\right]$$
$$\leq\frac{2}{n}\sigma^2+\frac{4L^2}{n}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right]+4\mathbb{E}\left[\|\frac{1}{n}\mathbb{1}_n^T\nabla\mathbf{f}(w^{(k)})\|^2\right]=\frac{2}{n}\sigma^2+\frac{4L^2}{n}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right]+4\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right], \tag{17}$$

we have that

$$\left(1-400\alpha^2 s_A^2 s_B^2 M_B^2 L^2-\frac{4L^2}{n}\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\right)\sum_{k=0}^{T}\|\Delta_x^{(k)}\|^2$$
$$\leq 40n\alpha^2 s_A^2 M_B s_B (T+1)\sigma^2$$
$$+\frac{2}{n}\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240c^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)(T+1)\sigma^2$$
$$+4\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\nabla f(w^{(k)})\|^2$$

We use $c\alpha L\leq 1$ to simplify the upper bound of α

$$400\alpha^2 s_A^2 s_B^2 M_B^2 L^2+\frac{4L^2}{n}\left(2\alpha^2 s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)$$

$$
\begin{aligned}
&\leq 400\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{4L^2}{n}\left(2\alpha^2 s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240n\alpha^2 s_A^2 s_B^2 M_B^2\right) \\
&< 1600\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{16\alpha^2 s_A^2 L^2\|n\pi_B - \mathbb{1}_n\|^2}{n}
\end{aligned}
$$

Since

$$
\begin{aligned}
&\frac{1}{2} \leq 1 - 400\alpha^2 s_A^2 s_B^2 M_B^2 L^2 - \frac{4L^2}{n}\left(2\alpha^2 s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right) \\
\iff & 400\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{4L^2}{n}\left(2\alpha^2 s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right) \leq \frac{1}{2} \\
\Longleftarrow & 3200\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{32\alpha^2 s_A^2 L^2\|n\pi_B - \mathbb{1}_n\|^2}{n} \leq 1
\end{aligned}
$$

So we can obtain the upper bound of α

$$
\alpha \leq \min\{\frac{1}{80 s_A s_B M_B L}, \frac{\sqrt{n}}{8 s_A\|n\pi_B - \mathbb{1}_n\|L}\}
$$

and we obtain the lemma.

$$
\begin{aligned}
\sum_{k=0}^{T}\|\Delta_x^{(k)}\|^2 \leq & 80n\alpha^2 s_A^2 M_B s_B (T+1)\sigma^2 + \frac{4}{n}\left(2\alpha^2 s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)(T+1)\sigma^2 \\
& + 8\left(2\alpha^2 s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240nc^2\alpha^4 s_A^2 s_B^2 M_B^2 L^2\right)\sum_{k=0}^{T}\|\nabla f(w^{(k)})\|^2
\end{aligned}
$$

□

2.7 Descent Lemma: Basic

Lemma 10.

$$
\begin{aligned}
&\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \\
\leq & \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \frac{2}{T+1}\left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] + \frac{4c\alpha L}{n}\sigma^2 \\
& + \frac{2\|\pi_A\|^2}{(T+1)c\alpha}\left(\frac{\alpha}{c} + \alpha^2 L\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] + \frac{2}{T+1}\left(4c\alpha L - \frac{1}{4}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

Proof. Since $w^{(k+1)} = w^{(k)} - \alpha\pi_A^T\mathbf{y}^{(k)}$, we can apply the descent lemma and obtain that

$$
f(w^{(k+1)}) \leq f(w^{(k)}) - \alpha\left\langle \pi_A^T\mathbf{y}^{(k)}, \nabla f(w^{(k)})\right\rangle + \frac{\alpha^2 L}{2}\|\pi_A^T\mathbf{y}^{(k)}\|^2
$$

Taking conditional expectation, we have

$$
\mathbb{E}\left[f(w^{(k+1)})\right] \leq \mathbb{E}\left[f(w^{(k)})\right] - \alpha\mathbb{E}\left[\left\langle \pi_A^T\mathbf{y}^{(k)}, \nabla f(w^{(k)})\right\rangle\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right]
$$

Noting that $\pi_A^T\mathbf{y}^{(k)} = c\bar{g}^{(k)} + \pi_A^T\Delta_y^{(k)}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[f(w^{(k+1)})\right] - \mathbb{E}\left[f(w^{(k)})\right] \\
\leq & - c\alpha\mathbb{E}\left[\left\langle \bar{g}^{(k)}, \nabla f(w^{(k)})\right\rangle\right] - \alpha\mathbb{E}\left[\left\langle \pi_A^T\Delta_y^{(k)}, \nabla f(w^{(k)})\right\rangle\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right] \\
= & - c\alpha\mathbb{E}\left[\left\langle \overline{\nabla f}^{(k)}, \nabla f(w^{(k)})\right\rangle\right] - \alpha\mathbb{E}\left[\left\langle \pi_A^T\Delta_y^{(k)}, \nabla f(w^{(k)})\right\rangle\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq -\frac{c\alpha}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] - \frac{c\alpha}{2}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] + \frac{c\alpha}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)} - \nabla f(w^{(k)})\|^2\right] \\
&\quad + \frac{\alpha}{c}\mathbb{E}\left[\|\pi_A^T\Delta_y^{(k)}\|^2\right] + \frac{c\alpha}{4}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right] \\
&= -\frac{c\alpha}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] - \frac{c\alpha}{4}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] + \frac{c\alpha}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)} - \nabla f(w^{(k)})\|^2\right] \\
&\quad + \frac{\alpha}{c}\mathbb{E}\left[\|\pi_A^T\Delta_y^{(k)}\|^2\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right]
\end{aligned}
$$

Notice that

$$
\mathbb{E}\left[\|\overline{\nabla f}^{(k)} - \nabla f(w^{(k)})\|^2\right] = \mathbb{E}\left[\|\frac{1}{n}\mathbb{1}_n^T(\nabla f(\mathbf{x}^{(k)}) - \nabla f(\mathbf{w}^{(k)}))\|^2\right] \leq \frac{2L^2}{n}\mathbb{E}\left[\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2\right]
$$

we can obtain that

$$
\begin{aligned}
&\mathbb{E}\left[f(w^{(k+1)})\right] - \mathbb{E}\left[f(w^{(k)})\right] \\
&\leq -\frac{c\alpha}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] - \frac{c\alpha}{4}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] + \frac{c\alpha L^2}{n}\mathbb{E}\left[\|\mathbf{x}^{(k)} - \mathbf{w}^{(k)}\|^2\right] \\
&\quad + \frac{\alpha}{c}\mathbb{E}\left[\|\pi_A^T\Delta_y^{(k)}\|^2\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right]
\end{aligned} \tag{18}
$$

Further notice that

$$
\|\pi_A^T\mathbf{y}^{(k)}\|^2 = \|\pi_A^T B_\infty \mathbf{y}^{(k)} + \pi_A^T(I - B_\infty)\mathbf{y}^{(k)}\|^2 \leq 2c^2\|\bar{g}^{(k)}\|^2 + 2\|\pi_A^T\Delta_y^{(k)}\|^2,
$$

where $c = n\pi_A^T\pi_B$, and the same as 17, we have

$$
\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right] \leq \frac{2}{n}\sigma^2 + \frac{4L^2}{n}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] + 4\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
$$

Where the last inequality utilizes the property that the gradients and stochastic gradients of each node are independent of each other. So we have that

$$
\begin{aligned}
&\mathbb{E}\left[\|\pi_A^T\mathbf{y}^{(k)}\|^2\right] \leq 2c^2\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right] + 2\|\pi_A^T\Delta_y^{(k)}\|^2 \\
&\leq 2\|\pi_A\|^2\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] + \frac{4c^2}{n}\sigma^2 + \frac{8c^2L^2}{n}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] + 8c^2\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned} \tag{19}
$$

Substitute 19 to 18, we have that

$$
\begin{aligned}
&\mathbb{E}\left[f(w^{(k+1)})\right] - \mathbb{E}\left[f(w^{(k)})\right] \\
&\leq -\frac{c\alpha}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] + \left(\frac{c\alpha L^2}{n} + \frac{4c^2\alpha^2 L^3}{n}\right)\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] \\
&\quad + \|\pi_A\|^2\left(\frac{\alpha}{c} + \alpha^2 L\right)\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] + \frac{2c^2\alpha^2 L}{n}\sigma^2 + \left(4c^2\alpha^2 L - \frac{c\alpha}{4}\right)\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

By summing up from $k = 0$ to T, we obtain the lemma.

$$
\begin{aligned}
&\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \\
&\leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \frac{2}{T+1}\left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] + \frac{4c\alpha L}{n}\sigma^2 \\
&\quad + \frac{2\|\pi_A\|^2}{(T+1)c\alpha}\left(\frac{\alpha}{c} + \alpha^2 L\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] + \frac{2}{T+1}\left(4c\alpha L - \frac{1}{4}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

We finish the proof of this lemma. $\square$

2.8 Main Theorem: Basic

Theorem 1. *By setting* $\alpha \leq \min\{\frac{1}{32cL}, \frac{1}{1600\sqrt{n}\|\pi_A\|M_B s_B L}, \frac{1}{80 s_A s_B M_B L}, \frac{\sqrt{n}}{8 s_A \|n\pi_B - \mathbb{1}_n\| L}, \frac{1}{4L\sqrt{\tilde{\mathbf{D}}_2}}\}$, *we have that*

$$\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \mathbf{D_1}(1)\sigma^2 \leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \tilde{\mathbf{D}}_\mathbf{1}\sigma^2$$

Where

$$\begin{aligned}\tilde{\mathbf{D}}_\mathbf{1} =& \frac{4}{n} + \frac{1200n\|\pi_A\|^2 M_B^2 s_B^2}{c^2} + \frac{17000}{c^4} n\|\pi_A\|^2 s_A^2 M_B^3 s_B^2 + 160 n s_A^2 M_B s_B\left(\frac{5}{nc^2} + \frac{400}{c^4}\|\pi_A\|^2 M_B^2 s_B^2\right)\\ &+\left(\frac{40}{n^2c^2} + \frac{12000}{nc^4}\|\pi_A\|^2 M_B^2 s_B^2\right)\left(2s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240 n s_A^2 s_B^2 M_B^2\right) \qquad (20)\end{aligned}$$

$$\tilde{\mathbf{D}}_\mathbf{2} = \left(\frac{80}{n} + \frac{24000}{c^2}\|\pi_A\|^2 M_B^2 s_B^2\right)\left(2s_A^2\|n\pi_B - \mathbb{1}_n\|^2 + 240 n s_A^2 s_B^2 M_B^2\right) \qquad (21)$$

