
New Proof

Anonymous Author(s)
Affiliation
Address
email

1	Contents	
2	1 Notations.	2
3	2 Analysis: Basic	2
4	2.1 Rolling Sum Lemma	2
5	2.2 Basic Transformation	3
6	2.3 Technical Lemmas	3
7	2.4 Gradient Consensus lemma	7
8	2.5 Consensus Lemma 1	9
9	2.6 Consensus Lemma 2	9
10	2.7 Descent Lemma: Basic 1	11
11	2.8 Main Theorem: Basic 1	12
12	2.9 Descent Lemma: Basic 2	14
13	2.10 Main Theorem: Basic 2	15
14	3 Convergence Analysis: Quadratic Term	18
15	3.1 Decomposition	18
16	3.2 Technical Lemmas	18
17	3.3 Main Theorem	22
18	4 Convergence Analysis: Inner Product Term	23
19	4.1 Decomposition	23
20	4.2 Technical Lemmas	23
21	4.3 Main Theorem	26
22	5 Convergence Analysis and Linear Speedup	27
23	5.1 Analysis	27
24	5.2 Substitution	28
25	5.3 Main Theorem	33

26 1 Notations.

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

$$28 \mathbf{x}^{(k)} = [(x_1^{(k)})^\top; (x_2^{(k)})^\top; \dots; (x_n^{(k)})^\top] \in \mathbb{R}^{n \times d}$$

$$29 \nabla F(\mathbf{x}^{(k)}; \boldsymbol{\xi}^{(k)}) := [\nabla F_1(x_1^{(k)}; \xi_1^{(k)})^\top; \dots; \nabla F_n(x_n^{(k)}; \xi_n^{(k)})^\top] \in \mathbb{R}^{n \times d}$$

$$30 w^{(k)} = \pi_A^T \mathbf{x}^{(k)}, \mathbf{w}^{(k)} = A_\infty \mathbf{x}^{(k)}$$

$$31 \bar{x} = \frac{1}{n} \mathbf{1}_n^T \mathbf{x}, \bar{\mathbf{x}} = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \mathbf{x}$$

$$32 \Delta_x^{(k)} = \mathbf{x}^{(k)} - \mathbf{w}^{(k)}$$

$$33 \Delta_y^{(k)} = \mathbf{y}^{(k)} - B_\infty \mathbf{y}^{(k)} = (I - B_\infty) \mathbf{y}^{(k)}$$

$$34 \Delta_g^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$$

$$35 \bar{y} = \frac{1}{n} \mathbf{1}_n^T \mathbf{y}, \bar{\mathbf{y}} = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \mathbf{y}$$

$$36 \nabla \bar{\mathbf{f}}_k = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \nabla \mathbf{f}(\mathbf{x}_k)$$

37 2 Analysis: Basic

38 2.1 Rolling Sum Lemma

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

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

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

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

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

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

43 *Proof.* First, we prove that

$$\|A^i - A_\infty\|_2 \leq \sqrt{\kappa(\pi_A)} \beta_A^i, \forall i \geq 0. \quad (4)$$

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

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

45 we have

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

46 which proves (4).

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

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

48 Towards this end, we define $M_A := \max_{k \geq 0} \|A^k - A_\infty\|_2$. M_A is well-defined because of

49 (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 $\|A^i - A_\infty\|_2 \leq$

50 $\min\{M_A, M_A \beta_A^{i-p}\}$. With this inequality, we can bound $\sum_{i=0}^k \|A^{k-i} - A_\infty\|_2$ as follows:

$$\sum_{i=0}^k \|A^{k-i} - A_\infty\|_2 = \sum_{i=0}^{\min\{\lfloor p \rfloor, k\}} \|A^i - A_\infty\|_2 + \sum_{i=\min\{\lfloor p \rfloor, k\}+1}^k \|A^i - A_\infty\|_2$$

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

51 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
52 $p > 0$, let $x = \min\{\lfloor p \rfloor, k\} + 1 - p \in [0, 1)$, (5) is simplified to

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

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

54 Finally, to obtain (1), we use Jensen's inequality. For positive numbers $a_i, i \in [k]$ satisfying
55 $\sum_{i=0}^k a_i = 1$, we have

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

56 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

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

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

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

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

60 □

61 2.2 Basic Transformation

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

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

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

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

66 4. $\lim_{m \rightarrow +\infty} (\sum_{i=0}^m B^i - m B_\infty) \cdot (I - B) = I - B_\infty$. [lly: Do we need this?]

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

68 2.3 Technical Lemmas

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

$$\|\Delta_y^{(k+1)}\|_F^2 = \sum_{i=0}^k \|(B - B_\infty)^{k-i} (I - B_\infty) \Delta_g^{(i)}\|_F^2$$

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

70 *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
 71 obtain that

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

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

Lemma 3.

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

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

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

75 For the first part, we have that

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

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

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

78 Note that

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

79 we can apply Cauchy's inequality and obtain that

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

80 So we obtain the lemma

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

81 \square

Lemma 4.

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

Proof. Notice that

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

82 and the basic transformation $(B - B_\infty)^{k-i}(I - B_\infty) = (B^{k-i} - B_\infty)(I - B_\infty) = B^{k-i} - B_\infty$,
83 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
84 product.