Proof. Substitute $\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]$ in Lemma 10 by Lemma 7, we have

$$\begin{aligned}&\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\ \leq&\frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \frac{2}{T+1}\left(4c\alpha L - \frac{1}{4}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]\\ &+\frac{2}{T+1}\left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2\left(\frac{1}{c^2} + \frac{\alpha L}{c}\right)\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right]\\ &+\left(\frac{4c\alpha L}{n} + 40n\|\pi_A\|^2 M_B s_B\left(\frac{1}{c^2} + \frac{\alpha L}{c}\right)\right)\sigma^2\\ &+\frac{240n\|\pi_A\|^2 M_B^2 s_B^2}{T+1}\left(\alpha^2L^2 + c\alpha^3L^3\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right]\end{aligned}$$

The same as 17, we have

$$\mathbb{E}\left[\|\bar{g}^{(k)}\|^2\right] \leq \frac{2}{n}\sigma^2 + \frac{4L^2}{n}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] + 4\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right],$$

and we have that

$$\begin{aligned}&\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\ \leq&\frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)}\\ &+\frac{2}{T+1}\left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2\left(\frac{1}{c^2} + \frac{\alpha L}{c}\right)\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right]\\ &+\frac{960\|\pi_A\|^2 M_B^2 s_B^2}{T+1}\left(\alpha^2L^4 + c\alpha^3L^5\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right]\\ &+\left(\frac{4c\alpha L}{n} + 40n\|\pi_A\|^2 M_B s_B\left(\frac{1}{c^2} + \frac{\alpha L}{c}\right) + 480\|\pi_A\|^2 M_B^2 s_B^2\left(\alpha^2L^2 + c\alpha^3L^3\right)\right)\sigma^2\\ &+\frac{2}{T+1}\left(480n\|\pi_A\|^2 M_B^2 s_B^2\left(\alpha^2L^2 + c\alpha^3L^3\right) + 4c\alpha L - \frac{1}{4}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]\end{aligned}$$

Substitute $\sum_{k=0}^{T}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right]$ by Lemma 9, we have

$$\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \leq \frac{2}{T+1}\left(480n\|\pi_A\|^2M_B^2s_B^2\left(\alpha^2L^2+c\alpha^3L^3\right)+\frac{\mathbf{D_2}(\alpha^2)}{2}+4c\alpha L-\frac{1}{4}\right)\sum_{k=0}^{T}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] + \frac{2(\nabla f(w^{(0)})-\nabla f(w^{(*)}))}{c\alpha(T+1)}+\mathbf{D_1}(1)\sigma^2$$

Where

$$\begin{aligned}\mathbf{D_1}(1) =& \frac{4c\alpha L}{n}+40n\|\pi_A\|^2M_Bs_B\left(\frac{1}{c^2}+\frac{\alpha L}{c}\right)+480n\|\pi_A\|^2M_B^2s_B^2\left(\alpha^2L^2+c\alpha^3L^3\right)\\ &+160n\alpha^2s_A^2M_Bs_B\left(\frac{L^2}{n}+\frac{4c\alpha L^3}{n}+200\|\pi_A\|^2M_B^2s_B^2L^2\left(\frac{1}{c^2}+\frac{\alpha L}{c}\right)\right)\\ &+\frac{8}{n}\left(\frac{L^2}{n}+\frac{4c\alpha L^3}{n}+200\|\pi_A\|^2M_B^2s_B^2L^2\left(\frac{1}{c^2}+\frac{\alpha L}{c}\right)\right)\left(2\alpha^2s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4s_A^2s_B^2M_B^2L^2\right)\\ &+7680n\|\pi_A\|^2s_A^2M_B^3s_B^3\left(\alpha^4L^4+c\alpha^5L^5\right)\\ &+\frac{3840\|\pi_A\|^2M_B^2s_B^2}{n}\left(\alpha^2L^4+c^3\alpha^3L^5\right)\left(2\alpha^2s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4s_A^2s_B^2M_B^2L^2\right)\end{aligned}$$

and

$$\begin{aligned}\mathbf{D_2}(\alpha^2) =& 16\left(\frac{L^2}{n}+\frac{4c\alpha L^3}{n}+200\|\pi_A\|^2M_B^2s_B^2L^2\left(\frac{1}{c^2}+\frac{\alpha L}{c}\right)\right)\left(2\alpha^2s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4s_A^2s_B^2M_B^2L^2\right)\\ &+7680\|\pi_A\|^2M_B^2s_B^2\left(\alpha^2L^4+c\alpha^3L^5\right)\left(2\alpha^2s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240nc^2\alpha^4s_A^2s_B^2M_B^2L^2\right)\end{aligned}$$

Use the bound $c\alpha L\leq 1$, we obtain that $\mathbf{D_1}(1)\leq\tilde{\mathbf{D}}_\mathbf{1}, \mathbf{D_2}(\alpha^2)\leq\alpha^2L^2\tilde{\mathbf{D}}_\mathbf{2}$, where

$$\begin{aligned}\tilde{\mathbf{D}}_\mathbf{1} =& \frac{4}{n}+\frac{1200n\|\pi_A\|^2M_B^2s_B^2}{c^2}+\frac{17000}{c^4}n\|\pi_A\|^2s_A^2M_B^3s_B^2+160ns_A^2M_Bs_B\left(\frac{5}{nc^2}+\frac{400}{c^4}\|\pi_A\|^2M_B^2s_B^2\right)\\ &+\left(\frac{40}{n^2c^2}+\frac{12000}{nc^4}\|\pi_A\|^2M_B^2s_B^2\right)\left(2s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240ns_A^2s_B^2M_B^2\right)\\ \tilde{\mathbf{D}}_\mathbf{2} =& \left(\frac{80}{n}+\frac{24000}{c^2}\|\pi_A\|^2M_B^2s_B^2\right)\left(2s_A^2\|n\pi_B-\mathbb{1}_n\|^2+240ns_A^2s_B^2M_B^2\right)\end{aligned}$$

We further set $c\alpha L\leq\frac{1}{32}$, and we can obtain that

$$480n\|\pi_A\|^2M_B^2s_B^2\left(\alpha^2L^2+c\alpha^3L^3\right)+\frac{\mathbf{D_2}(\alpha^2)}{2}+4c\alpha L-\frac{1}{4}\leq 1000n\|\pi_A\|^2M_B^2s_B^2\alpha^2L^2+\frac{\alpha^2L^2\tilde{\mathbf{D}}_\mathbf{2}}{2}-\frac{1}{8}$$

So we have that

$$\alpha\leq\min\{\frac{1}{1600\sqrt{n}\|\pi_A\|M_Bs_BL},\ \frac{1}{4L\sqrt{\tilde{\mathbf{D}}_\mathbf{2}}}\}\implies 8000n\|\pi_A\|^2M_B^2s_B^2\alpha^2L^2+4\alpha^2\tilde{\mathbf{D}}_\mathbf{2}-1\leq 0$$

So use the upper bound we obtain the lemma.

$$\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\leq\frac{2(\nabla f(w^{(0)})-\nabla f(w^{(*)}))}{c\alpha(T+1)}+\mathbf{D_1}(1)\sigma^2\leq\frac{2(\nabla f(w^{(0)})-\nabla f(w^{(*)}))}{c\alpha(T+1)}+\tilde{\mathbf{D}}_\mathbf{1}\sigma^2$$

□

3 Analysis

Note that in the Theorem 1, the coefficient $\tilde{\mathbf{D}}_\mathbf{1}$ of σ^2 is independent of the learning rate α, which is unrealistic in the analysis of the convergence. To address this issue, we consider applying the inequality derived from the L-smooth property to the m-step update equation of w.

Lemma 11.

$$\frac{f(w^{(*)})-f(w^{(0)})}{m(K+1)} \leq -\frac{\alpha}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)},\nabla f(w^{(k)})\right\rangle\right] + \frac{\alpha^2 L}{2m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2\right]$$

Proof. Since $w^{(k+m)} = w^{(k)} - \alpha\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}$, we can apply the descent lemma and obtain that

$$f(w^{(k+m)}) \leq f(w^{(k)}) - \alpha\left\langle \pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)},\nabla f(w^{(k)})\right\rangle + \frac{\alpha^2 L}{2}\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2$$

Taking conditional expectation, we have

$$\mathbb{E}\left[f(w^{(k+m)})\right] \leq \mathbb{E}\left[f(w^{(k)})\right] - \alpha\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)},\nabla f(w^{(k)})\right\rangle\right] + \frac{\alpha^2 L}{2}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2\right]$$

By summing over $k=0,m,\cdots,mK$, we have $T=m(K+1)$, and we have

$$\frac{f(w^{(*)})-f(w^{(0)})}{m(K+1)} \leq -\frac{\alpha}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)},\nabla f(w^{(k)})\right\rangle\right] + \frac{\alpha^2 L}{2m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2\right]$$

Then we finish the proof of the lemma. □

Next, we decompose the expectations of the separated quadratic term and the inner product term.

Lemma 12.

$$\frac{\alpha^2 L}{2}\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2 \leq c^2\alpha^2 L\|\sum_{i=0}^{m-1}\bar{g}^{(k+i)}\|^2 + 2\alpha^2 L\|\pi_A^T(\sum_{i=0}^{m-1}B^i - mB_\infty)\mathbf{y}^{(k)}\|^2 + 2\alpha^2 L\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2$$

Proof. Since $\sum_{i=0}^{m-1}\pi_A^T\mathbf{y}^{(k+i)} = c\sum_{i=0}^{m-1}\bar{g}^{(k+i)} + \sum_{i=0}^{m-1}\pi_A^T(I-B_\infty)\mathbf{y}^{(k+i)}$, the squared norm term can be decomposed as follows.

$$\frac{\alpha^2 L}{2}\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2 \leq c^2\alpha^2 L\|\sum_{i=0}^{m-1}\bar{g}^{(k+i)}\|^2 + \alpha^2 L\|\sum_{i=0}^{m-1}\pi_A^T(I-B_\infty)\mathbf{y}^{(k+i)}\|^2$$

Since $\sum_{i=0}^{m-1}\pi_A^T(I-B_\infty)\mathbf{y}^{(k+i)} = \pi_A^T(\sum_{i=0}^{m-1}B^i - mB_\infty)\mathbf{y}^{(k)} + \pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i} - B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})$, we have

$$\frac{\alpha^2 L}{2}\|\pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)}\|^2 \leq c^2\alpha^2 L\|\sum_{i=0}^{m-1}\bar{g}^{(k+i)}\|^2 + 2\alpha^2 L\|\pi_A^T(\sum_{i=0}^{m-1}B^i - mB_\infty)\mathbf{y}^{(k)}\|^2 + 2\alpha^2 L\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2$$

We finish the proof of the lemma. □

Lemma 13.

$$
\begin{aligned}
&-\alpha\mathbb{E}\left[\left\langle \pi_A^T \sum_{i=0}^{m-1} \mathbf{y}^{(k+i)}, \nabla f(w^{(k)}) \right\rangle\right] \\
=&-\alpha\mathbb{E}\left[\left\langle \pi_A^T (\sum_{i=0}^{m-1} B^i - B_\infty)\mathbf{y}^{(k)}, \nabla f(w^{(k)}) \right\rangle\right] - c\alpha m \mathbb{E}\left[\left\langle \bar{g}^{(k)}, \nabla f(w^{(k)}) \right\rangle\right] \\
&-\alpha\mathbb{E}\left[\left\langle \pi_A^T \sum_{i=0}^{m-1} B^{m-1-i}(\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)}), \nabla f(w^{(k)}) \right\rangle\right]
\end{aligned}
$$

Proof. Consider the inner product term $-\alpha \left\langle \pi_A^T \sum_{i=0}^{m-1} \mathbf{y}^{(k+i)}, \nabla f(w^{(k)}) \right\rangle$, we have that

$$
\begin{aligned}
&-\alpha \left\langle \pi_A^T \sum_{i=0}^{m-1} \mathbf{y}^{(k+i)}, \nabla f(w^{(k)}) \right\rangle \\
=&-\alpha \left\langle \pi_A^T (\sum_{i=0}^{m-1} B^i - B_\infty)\mathbf{y}^{(k)}, \nabla f(w^{(k)}) \right\rangle - c\alpha m \left\langle \bar{g}^{(k)}, \nabla f(w^{(k)}) \right\rangle \\
&-\alpha \left\langle \pi_A^T \sum_{i=0}^{m-1} B^{m-1-i}(\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)}), \nabla f(w^{(k)}) \right\rangle
\end{aligned}
$$

taking conditional expectation, we obtain the lemma. ☐

4 Convergence Analysis: Quadratic Term

4.1 Technical Lemmas

Now, we perform upper bound estimates for the decomposed terms of the expectation of the quadratic term in Lemma 12.

Lemma 14.

$$
\begin{aligned}
&\frac{c^2\alpha^2 L}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\sum_{i=0}^{m-1} \bar{g}^{(k+i)}\|^2\right] \\
\le & 2c^2\alpha^2 L\sigma^2 + \frac{4c^2\alpha^2 mL}{K+1} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] \\
&+ \frac{8c^2\alpha^2 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{16c^4 m^2\alpha^4 L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right] \\
&+ \frac{16c^2 m^2\alpha^4 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
\end{aligned}
$$

Proof. Consider $c^2\alpha^2 L\|\sum_{i=0}^{m-1} \bar{g}^{(k+i)}\|^2$, taking conditional expectation, we have

$$
c^2\alpha^2 L\mathbb{E}\left[\|\sum_{i=0}^{m-1} \bar{g}^{(k+i)}\|^2\right] \le 2c^2\alpha^2 L\mathbb{E}\left[\|\sum_{i=0}^{m-1} (\bar{g}^{(k+i)} - \overline{\nabla f}^{(k+i)})\|^2\right] + 2c^2\alpha^2 mL \sum_{i=0}^{m-1} \mathbb{E}\left[\|\overline{\nabla f}^{(k+i)}\|^2\right]
$$

Notice that, $\forall O \in \mathbb{R}^{n\times n},\ \|\mathbb{1}_n^T O\|^2 \le n\|O\|^2$, so we have that

$$
\mathbb{E}\left[\|\sum_{i=0}^{m-1} (\bar{g}^{(k+i)} - \overline{\nabla f}^{(k+i)})\|^2\right] = \mathbb{E}\left[\|\frac{1}{n}\mathbb{1}_n^T \sum_{i=0}^{m-1} (\mathbf{g}^{(k+i)} - \nabla \mathbf{f}^{(k+i)})\|^2\right] \le \frac{1}{n}\mathbb{E}\left[\|\sum_{i=0}^{m-1} (\mathbf{g}^{(k+i)} - \nabla \mathbf{f}^{(k+i)})\|^2\right]
$$

Since the $\mathbf{g}^{(k+i)} - \nabla \mathbf{f}^{(k+i)}$ of each step $k+i, i \in [0, m-1]$ is independent of each other, we have that

$$\frac{1}{n}\mathbb{E}\left[\|\sum_{i=0}^{m-1}(\mathbf{g}^{(k+i)} - \nabla \mathbf{f}^{(k+i)})\|^2\right] = \frac{1}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\|\mathbf{g}^{(k+i)} - \nabla \mathbf{f}^{(k+i)}\|^2\right] = \frac{1}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\left\|\begin{pmatrix} g_1^{(k+i)} - \nabla f_1^{(k+i)} \\ g_2^{(k+i)} - \nabla f_2^{(k+i)} \\ \vdots \\ g_n^{(k+i)} - \nabla f_n^{(k+i)} \end{pmatrix}\right\|^2\right]$$

Since on each step $k+i, i \in [0, m-1]$, the element $g_j^{(k+i)} - \nabla f_j^{(k+i)}, j \in [1, n]$ is independent of each other, we have that

$$\frac{1}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\left\|\begin{pmatrix} g_1^{(k+i)} - \nabla f_1^{(k+i)} \\ g_2^{(k+i)} - \nabla f_2^{(k+i)} \\ \vdots \\ g_n^{(k+i)} - \nabla f_n^{(k+i)} \end{pmatrix}\right\|^2\right] = \frac{1}{n}\sum_{i=0}^{m-1}\sum_{j=1}^{n}\mathbb{E}\left[\|g_j^{(k+i)} - \nabla f_j^{(k+i)}\|^2\right] \leq m\sigma^2$$

So we have that

$$2c^2\alpha^2 L\mathbb{E}\left[\|\sum_{i=0}^{m-1}(\bar{g}^{(k+i)} - \overline{\nabla f}^{(k+i)})\|^2\right] \leq \frac{2c^2\alpha^2 L}{n}\mathbb{E}\left[\|\sum_{i=0}^{m-1}(\mathbf{g}^{(k+i)} - \nabla \mathbf{f}^{(k+i)})\|^2\right] \leq 2c^2\alpha^2 mL\sigma^2$$

So we have

$$c^2\alpha^2 L\mathbb{E}\left[\|\sum_{i=0}^{m-1}\bar{g}^{(k+i)}\|^2\right] \leq 2c^2\alpha^2 mL\sigma^2 + 2c^2\alpha^2 mL\sum_{i=0}^{m-1}\mathbb{E}\left[\|\overline{\nabla f}^{(k+i)}\|^2\right]$$

Noting that $\|\overline{\nabla f}^{(k+i)}\|^2 \leq 2\|\overline{\nabla f}^{(k+i)} - \nabla f(w^{(k)})\|^2 + 2\|\nabla f(w^{(k)})\|^2$, we have

$$\begin{aligned}
&2c^2\alpha^2 mL\sum_{i=0}^{m-1}\mathbb{E}\left[\|\overline{\nabla f}^{(k+i)}\|^2\right] \\
\leq &4c^2\alpha^2 mL\sum_{i=0}^{m-1}\mathbb{E}\left[\|\overline{\nabla f}^{(k+i)} - \nabla f(w^{(k)})\|^2\right] + 4c^2\alpha^2 mL\sum_{i=0}^{m-1}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] \\
\leq &\frac{4c^2\alpha^2 mL^3}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\|\mathbf{x}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right] + 4c^2\alpha^2 m^2 L\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}$$

The last inequality above is because of

$$\begin{aligned}
&\sum_{i=0}^{m-1}\mathbb{E}\left[\|\overline{\nabla f}^{(k+i)} - \nabla f(w^{(k)})\|^2\right] = \sum_{i=0}^{m-1}\mathbb{E}\left[\|\frac{1}{n}\mathbb{1}_n^T\left(\nabla \mathbf{f}^{(k+i)} - \nabla \mathbf{f}(w^{(k)})\right)\|^2\right] \\
\leq &\frac{1}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\|\nabla \mathbf{f}^{(k+i)} - \nabla \mathbf{f}(w^{(k)})\|^2\right] = \frac{1}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\left\|\begin{pmatrix} \nabla f_1(x_1^{(k+i)}) - \nabla f_1(w^{(k)}) \\ \nabla f_2(x_2^{(k+i)}) - \nabla f_2(w^{(k)}) \\ \vdots \\ \nabla f_n(x_n^{(k+i)}) - \nabla f_n(w^{(k)}) \end{pmatrix}\right\|^2\right] \\
\leq &\frac{L^2}{n}\sum_{i=0}^{m-1}\mathbb{E}\left[\left\|\begin{pmatrix} x_1^{(k+i)} - w^{(k)} \\ x_2^{(k+i)} - w^{(k)} \\ \vdots \\ x_n^{(k+i)} - w^{(k)} \end{pmatrix}\right\|^2\right] = \frac{L^2}{n}\sum_{i=1}^{m}\mathbb{E}\left[\|\mathbf{x}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right]
\end{aligned}$$

By summing over $k = 0,\ m, \cdots,\ mK$, we have $T = m(K+1)$, and we have

$$\frac{c^2\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\sum_{i=0}^{m-1}\bar{g}^{(k+i)}\|^2\right]$$

$$
\leq 2c^2\alpha^2 L\sigma^2 + \frac{4c^2\alpha^2 L^3}{n(K+1)} \sum_{k=0,m,\cdots,mK} \sum_{i=0}^{m-1} \mathbb{E}\left[\|\mathbf{x}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right]
+ \frac{4c^2\alpha^2 mL}{K+1} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
$$

Note that $\mathbb{E}\left[\|\mathbf{x}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right] \leq 2\mathbb{E}\left[\|\Delta_x^{(k+i)}\|^2\right] + 2\mathbb{E}\left[\|\mathbf{w}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right]$ and that

$$
\mathbb{E}\left[\|\mathbf{w}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right] = \mathbb{E}\left[\|\alpha A_\infty \sum_{j=0}^{i-1} \mathbf{y}^{(k+j)}\|^2\right] \leq 2\mathbb{E}\left[\|\alpha A_\infty B_\infty \sum_{j=0}^{i-1} \mathbf{y}^{(k+j)}\|^2\right] + 2\mathbb{E}\left[\|\alpha A_\infty \sum_{j=0}^{i-1} \Delta_y^{(k+j)}\|^2\right]
$$
$$
= 2c^2\alpha^2 \mathbb{E}\left[\|\mathbb{1}_n^T \sum_{j=0}^{i-1} \bar{g}^{(k+j)}\|^2\right] + 2\alpha^2 \mathbb{E}\left[\|A_\infty \sum_{j=0}^{i-1} \Delta_y^{(k+j)}\|^2\right]
$$
$$
\leq 2nc^2\alpha^2 \cdot i \sum_{j=0}^{i-1} \mathbb{E}\left[\|\bar{g}^{(k+j)}\|^2\right] + 2n\alpha^2\|\pi_A\|^2 \cdot i \sum_{j=0}^{i-1} \mathbb{E}\left[\|\Delta_y^{(k+j)}\|^2\right]
$$

Where $c = n\pi_A^T\pi_B$, then consider the simplified summations.

$$
\sum_{k=0,m,\cdots,mK} \sum_{i=0}^{m-1} i \sum_{j=0}^{i-1} \mathbb{E}\left[\|\bar{g}^{(k+j)}\|^2\right] \leq \sum_{k=0,m,\cdots,mK} \sum_{i=0}^{m-1} i \sum_{j=0}^{i} \mathbb{E}\left[\|\bar{g}^{(k+j)}\|^2\right]
$$
$$
= \sum_{k=0,m,\cdots,mK} \sum_{j=0}^{m-1} \mathbb{E}\left[\|\bar{g}^{(k+j)}\|^2\right] \sum_{i=j}^{m-1} i \leq \sum_{k=0,m,\cdots,mK} \sum_{j=0}^{m-1} m\mathbb{E}\left[\|\bar{g}^{(k+j)}\|^2\right] \leq m^2 \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
$$

Similarly, we have that

$$
\sum_{k=0,m,\cdots,mK} \sum_{i=0}^{m-1} i \sum_{j=0}^{i-1} \mathbb{E}\left[\|\Delta_y^{(k+j)}\|^2\right] \leq m^2 \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
$$

Summarize the discussion above, we can obtain the estimation

$$
\frac{4c^2\alpha^2 L^3}{n(K+1)} \sum_{k=0,m,\cdots,mK} \sum_{i=0}^{m-1} \mathbb{E}\left[\|\mathbf{x}^{(k+i)} - \mathbf{w}^{(k)}\|^2\right]
$$
$$
\leq \frac{8c^2\alpha^2 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{16c^4 m^2\alpha^4 L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
+ \frac{16c^2 m^2\alpha^4 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
$$

then we obtain the lemma.

$$
\frac{c^2\alpha^2 L}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\sum_{i=0}^{m-1} \bar{g}^{(k+i)}\|^2\right]
$$
$$
\leq 2c^2\alpha^2 L\sigma^2 + \frac{4c^2\alpha^2 mL}{K+1} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
$$
$$
+ \frac{8c^2\alpha^2 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{16c^4 m^2\alpha^4 L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
$$
$$
+ \frac{16c^2 m^2\alpha^4 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
$$

We finish the proof of the lemma. $\square$

Lemma 15.

$$\frac{\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T(\sum_{i=0}^{m-1}B^i-mB_\infty)\mathbf{y}^{(k)}\|^2\right]\leq\frac{\alpha^2 s_B^2 L\|\pi_A\|^2}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]$$

Proof. Consider $\alpha^2 L\|\pi_A^T(\sum_{i=0}^{m-1}B^i-mB_\infty)\mathbf{y}^{(k)}\|^2$, taking conditional expectation, we have

$$\begin{aligned}\alpha^2 L\mathbb{E}\left[\|\pi_A^T(\sum_{i=0}^{m-1}B^i-mB_\infty)\mathbf{y}^{(k)}\|^2\right]=&\alpha^2 L\mathbb{E}\left[\|\pi_A^T(\sum_{i=0}^{m-1}B^i-mB_\infty)(I-B_\infty)\mathbf{y}^{(k)}\|^2\right]\\\leq&\alpha^2 L\|\pi_A\|^2\|\sum_{i=0}^{m-1}(B^i-B_\infty)\|^2\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]\\\leq&\alpha^2 s_B^2 L\|\pi_A\|^2\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]\end{aligned}$$

By summing over $k=0,\ m,\cdots,\ mK$, we have $T=m(K+1)$, and we have

$$\frac{\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T(\sum_{i=0}^{m-1}B^i-mB_\infty)\mathbf{y}^{(k)}\|^2\right]\leq\frac{\alpha^2 s_B^2 L\|\pi_A\|^2}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]$$

We finish the proof of the lemma. □

Lemma 16.

$$\begin{aligned}&\frac{\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2\right]\\\leq&\frac{6\alpha^2 L\|\pi_A\|^2 s_B^2}{m}\sigma^2+\frac{18\alpha^2 L^3\|\pi_A\|^2 s_B^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_x^{(t)}\|^2\\&+\frac{18\alpha^4 L^3\|\pi_A\|^2 s_B^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_y^{(t)}\|^2+\frac{18n^2\alpha^4 L^3 s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\bar{g}^{(t)}\|^2\end{aligned}$$

Proof. Consider $\alpha^2 L\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2$, and let $\mathbf{g}^{(k)}-\nabla f(\mathbf{x}^{(k)})$ be denoted as $\mathbf{G}^{(k)}=\mathbf{g}^{(k)}-\nabla f(\mathbf{x}^{(k)})$, taking conditional expectation, we have

$$\begin{aligned}&\alpha^2 L\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2\right]\\\leq&3\alpha^2 L\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\mathbf{G}^{(k+i)}\|^2\right]+3\alpha^2 L\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\mathbf{G}^{(k)}\|^2\right]\\&+3\alpha^2 L\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k)}))\|^2\right]\end{aligned}$$

Based on the independence in the expectation calculation, we have

$$3\alpha^2 L\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\mathbf{G}^{(k+i)}\|^2\right]\leq 3\alpha^2 L\sigma^2\|\pi_A\|^2\sum_{i=0}^{m-1}\|B^{m-1-i}-B_\infty\|^2$$

And we have

$$3\alpha^2 L\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\mathbf{G}^{(k)}\|^2\right]\leq 3\alpha^2 L\sigma^2\|\pi_A\|^2\|\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\|^2$$

By summing over $k=0,\ m,\cdots,\ mK$, we have $T=m(K+1)$, and we have

$$\begin{aligned}
&\frac{\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2\right]\\
\le&\frac{3\alpha^2 L\|\pi_A\|^2}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\mathbf{G}^{(k+i)}\|^2\right]\\
&+\frac{3\alpha^2 L\|\pi_A\|^2}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\mathbf{G}^{(k)}\|^2\right]\\
&+\frac{3\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k)}))\|^2\right]\\
\le&\frac{3\alpha^2 L\|\pi_A\|^2\sigma^2}{m}\sum_{i=0}^{m-1}\|B^{m-1-i}-B_\infty\|^2+\frac{3\alpha^2 L\|\pi_A\|^2\sigma^2}{m}\|\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)\|^2\\
&+\frac{3\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k)}))\|^2\right]\\
\le&\frac{3\alpha^2 L\|\pi_A\|^2 s_B^2\sigma^2}{m}+\frac{3\alpha^2 L\|\pi_A\|^2 s_B^2\sigma^2}{m}\\
&+\frac{3\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k)}))\|^2\right]
\end{aligned}$$

Noticing that

$$\begin{aligned}
&\frac{3\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k)}))\|^2\\
=&\frac{3\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\|\pi_A^T\sum_{i=1}^{m-1}(\sum_{j=i}^{m-1}(B^{m-1-j}-B_\infty))(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k+i-1)}))\|^2\\
\le&\frac{3\alpha^2 L\|\pi_A\|^2}{K+1}\sum_{k=0,m,\cdots,mK}\sum_{i=1}^{m-1}\|\sum_{j=i}^{m-1}(B^{m-1-j}-B_\infty)\|^2\|(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k+i-1)}))\|^2\\
\le&\frac{3\alpha^2 L\|\pi_A\|^2 s_B^2}{K+1}\sum_{k=0,m,\cdots,mK}\sum_{i=1}^{m-1}\|(\nabla f(\mathbf{x}^{(k+i)})-\nabla f(\mathbf{x}^{(k+i-1)}))\|^2\\
\le&\frac{3\alpha^2 L^3\|\pi_A\|^2 s_B^2}{K+1}\sum_{t=0}^{m(K+1)}\|\mathbf{x}^{(t+1)}-\mathbf{x}^{(t)}\|^2\\
\le&\frac{18\alpha^2 L^3\|\pi_A\|^2 s_B^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_x^{(t)}\|^2+\frac{9\alpha^4 L^3\|\pi_A\|^2 s_B^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\mathbf{y}^{(t)}\|^2
\end{aligned}$$

Since

$$\begin{aligned}
&\frac{9\alpha^4 L^3\|\pi_A\|^2 s_B^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\mathbf{y}^{(t)}\|^2\\
\le&\frac{18\alpha^4 L^3\|\pi_A\|^2 s_B^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_y^{(t)}\|^2+\frac{18n^2\alpha^4 L^3 s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\bar{g}^{(t)}\|^2
\end{aligned}$$

Then we have

$$\frac{\alpha^2 L}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\pi_A^T\sum_{i=0}^{m-1}(B^{m-1-i}-B_\infty)(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)})\|^2\right]$$

$$
\leq \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m}\sigma^2 + \frac{18\alpha^2 L^3 \|\pi_A\|^2 s_B^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_x^{(t)}\|^2
$$
$$
+ \frac{18\alpha^4 L^3 \|\pi_A\|^2 s_B^2 \|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_y^{(t)}\|^2 + \frac{18n^2\alpha^4 L^3 s_B^2 \|\pi_A\|^2 \|\pi_B\|^2 \|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\bar{g}^{(t)}\|^2
$$

We finish the proof of the lemma. □

4.2 Main Theorem

Theorem 2.

$$
\frac{\alpha^2 L}{2m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\pi_A^T \sum_{i=0}^{m-1} \mathbf{y}^{(k+i)}\|^2\right]
$$
$$
\leq 2c^2\alpha^2 L\sigma^2 + \frac{4c^2\alpha^2 mL}{K+1} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
$$
$$
+ \frac{8c^2\alpha^2 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{16c^4 m^2\alpha^4 L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
$$
$$
+ \frac{16c^2 m^2\alpha^4 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
$$
$$
+ \frac{\alpha^2 s_B^2 L \|\pi_A\|^2}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]
$$
$$
+ \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m}\sigma^2 + \frac{18\alpha^2 L^3 \|\pi_A\|^2 s_B^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_x^{(t)}\|^2
$$
$$
+ \frac{18\alpha^4 L^3 \|\pi_A\|^2 s_B^2 \|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_y^{(t)}\|^2 + \frac{18n^2\alpha^4 L^3 s_B^2 \|\pi_A\|^2 \|\pi_B\|^2 \|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\bar{g}^{(t)}\|^2
$$

Proof. Substitute Lemma 14, 15, and 16 to Lemma 12, we obtain that

$$
\frac{\alpha^2 L}{2m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\pi_A^T \sum_{i=0}^{m-1} \mathbf{y}^{(k+i)}\|^2\right]
$$
$$
\leq 2c^2\alpha^2 L\sigma^2 + \frac{4c^2\alpha^2 mL}{K+1} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
$$
$$
+ \frac{8c^2\alpha^2 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{16c^4 m^2\alpha^4 L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
$$
$$
+ \frac{16c^2 m^2\alpha^4 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
$$
$$
+ \frac{\alpha^2 s_B^2 L \|\pi_A\|^2}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]
$$
$$
+ \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m}\sigma^2 + \frac{18\alpha^2 L^3 \|\pi_A\|^2 s_B^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_x^{(t)}\|^2
$$
$$
+ \frac{18\alpha^4 L^3 \|\pi_A\|^2 s_B^2 \|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_y^{(t)}\|^2 + \frac{18n^2\alpha^4 L^3 s_B^2 \|\pi_A\|^2 \|\pi_B\|^2 \|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\bar{g}^{(t)}\|^2
$$

We finish the proof of the theorem. □

5 Convergence Analysis: Inner Product Term

5.1 Technical Lemmas

Now, we perform upper bound estimates for the decomposed terms of the expectation of the quadratic term in Lemma 13.

Lemma 17.

$$
\begin{aligned}
&-\frac{\alpha}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle\pi_A^T(\sum_{i=0}^{m-1}B^i-B_\infty)\mathbf{y}^{(k)},\nabla f(w^{(k)})\right\rangle\right]\\
\leq&\frac{c\alpha}{4(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2+\frac{\alpha\|\pi_A\|^2s_B^2}{cm^2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]
\end{aligned}
$$

Proof. Consider $-\alpha\mathbb{E}\left[\left\langle\pi_A^T(\sum_{i=0}^{m-1}B^i-B_\infty)\mathbf{y}^{(k)},\nabla f(w^{(k)})\right\rangle\right]$, we have that

$$
\begin{aligned}
&-\alpha\mathbb{E}\left[\left\langle\pi_A^T(\sum_{i=0}^{m-1}B^i-B_\infty)\mathbf{y}^{(k)},\nabla f(w^{(k)})\right\rangle\right]\\
=&\alpha\mathbb{E}\left[\left\langle-\pi_A^T(\sum_{i=0}^{m-1}B^i-B_\infty)(I-B_\infty)\mathbf{y}^{(k)},\nabla f(w^{(k)})\right\rangle\right]\\
\leq&\alpha\|\pi_A\|s_B\mathbb{E}\left[\|\Delta_y^{(k)}\|\|\nabla f(w^{(k)})\|\right]\\
\leq&\alpha\|\pi_A\|s_B\cdot\frac{\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]}{2}\cdot\frac{cm}{2\|\pi_A\|s_B}+\alpha\|\pi_A\|s_B\cdot\frac{\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]}{2}\cdot\frac{2\|\pi_A\|s_B}{cm}\\
\leq&\frac{cm\alpha}{4}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]+\frac{\alpha\|\pi_A\|^2s_B^2}{cm}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]
\end{aligned}
$$

By summing over $k=0,\ m,\cdots,\ mK$, we have $T=m(K+1)$, and we have

$$
\begin{aligned}
&-\frac{\alpha}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle\pi_A^T(\sum_{i=0}^{m-1}B^i-B_\infty)\mathbf{y}^{(k)},\nabla f(w^{(k)})\right\rangle\right]\\
\leq&\frac{c\alpha}{4(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2+\frac{\alpha\|\pi_A\|^2s_B^2}{cm^2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]
\end{aligned}
$$

We finish the proof of the lemma. □

Lemma 18.

$$
\begin{aligned}
&-\frac{c\alpha}{K+1}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle\bar{g}^{(k)},\nabla f(w^{(k)})\right\rangle\right]\\
\leq&-\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]+\frac{c\alpha L^2}{2n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&-\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

Proof. Consider $-c\alpha m\mathbb{E}\left[\langle\bar{g}^{(k)},\nabla f(w^{(k)})\rangle\right]$, we have that

$$
\begin{aligned}
&-c\alpha m\mathbb{E}\left[\left\langle\bar{g}^{(k)},\nabla f(w^{(k)})\right\rangle\right]\\
=&-c\alpha m\mathbb{E}\left[\left\langle\overline{\nabla f}^{(k)},\nabla f(w^{(k)})\right\rangle\right]
\end{aligned}
$$

$$
\begin{aligned}
&\leq -\frac{c\alpha m}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \underbrace{-\frac{c\alpha m}{2}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]}_{\text{do not ignore}} + \frac{c\alpha m}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)} - \nabla f(w^{(k)})\|^2\right] \\
&\leq -\frac{c\alpha m}{2}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] + \frac{c\alpha m L^2}{2n}\mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] - \frac{c\alpha m}{2}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

225 By summing over $k = 0,\ m, \cdots,\ mK$, we have $T = m(K+1)$, and we have

$$
\begin{aligned}
&-\frac{c\alpha}{K+1}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \bar{g}^{(k)}, \nabla f(w^{(k)})\right\rangle\right] \\
\leq& -\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] + \frac{c\alpha L^2}{2n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&-\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

226 We finish the proof of the lemma. □

Lemma 19.

$$
\begin{aligned}
&-\frac{\alpha}{mK}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=1}^{m-1}B^{m-1-i}(\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)}), \nabla f(w^{(k)})\right\rangle\right] \\
\leq& \frac{54\alpha L^2\|\pi_A\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&+\frac{144\alpha^3L^2\|\pi_A\|^2\|A_\infty\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] \\
&+\frac{144n^2\alpha^3L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right] \\
&+\frac{7c\alpha}{32(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

227 *Proof.* Consider $-\alpha\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=0}^{m-1}B^{m-1-i}(\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)}), \nabla f(w^{(k)})\right\rangle\right]$, we have

$$
\begin{aligned}
&-\alpha\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=1}^{m-1}B^{m-1-i}(\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)}), \nabla f(w^{(k)})\right\rangle\right] \\
=&\alpha\mathbb{E}\left[\left\langle -\pi_A^T\sum_{i=1}^{m-1}B^{m-1-i}(\nabla f(\mathbf{x}^{(k+i)}) - \nabla f(\mathbf{x}^{(k)})), \nabla f(w^{(k)})\right\rangle\right] \\
\leq&\alpha L\|\pi_A\|\sum_{i=1}^{m-1}\|B^{m-1-i}\|\mathbb{E}\left[\|\mathbf{x}^{(k+i)} - \mathbf{x}^{(k)}\|\cdot\|\nabla f(w^{(k)})\|\right] \\
\leq&3\alpha L\|\pi_A\|\sum_{i=1}^{m-1}\|B^{m-1-i}\|\mathbb{E}\left[\|\Delta_x^{(k+i)}\|\cdot\|\nabla f(w^{(k)})\|\right] \\
&+3\alpha L\|\pi_A\|\mathbb{E}\left[\|\Delta_x^{(k)}\|\cdot\|\nabla f(w^{(k)})\|\right]\sum_{i=1}^{m-1}\|B^{m-1-i}\| \\
&+3\alpha^2L\|\pi_A\|\|A_\infty\|\sum_{i=1}^{m-1}\|B^{m-1-i}\|\mathbb{E}\left[\|\sum_{j=0}^{i-1}\mathbf{y}^{(k+j)}\|\cdot\|\nabla f(w^{(k)})\|\right]
\end{aligned}
$$

228 Noting that

$$
\begin{aligned}
&\frac{3\alpha L\|\pi_A\|}{m(K+1)} \sum_{k=0,m,\cdots,mK} \sum_{i=1}^{m-1} \|B^{m-1-i}\| \mathbb{E}\left[\|\Delta_x^{(k+i)}\| \cdot \|\nabla f(w^{(k)})\|\right] \\
\le & \frac{3\alpha L\|\pi_A\|}{2m(K+1)} \cdot \frac{12L\|\pi_A\|(s_B + m\|B_\infty\|)}{mc} \sum_{k=0,m,\cdots,mK} \sum_{i=1}^{m-1} \|B^{m-1-i}\| \mathbb{E}\left[\|\Delta_x^{(k+i)}\|^2\right] \\
& + \frac{3\alpha L\|\pi_A\|}{2m(K+1)} \cdot (s_B + m\|B_\infty\|) \cdot \frac{mc}{12L\|\pi_A\|(s_B + m\|B_\infty\|)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] \\
\le & \frac{18\alpha L^2\|\pi_A\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
& + \frac{c\alpha}{8(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

229 and that

$$
\begin{aligned}
&\frac{3\alpha L\|\pi_A\|}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_x^{(k)}\| \cdot \|\nabla f(w^{(k)})\|\right] \sum_{i=1}^{m-1} \|B^{m-1-i}\| \\
\le & \frac{3\alpha L\|\pi_A\|(s_B + m\|B_\infty\|)}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_x^{(k)}\| \cdot \|\nabla f(w^{(k)})\|\right] \\
\le & \frac{3\alpha L\|\pi_A\|(s_B + m\|B_\infty\|)}{2m(K+1)} \cdot \frac{24L\|\pi_A\|(s_B + m\|B_\infty\|)}{cm} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_x^{(k)}\|^2\right] \\
& + \frac{3\alpha L\|\pi_A\|(s_B + m\|B_\infty\|)}{2m(K+1)} \cdot \frac{cm}{24L\|\pi_A\|(s_B + m\|B_\infty\|)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] \\
\le & \frac{36\alpha L^2\|\pi_A\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
& + \frac{c\alpha}{16(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}
$$

230 and that

$$
\begin{aligned}
&\frac{3\alpha^2 L\|\pi_A\|\|A_\infty\|}{m(K+1)} \sum_{k=0,m,\cdots,mK} \sum_{i=1}^{m-1} \|B^{m-1-i}\| \mathbb{E}\left[\|\sum_{j=0}^{i-1} \mathbf{y}^{(k+j)}\| \cdot \|\nabla f(w^{(k)})\|\right] \\
\le & \frac{3\alpha^2 L\|\pi_A\|\|A_\infty\|(s_B + m\|B_\infty\|)}{2m(K+1)} \cdot \frac{48\alpha L\|\pi_A\|\|A_\infty\|(s_B + m\|B_\infty\|)}{cm} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] \\
& + \frac{3\alpha^2 L\|\pi_A\|\|A_\infty\|(s_B + m\|B_\infty\|)}{2m(K+1)} \cdot \frac{cm}{48\alpha L\|\pi_A\|\|A_\infty\|(s_B + m\|B_\infty\|)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] \\
\le & \frac{72\alpha^3 L^2\|\pi_A\|^2\|A_\infty\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\mathbf{y}^{(t)}\|^2\right] \\
& + \frac{c\alpha}{32(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right] \\
\le & \frac{144\alpha^3 L^2\|\pi_A\|^2\|A_\infty\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] \\
& + \frac{144n^2\alpha^3 L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}
$$

$$+\frac{c\alpha}{32(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]$$

231 Then we obtain the lemma.

$$\begin{aligned}
&-\frac{\alpha}{mK}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=1}^{m-1}B^{m-1-i}(\mathbf{g}^{(k+i)}-\mathbf{g}^{(k)}),\nabla f(w^{(k)})\right\rangle\right]\\
\leq&\frac{54\alpha L^2\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&+\frac{144\alpha^3L^2\|\pi_A\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]\\
&+\frac{144n^2\alpha^3L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]\\
&+\frac{7c\alpha}{32(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|^2\right]
\end{aligned}$$

232 We finish the proof of the lemma. □

233 5.2 Main Theorem

Theorem 3.

$$\begin{aligned}
&-\frac{\alpha}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)},\nabla f(w^{(k)})\right\rangle\right]\\
\leq&-\frac{c\alpha}{32(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2+\frac{\alpha\|\pi_A\|^2s_B^2}{cm^2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]\\
&-\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]+\frac{c\alpha L^2}{2n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&+\frac{54\alpha L^2\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&+\frac{144\alpha^3L^2\|\pi_A\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]\\
&+\frac{144n^2\alpha^3L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}$$

234 *Proof.* Substitute Lemma 17, 18 and 19 to Lemma 13, we obtain that

$$\begin{aligned}
&-\frac{\alpha}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\left\langle \pi_A^T\sum_{i=0}^{m-1}\mathbf{y}^{(k+i)},\nabla f(w^{(k)})\right\rangle\right]\\
\leq&-\frac{c\alpha}{32(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2+\frac{\alpha\|\pi_A\|^2s_B^2}{cm^2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]\\
&-\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]+\frac{c\alpha L^2}{2n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&+\frac{54\alpha L^2\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]
\end{aligned}$$

$$
+ \frac{144\alpha^3 L^2 \|\pi_A\|^2 \|A_\infty\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
$$

$$
+ \frac{144 n^2 \alpha^3 L^2 \|\pi_A\|^2 \|\pi_B\|^2 \|A_\infty\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
$$

Then we finish the proof of the theorem. □

6 Convergence Analysis and Linear Speedup

6.1 Substitution

Lemma 20. *By setting* $\alpha \le \min\{\frac{1}{cL}, \frac{1}{128cmL}, \frac{1}{25M_B s_B L\|A_\infty\|}, \sqrt{\frac{n}{1360 n s_A^2 s_B^2 M_B^2 L^2 + 8 s_A^2 L^2 \|n\pi_B - \mathbb{1}_n\|^2}}, \frac{-8cL + \sqrt{16c^2L^2 + 2\left(960n\|\pi_A\|^2 M_B^2 s_B^2 L^2 (1+c^2) + \tilde{\mathbf{D}}_\mathbf{2}\right)}}{2\left(960n\|\pi_A\|^2 M_B^2 s_B^2 L^2(1+c^2) + \tilde{\mathbf{D}}_\mathbf{2}\right)}\}$ *and* $m \ge 1$, *we have*

$$
\begin{aligned}
&\frac{c\alpha}{2(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \\
\le& \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{4(m^2\alpha^2 L \mathbf{I_1} + \alpha L \mathbf{H_3})(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{cm^2(K+1)} \\
&+ \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2}\sigma^2 + \frac{2c^2\alpha^2 L}{n}\sigma^2 + \frac{6\alpha^2 L\|\pi_A\|^2 s_B^2}{m}\sigma^2 + \frac{8\alpha^2 s_B^4 L\|\pi_A\|^2}{m}\sigma^2 \\
&+ \frac{20\alpha^2 L\mathbf{H_2}}{m} M_B s_B n\sigma^2 + 320 m^2 c^2 \alpha^4 L^3 M_B s_B \sigma^2 \\
&+ \frac{30000\alpha^3 s_A^2 M_B s_B L\left(cm^3 n\mathbf{H_1} + ncLs_B^2 M_B^2 \mathbf{H_2} + mLs_B^2 M_B^2\right)}{cm^2}\sigma^2 \\
&+ \frac{2(m^2\alpha^3 L\mathbf{I_1} + \alpha^2 L\mathbf{H_3})}{m}\sigma^2 + \frac{2(m^2\alpha^3 L\mathbf{I_1} + \alpha^2 L\mathbf{H_3})}{m}\tilde{\mathbf{D}}_\mathbf{1}\sigma^2
\end{aligned} \tag{22}
$$

Proof. Substitute Theorem 2 and 3 to Lemma 11, we have

$$
\begin{aligned}
&\frac{f(w^{(*)}) - f(w^{(0)})}{m(K+1)} \\
\le& \left(\frac{4c^2\alpha^2 mL}{K+1} - \frac{c\alpha}{32(K+1)}\right) \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2 \\
&+ \frac{\alpha\|\pi_A\|^2 s_B^2}{cm^2(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] + \frac{2c^2\alpha^2 L}{n}\sigma^2 + \frac{6\alpha^2 L\|\pi_A\|^2 s_B^2}{m}\sigma^2 \\
&+ \frac{\alpha^2 s_B^2 L\|\pi_A\|^2}{m(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right] \\
&- \frac{c\alpha}{2(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] + \frac{c\alpha L^2}{2n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&+ \frac{54\alpha L^2\|\pi_A\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&+ \frac{144\alpha^3 L^2\|\pi_A\|^2\|A_\infty\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] \\
&+ \frac{144n^2\alpha^3 L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{8c^2\alpha^2L^3}{n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]+\frac{18\alpha^2L^3\|\pi_A\|^2s_B^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_x^{(t)}\|^2\\
&+\frac{16c^2m\alpha^4L^3}{n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]+\frac{18\alpha^4L^3\|\pi_A\|^2s_B^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_y^{(t)}\|^2\\
&+\frac{18n^2\alpha^4L^3s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\bar{g}^{(t)}\|^2+\frac{16c^4m\alpha^4L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}
$$

241 For $\left(\frac{4c^2\alpha^2mL}{K+1}-\frac{c\alpha}{32(K+1)}\right)\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2$, by setting $\alpha\leq\frac{1}{128cmL}$, we have
242 $\left(\frac{4c^2\alpha^2mL}{K+1}-\frac{c\alpha}{32(K+1)}\right)\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\nabla f(w^{(k)})\|\right]^2\leq 0.$

243 Moving $\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}$ to the left side of inequality, and moving $\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}$ to the right
244 side of inequality, then simplify the remaining terms, we have

$$
\begin{aligned}
&\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\
\leq&\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}\\
&+\frac{\alpha\|\pi_A\|^2s_B^2}{cm^2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]+\frac{2c^2\alpha^2L}{n}\sigma^2+\frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m}\sigma^2\\
&+\frac{\alpha^2s_B^2L\|\pi_A\|^2}{m(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]\\
&+\frac{c\alpha L^2}{2n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&+\frac{54\alpha L^2\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]\\
&+\frac{144\alpha^3L^2\|\pi_A\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]\\
&+\frac{144n^2\alpha^3L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]\\
&+\frac{8c^2\alpha^2L^3}{n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]+\frac{18\alpha^2L^3\|\pi_A\|^2s_B^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_x^{(t)}\|^2\\
&+\frac{16c^2m\alpha^4L^3}{n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]+\frac{18\alpha^4L^3\|\pi_A\|^2s_B^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\Delta_y^{(t)}\|^2\\
&+\frac{18n^2\alpha^4L^3s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\bar{g}^{(t)}\|^2+\frac{16c^4m\alpha^4L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}
$$

245 We denote $\mathbf{g}^{(i)}-\nabla f(\mathbf{x}^{(i)})$ as $\mathbf{G}^{(i)}$, we have

$$
\begin{aligned}
&\frac{\alpha^2s_B^2L\|\pi_A\|^2}{m(K+1)}\sum_{k=0,m,\cdot,mK}\mathbb{E}\left[\|\Delta_y^{(k)}\|^2\right]\\
=&\frac{\alpha^2s_B^2L\|\pi_A\|^2}{m(K+1)}\sum_{k=m,\cdots,mK}\mathbb{E}\left[\|\sum_{i=0}^{k-1}(B^{k-1-i}-B_\infty)(\mathbf{g}^{(i+1)}-\mathbf{g}^{(i)})\|^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&\le \frac{2\alpha^2 s_B^2 L\|\pi_A\|^2}{m(K+1)} \sum_{k=m,\cdots,mK} \mathbb{E}\left[\|\sum_{i=0}^{k-1}(B^{k-1-i}-B_\infty)(\mathbf{G}^{(i+1)}-\mathbf{G}^{(i)})\|^2\right] \\
&\quad + \frac{2\alpha^2 s_B^2 L\|\pi_A\|^2}{m(K+1)} \sum_{k=m,\cdots,mK} \mathbb{E}\left[\|\sum_{i=0}^{k-1}(B^{k-1-i}-B_\infty)(\nabla f(\mathbf{x}^{(i+1)})-\nabla f(\mathbf{x}^{(i)}))\|^2\right] \\
&\le \frac{8\alpha^2 s_B^4 L\|\pi_A\|^2}{m}\sigma^2 + \frac{2\alpha^2 s_B^4 L^3\|\pi_A\|^2}{m(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\mathbf{x}^{(t+1)}-\mathbf{x}^{(t)}\|^2\right] \\
&\le \frac{8\alpha^2 s_B^4 L\|\pi_A\|^2}{m}\sigma^2 + \frac{12\alpha^2 s_B^4 L^3\|\pi_A\|^2}{m(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&\quad + \frac{12\alpha^4 s_B^4 L^3\|\pi_A\|^2\|A_\infty\|^2}{m(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] \\
&\quad + \frac{12n^2\alpha^4 s_B^4 L^3\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{m(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}
$$

246 And

$$
\begin{aligned}
&\frac{\alpha\|\pi_A\|^2 s_B^2}{cm^2(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\Delta_y^{(k)}\|\right]^2 \\
&\le \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2}\sigma^2 + \frac{12\alpha\|\pi_A\|^2 s_B^4 L^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&\quad + \frac{12\alpha^3\|\pi_A\|^2 s_B^4 L^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] \\
&\quad + \frac{12n^2\alpha^3\|\pi_A\|^2\|\pi_B\|^2 s_B^4 L^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]
\end{aligned}
$$

247 So we have that

$$
\begin{aligned}
&\frac{c\alpha}{2(K+1)} \sum_{k=0,m,\cdots,mK} \mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \\
&\le \frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)} + \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2}\sigma^2 + \frac{2c^2\alpha^2 L}{n}\sigma^2 + \frac{6\alpha^2 L\|\pi_A\|^2 s_B^2}{m}\sigma^2 + \frac{8\alpha^2 s_B^4 L\|\pi_A\|^2}{m}\sigma^2 \\
&\quad + \frac{12\alpha\|\pi_A\|^2 s_B^4 L^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{12\alpha^2 s_B^4 L^3\|\pi_A\|^2}{m(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&\quad + \frac{c\alpha L^2}{2n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{54\alpha L^2\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] \\
&\quad + \frac{8c^2\alpha^2 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right] + \frac{18\alpha^2 L^3\|\pi_A\|^2 s_B^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_x^{(t)}\|^2 \\
&\quad + \frac{12\alpha^3\|\pi_A\|^2 s_B^4 L^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] + \frac{12\alpha^4 s_B^4 L^3\|\pi_A\|^2\|A_\infty\|^2}{m(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] \\
&\quad + \frac{16c^2 m\alpha^4 L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right] + \frac{18\alpha^4 L^3\|\pi_A\|^2 s_B^2\|A_\infty\|^2}{K+1} \sum_{t=0}^{m(K+1)} \|\Delta_y^{(t)}\|^2 \\
&\quad + \frac{144\alpha^3 L^2\|\pi_A\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]
\end{aligned}
$$

$$+\frac{12n^2\alpha^3\|\pi_A\|^2\|\pi_B\|^2 s_B^4 L^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]+\frac{12n^2\alpha^4 s_B^4 L^3\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{m(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]$$
$$+\frac{144n^2\alpha^3 L^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]$$
$$+\frac{18n^2\alpha^4 L^3 s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{K+1}\sum_{t=0}^{m(K+1)}\|\bar{g}^{(t)}\|^2+\frac{16c^4 m\alpha^4 L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]$$

248 By setting $\alpha \le \frac{1}{12cmL}$, the coefficient of $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]$ can be simplified to $\frac{\alpha L^2\mathbf{H_1}}{K+1}$, where

$$\mathbf{H_1}=\frac{13\|\pi_A\|^2 s_B^4}{cm^2}+\frac{c}{2n}+\frac{54\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2}+\frac{2c}{3mn}+\frac{3\|\pi_A\|^2 s_B^2}{2cm}$$

249 By setting $\alpha \le \frac{1}{2cmL}$, the coefficient of $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]$ can be simplified to $\frac{\alpha^2 L\mathbf{H_2}}{m^2(K+1)}+$
250 $\frac{16c^2 m\alpha^4 L^3}{n(K+1)}$, where

$$\mathbf{H_2}=\frac{6\|\pi_A\|^2 s_B^4}{c^2 m}+\frac{3s_B^2\|\pi_A\|^2\|A_\infty\|^2}{c^2 m}+\frac{9\|\pi_A\|^2 s_B^2\|A_\infty\|^2}{2c^2}+\frac{72\|\pi_A\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{c^2 m}$$

251 By setting $\alpha \le \frac{1}{2cmL}$, the coefficient of $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]$ can be simplified to $\frac{\alpha^2 L\mathbf{H_3}}{m^2(K+1)}+$
252 $\frac{16c^4 m\alpha^4 L^3}{K+1}$, where

$$\mathbf{H_3}=\frac{6n^2\|\pi_A\|^2\|\pi_B\|^2 s_B^4}{c^2 m}+\frac{3n^2 s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|}{c^2 m}+\frac{72n^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{c^2 m}$$
$$+\frac{9n^2 s_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{2c^2}$$

253 Then we have that

$$\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]$$
$$\le\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2 L}{n}\sigma^2+\frac{6\alpha^2 L\|\pi_A\|^2 s_B^2}{m}\sigma^2+\frac{8\alpha^2 s_B^4 L\|\pi_A\|^2}{m}\sigma^2$$
$$+\frac{\alpha L^2\mathbf{H_1}}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]+\frac{\alpha^2 L\mathbf{H_2}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]$$
$$+\frac{16c^2 m\alpha^4 L^3}{n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]$$
$$+\frac{\alpha^2 L\mathbf{H_3}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]+\frac{16c^4 m\alpha^4 L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]$$

254 Then we substitute $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_y^{(t)}\|^2\right]$ by Lemma 7. And we set $\alpha \le$
255 $\min\{\frac{1}{25M_B s_B L\|A_\infty\|},\frac{1}{cmL}\}$, we have that

$$\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]$$
$$\le\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2 L}{n}\sigma^2+\frac{6\alpha^2 L\|\pi_A\|^2 s_B^2}{m}\sigma^2+\frac{8\alpha^2 s_B^4 L\|\pi_A\|^2}{m}\sigma^2$$
$$+\frac{20\alpha^2 L\mathbf{H_2}}{m}M_B s_B n\sigma^2+320m^2c^2\alpha^4L^3M_B s_B\sigma^2$$

$$+\frac{\alpha m^2 nL\mathbf{H_1}+200nc\alpha^2L^3s_B^2M_B^2\mathbf{H_2}+3200c^2m^3\alpha^4L^5s_B^2M_B^2}{m^2n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t+1)}\|^2\right]$$
$$+\frac{120nc^2\alpha^4L^3s_B^2M_B^2\mathbf{H_2}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]+\frac{1920c^4m\alpha^6L^5s_B^2M_B^2}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$$
$$+\frac{\alpha^2L\mathbf{H_3}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]+\frac{16c^4m\alpha^4L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$$

And the coefficient of $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t+1)}\|^2\right]$ can be simplified to

$$\frac{3200\alpha L\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{cm^3n(K+1)}$$

So we have

$$\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]$$
$$\le\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2L}{n}\sigma^2+\frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m}\sigma^2+\frac{8\alpha^2s_B^4L\|\pi_A\|^2}{m}\sigma^2$$
$$+\frac{20\alpha^2L\mathbf{H_2}}{m}M_Bs_Bn\sigma^2+320m^2c^2\alpha^4L^3M_Bs_B\sigma^2$$
$$+\frac{3200\alpha L\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{cm^3n(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t+1)}\|^2\right]$$
$$+\frac{120nc^2\alpha^4L^3s_B^2M_B^2\mathbf{H_2}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]+\frac{2000c^4m\alpha^6L^5s_B^2M_B^2}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$$
$$+\frac{\alpha^2L\mathbf{H_3}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]+\frac{16c^4m\alpha^4L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$$

Then we substitute $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\Delta_x^{(t)}\|^2\right]$ by Lemma 8. And we set $\alpha\le\min\{\frac{1}{16cmL},\frac{1}{cL}\}$, so we have that

$$\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]$$
$$\le\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2L}{n}\sigma^2+\frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m}\sigma^2+\frac{8\alpha^2s_B^4L\|\pi_A\|^2}{m}\sigma^2$$
$$+\frac{20\alpha^2L\mathbf{H_2}}{m}M_Bs_Bn\sigma^2+320m^2c^2\alpha^4L^3M_Bs_B\sigma^2$$
$$+\frac{30000\alpha^3s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{cm^2}\sigma^2$$
$$+\frac{20000\alpha^3L\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)\left(s_A^2\|n\pi_B-\mathbb{1}_n\|^2+120ns_A^2s_B^2M_B^2\right)}{cm^3n(K+1)}\sum_{t=0}^{m(K+1)}\|\bar g^{(t)}\|^2$$
$$+\frac{120nc^2\alpha^4L^3s_B^2M_B^2\mathbf{H_2}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]+\frac{2000c^4m\alpha^6L^5s_B^2M_B^2}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$$
$$+\frac{\alpha^2L\mathbf{H_3}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]+\frac{16c^4m\alpha^4L^3}{K+1}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$$

By setting $\alpha\le\frac{1}{cmL}$ and $m\ge1$, we can simplify the coefficient of $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\bar g^{(t)}\|^2\right]$ as follows.

$$\frac{20000\alpha^3L\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)\left(s_A^2\|n\pi_B-\mathbb{1}_n\|^2+120ns_A^2s_B^2M_B^2\right)}{cm^3n(K+1)}$$

$$
\begin{aligned}
&+\frac{120nc^2\alpha^4L^3s_B^2M_B^2\mathbf{H_2}}{m^2(K+1)}+\frac{2000c^4m\alpha^6L^5s_B^2M_B^2}{K+1}+\frac{\alpha^2L\mathbf{H_3}}{m^2(K+1)}+\frac{16c^4m\alpha^4L^3}{K+1}\\
\leq&\frac{\alpha^3L\mathbf{I_1}}{K+1}+\frac{\alpha^2L\mathbf{H_3}}{m^2(K+1)}=\frac{m^2\alpha^3L\mathbf{I_1}+\alpha^2L\mathbf{H_3}}{m^2(K+1)}
\end{aligned}
$$

Where

$$
\begin{aligned}
\mathbf{I_1}=&\frac{20000\left(cn\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+Ls_B^2M_B^2\right)\left(s_A^2\|n\pi_B-\mathbb{1}_n\|^2+120ns_A^2s_B^2M_B^2\right)}{cn}\\
&+120ncLs_B^2M_B^2\mathbf{H_2}+2000cLs_B^2M_B^2+16c^3L
\end{aligned}
$$

And we have

$$
\begin{aligned}
&\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\
\leq&\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2L}{n}\sigma^2+\frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m}\sigma^2+\frac{8\alpha^2s_B^4L\|\pi_A\|^2}{m}\sigma^2\\
&+\frac{20\alpha^2L\mathbf{H_2}}{m}M_Bs_Bn\sigma^2+320m^2c^2\alpha^4L^3M_Bs_B\sigma^2\\
&+\frac{30000\alpha^3s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{cm^2}\sigma^2\\
&+\frac{m^2\alpha^3L\mathbf{I_1}+\alpha^2L\mathbf{H_3}}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\|\bar{g}^{(t)}\|^2
\end{aligned}
$$

Since $\mathbb{E}\left[\|\bar{g}^{(t)}\|^2\right]\leq 2\mathbb{E}\left[\|\bar{g}^{(t)}-\overline{\nabla f}^{(t)}\|^2\right]+2\mathbb{E}\left[\|\overline{\nabla f}^{(t)}\|^2\right]\leq 2\sigma^2+2\mathbb{E}\left[\|\overline{\nabla f}^{(t)}\|^2\right]$, we have

$$
\begin{aligned}
&\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\
\leq&\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2L}{n}\sigma^2+\frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m}\sigma^2+\frac{8\alpha^2s_B^4L\|\pi_A\|^2}{m}\sigma^2\\
&+\frac{20\alpha^2L\mathbf{H_2}}{m}M_Bs_Bn\sigma^2+320m^2c^2\alpha^4L^3M_Bs_B\sigma^2\\
&+\frac{30000\alpha^3s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{cm^2}\sigma^2\\
&+\frac{2(m^2\alpha^3L\mathbf{I_1}+\alpha^2L\mathbf{H_3})}{m}\sigma^2+\frac{2(m^2\alpha^3L\mathbf{I_1}+\alpha^2L\mathbf{H_3})}{m^2(K+1)}\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\overline{\nabla f}^{(t)}\|^2\right]
\end{aligned}
$$

Substituting $\sum_{t=0}^{m(K+1)}\mathbb{E}\left[\|\overline{\nabla f}^{(t)}\|^2\right]$ by Theorem 1, we have

$$
\begin{aligned}
&\frac{c\alpha}{2(K+1)}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\
\leq&\frac{f(w^{(0)})-f(w^{(*)})}{m(K+1)}+\frac{8\alpha\|\pi_A\|^2s_B^4}{cm^2}\sigma^2+\frac{2c^2\alpha^2L}{n}\sigma^2+\frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m}\sigma^2+\frac{8\alpha^2s_B^4L\|\pi_A\|^2}{m}\sigma^2\\
&+\frac{20\alpha^2L\mathbf{H_2}}{m}M_Bs_Bn\sigma^2+320m^2c^2\alpha^4L^3M_Bs_B\sigma^2\\
&+\frac{30000\alpha^3s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{cm^2}\sigma^2\\
&+\frac{2(m^2\alpha^3L\mathbf{I_1}+\alpha^2L\mathbf{H_3})}{m}\sigma^2+\frac{2(m^2\alpha^3L\mathbf{I_1}+\alpha^2L\mathbf{H_3})}{m}\mathbf{\tilde{D}_1}\sigma^2\\
&+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})(\nabla f(w^{(0)})-\nabla f(w^{(*)}))}{cm^2(K+1)}
\end{aligned}
$$

We finish the proof of the lemma. □

6.2 Main Theorem

Theorem 4.

$$
\begin{aligned}
&\frac{1}{K+1}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\
\leq&\frac{2(f(w^{(0)})-f(w^{(*)}))}{c\alpha m(K+1)}+\frac{8L(m^2\alpha\mathbf{I_1}+\mathbf{H_3})(\nabla f(w^{(0)})-\nabla f(w^{(*)}))}{c^2m^2(K+1)}\\
&+\frac{16\|\pi_A\|^2s_B^4}{c^2m^2}\sigma^2+\frac{4c\alpha L}{n}\sigma^2+\frac{12\alpha L\|\pi_A\|^2s_B^2}{cm}\sigma^2+\frac{16\alpha s_B^4L\|\pi_A\|^2}{cm}\sigma^2\\
&+\frac{40\alpha L\mathbf{H_2}}{cm}M_Bs_Bn\sigma^2+640m^2c\alpha^3L^3M_Bs_B\sigma^2\\
&+\frac{60000\alpha^2s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{c^2m^2}\sigma^2\\
&+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})}{cm}\sigma^2+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})}{cm}\tilde{\mathbf{D}}_\mathbf{1}\sigma^2\\
\sim&\frac{3(f(w^{(0)})-f(w^{(*)}))}{\sqrt{nT}}+\frac{\sigma^2}{\sqrt{nT}}+\mathbf{J_2}(\frac{1}{T^{\frac{3}{4}}})\sigma^2
\end{aligned}
$$

Proof. Multiple $\frac{2}{c\alpha}$ on both sides of 22, and we have

$$
\begin{aligned}
&\frac{1}{K+1}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right]\\
\leq&\frac{2(f(w^{(0)})-f(w^{(*)}))}{c\alpha m(K+1)}+\frac{8L(m^2\alpha\mathbf{I_1}+\mathbf{H_3})(\nabla f(w^{(0)})-\nabla f(w^{(*)}))}{c^2m^2(K+1)}\\
&+\frac{16\|\pi_A\|^2s_B^4}{c^2m^2}\sigma^2+\frac{4c\alpha L}{n}\sigma^2+\frac{12\alpha L\|\pi_A\|^2s_B^2}{cm}\sigma^2+\frac{16\alpha s_B^4L\|\pi_A\|^2}{cm}\sigma^2\\
&+\frac{40\alpha L\mathbf{H_2}}{cm}M_Bs_Bn\sigma^2+640m^2c\alpha^3L^3M_Bs_B\sigma^2\\
&+\frac{60000\alpha^2s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{c^2m^2}\sigma^2\\
&+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})}{cm}\sigma^2+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})}{cm}\tilde{\mathbf{D}}_\mathbf{1}\sigma^2
\end{aligned}
$$

Consider the coefficient of $\frac{f(w^{(0)})-f(w^{(*)})}{c\alpha m(K+1)}=\frac{f(w^{(0)})-f(w^{(*)})}{c\alpha T}$

$$
\mathbf{J_1}=2+\frac{8\alpha L(m^2\alpha\mathbf{I_1}+\mathbf{H_3})}{cm}
$$

and the coefficient of the non-red term σ^2

$$
\begin{aligned}
\mathbf{J_2}=&\frac{12\alpha L\|\pi_A\|^2s_B^2}{cm}\sigma^2+\frac{16\alpha s_B^4L\|\pi_A\|^2}{cm}\sigma^2\\
&+\frac{40\alpha L\mathbf{H_2}}{cm}M_Bs_Bn\sigma^2+640m^2c\alpha^3L^3M_Bs_B\sigma^2\\
&+\frac{60000\alpha^2s_A^2M_Bs_BL\left(cm^3n\mathbf{H_1}+ncLs_B^2M_B^2\mathbf{H_2}+mLs_B^2M_B^2\right)}{c^2m^2}\sigma^2\\
&+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})}{cm}\sigma^2+\frac{4(m^2\alpha^2L\mathbf{I_1}+\alpha L\mathbf{H_3})}{cm}\tilde{\mathbf{D}}_\mathbf{1}\sigma^2
\end{aligned}
$$

So when $m\geq\frac{4\sqrt{2}\|\pi_A\|s_Bn^{\frac{1}{4}}T^{\frac{1}{4}}}{c}$, we have that $\frac{16\|\pi_A\|^2s_B^4}{c^2m^2}\sigma^2\leq\frac{\sigma^2}{2\sqrt{nT}}$. When $\alpha\leq\frac{\sqrt{n}}{8cL\sqrt{T}}$, we have $\frac{4c\alpha L}{n}\sigma^2\leq\frac{\sigma^2}{2\sqrt{nT}}$. Then we have that $\frac{16\|\pi_A\|^2s_B^4}{c^2m^2}\sigma^2+\frac{4c\alpha L}{n}\sigma^2\leq\frac{\sigma^2}{\sqrt{nT}}$, this is the linear speedup term.

Furthermore, by setting $\frac{4\sqrt{2}\|\pi_A\|s_B n^{\frac{1}{4}} T^{\frac{1}{4}}}{c} \leq m \leq \frac{8\sqrt{2}\|\pi_A\|s_B n^{\frac{1}{4}} T^{\frac{1}{4}}}{c}$, $\mathbf{B_0} = \min\{\frac{1}{cL}, \frac{1}{128cmL}, \frac{1}{25M_B s_B L\|A_\infty\|}, \sqrt{\frac{n}{1360ns_A^2 s_B^2 M_B^2 L^2 + 8s_A^2 L^2\|n\pi_B - \mathbb{1}_n\|^2}}, \frac{-8cL + \sqrt{16c^2L^2 + 2\left(960n\|\pi_A\|^2 M_B^2 s_B^2 L^2(1+c^2) + \tilde{\mathbf{D}}_\mathbf{2}\right)}}{2\left(960n\|\pi_A\|^2 M_B^2 s_B^2 L^2(1+c^2) + \tilde{\mathbf{D}}_\mathbf{2}\right)}\}$, and $0.5\mathbf{B_0} \leq \alpha \leq \mathbf{B_0}$. Since T can be sufficiently large to make $\frac{\sqrt{n}}{8cL\sqrt{T}}$ be the minimum, we have that $\alpha \sim O(\frac{1}{T^{\frac{1}{2}}}), m \sim O(T^{\frac{1}{4}})$. With the help of this,we have that

$$\mathbf{J_1} = 2 + \frac{8\alpha L(m^2\alpha\mathbf{I_1} + \mathbf{H_3})}{cm} \sim 2 + O(\frac{1}{T^{\frac{1}{4}}})$$

so $\mathbf{J_1}$ have a constant upper bound 3. And we have that

$$\begin{aligned}
\mathbf{J_2} =& \frac{12\alpha L\|\pi_A\|^2 s_B^2}{cm}\sigma^2 + \frac{16\alpha s_B^4 L\|\pi_A\|^2}{cm}\sigma^2 \\
&+ \frac{40\alpha L\mathbf{H_2}}{cm} M_B s_B n\sigma^2 + 640m^2 c\alpha^3 L^3 M_B s_B \sigma^2 \\
&+ \frac{60000\alpha^2 s_A^2 M_B s_B L\left(cm^3 n\mathbf{H_1} + ncLs_B^2 M_B^2 \mathbf{H_2} + mLs_B^2 M_B^2\right)}{c^2m^2}\sigma^2 \\
&+ \frac{4(m^2\alpha^2 L\mathbf{I_1} + \alpha L\mathbf{H_3})}{cm}\sigma^2 + \frac{4(m^2\alpha^2 L\mathbf{I_1} + \alpha L\mathbf{H_3})}{cm}\tilde{\mathbf{D}}_\mathbf{1}\sigma^2 \\
\sim& O(\frac{1}{T^{\frac{3}{4}}})\sigma^2
\end{aligned}$$

So we obtain the main theorem

$$\begin{aligned}
&\frac{1}{K+1}\sum_{k=0,m,\cdots,mK}\mathbb{E}\left[\|\overline{\nabla f}^{(k)}\|^2\right] \\
\leq& \frac{2(f(w^{(0)}) - f(w^{(*)}))}{c\alpha m(K+1)} + \frac{8L(m^2\alpha\mathbf{I_1} + \mathbf{H_3})(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c^2m^2(K+1)} \\
&+ \frac{16\|\pi_A\|^2 s_B^4}{c^2m^2}\sigma^2 + \frac{4c\alpha L}{n}\sigma^2 + \frac{12\alpha L\|\pi_A\|^2 s_B^2}{cm}\sigma^2 + \frac{16\alpha s_B^4 L\|\pi_A\|^2}{cm}\sigma^2 \\
&+ \frac{40\alpha L\mathbf{H_2}}{cm} M_B s_B n\sigma^2 + 640m^2 c\alpha^3 L^3 M_B s_B \sigma^2 \\
&+ \frac{60000\alpha^2 s_A^2 M_B s_B L\left(cm^3 n\mathbf{H_1} + ncLs_B^2 M_B^2 \mathbf{H_2} + mLs_B^2 M_B^2\right)}{c^2m^2}\sigma^2 \\
&+ \frac{4(m^2\alpha^2 L\mathbf{I_1} + \alpha L\mathbf{H_3})}{cm}\sigma^2 + \frac{4(m^2\alpha^2 L\mathbf{I_1} + \alpha L\mathbf{H_3})}{cm}\tilde{\mathbf{D}}_\mathbf{1}\sigma^2 \\
\sim& \frac{3(f(w^{(0)}) - f(w^{(*)}))}{\sqrt{nT}} + \frac{\sigma^2}{\sqrt{nT}} + \mathbf{J_2}(\frac{1}{T^{\frac{3}{4}}})\sigma^2
\end{aligned}$$

$\square$