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

85 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
86 bounded by the Cauchy-Schwarz inequality as follows

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

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:

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

87 so we obtain that

$$\begin{aligned}
& \mathbb{E} \left[\left\langle (B - B_\infty)^{k-i+1} \Delta_y^{(i)}, (B - B_\infty)^{k-i} (\nabla f^{(i+1)} - \nabla f^{(i)}) \right\rangle \right] \tag{10} \\
& \leq \alpha L b_{k-i} \mathbb{E} \left[\|A_\infty \mathbf{y}^{(i)}\| \|\Delta_y^{(i)}\| \right] + L b_{k-i} \mathbb{E} \left[\|\mathbf{x}^{(i+1)} - \mathbf{w}^{(i+1)}\| \|\Delta_y^{(i)}\| \right] \\
& \quad + L b_{k-i} \mathbb{E} \left[\|\mathbf{x}^{(i)} - \mathbf{w}^{(i)}\| \|\Delta_y^{(i)}\| \right]
\end{aligned}$$

88 By Young inequality, we can further bound 10 as

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

89 For the second term decomposed from $\mathbb{E} \left[\left\langle (B - B_\infty)^{k-i+1} \Delta_y^{(i)}, (B - B_\infty)^{k-i} (I - B_\infty) \Delta_g^{(i)} \right\rangle \right]$,

90 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

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

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

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

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

93 The above expression can be reduced to

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

94 Combine 11 and 12, we obtain the lemma. \square

95 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
96 size of $\sum_{k=l}^T c_{k-l}$.

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

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

98 *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),
99 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
100 verify that $b_i \leq \min M_B^2, M_B^2 \beta_B^{2i+1-p}, \forall i \geq 0$. With this inequality, we can bound $\sum_{k=l}^T b_{k-l}$ as
101 follows:

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

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

Lemma 6.

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

103 *Proof.* Notice that

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

104 We substitute Lemma 5 in Lemma 4, and obtain that

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

105 So we obtain the lemma

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

106 □

107 2.4 Gradient Consensus lemma

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

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

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

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

110 Noting that $A_\infty \mathbf{y}^{(k)} = A_\infty B_\infty \mathbf{y}^{(k)} + A_\infty (I - B_\infty) \mathbf{y}^{(k)} = c \mathbf{1}_n \bar{g}^{(k)} + A_\infty \Delta_y^{(k)}$, we have
 111 $\|A_\infty \mathbf{y}^{(k)}\|_F^2 \leq 2c^2 \|\mathbf{1}_n \bar{g}^{(k)}\|_F^2 + 2\|A_\infty\|_2^2 \|\Delta_y^{(k)}\|_F^2 = 2nc^2 \|\bar{g}^{(k)}\|^2 + 2\|A_\infty\|_2^2 \|\Delta_y^{(k)}\|_F^2$, so we
 112 have

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

113 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

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

114 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
 115 is equivalent to the following inequality

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

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

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

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

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

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

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

118 and the inequality

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

119 We finish the proof of the lemma. □

120 2.5 Consensus Lemma 1

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

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

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

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

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

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

125 By further applying Lemma 7 in 15, we have

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

By setting

$$\alpha \leq \min\left\{\frac{1}{20s_A s_B M_B L}, \frac{1}{25M_B s_B L \|A_\infty\|}\right\}$$

126 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
 127 proof. \square

128 2.6 Consensus Lemma 2

129 **Lemma 9.** By setting $\alpha \leq \min\{\frac{1}{cL}, \frac{1}{25M_B s_B L \|A_\infty\|}, \sqrt{\frac{n}{1360ns_A^2 s_B^2 M_B^2 L^2 + 8s_A^2 L^2 \|n\pi_B - \mathbb{1}_n\|^2}}\}$, we
 130 have

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

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

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

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

$$\begin{aligned} & \sum_{k=0}^T \|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2 \\ &\leq 2\alpha^2 \sum_{k=0}^T \left\| \sum_{i=0}^k (A^{k-i} - A_\infty)(I - B_\infty) \mathbf{y}^{(i)} \right\|^2 + 2\alpha^2 \sum_{k=0}^T \left\| \sum_{i=0}^k (A^{k-i} - A_\infty)(n\pi_B^T - \mathbf{1}_n) \bar{\mathbf{y}}^{(i)} \right\|^2 \\ &\leq 2\alpha^2 s_A^2 \sum_{k=0}^T \|\Delta_y^{(k)}\|^2 + 2\alpha^2 s_A^2 \|n\pi_B^T - \mathbf{1}_n\|^2 \sum_{k=0}^T \|\bar{\mathbf{g}}^{(k)}\|^2 \end{aligned} \quad (17)$$

134 By further applying Lemma 7 in 17, we have

$$\begin{aligned} & (1 - 400\alpha^2 s_A^2 s_B^2 M_B^2 L^2) \sum_{k=0}^T \|\mathbf{x}^{(k+1)} - \mathbf{w}^{(k+1)}\|^2 \\ &\leq (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240nc^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2) \sum_{k=0}^T \|\bar{\mathbf{g}}^{(k)}\|^2 \\ &\quad + 40n\alpha^2 s_A^2 M_B s_B (T+1) \sigma^2 \end{aligned} \quad (18)$$

135 Noting that $\mathbb{E} [\|\bar{\mathbf{g}}^k\|^2] \leq 2\sigma^2 + \frac{4L^2}{n} \mathbb{E} [\|\Delta_x^{(k)}\|^2] + 4\mathbb{E} [\|\nabla f(w^{(k)})\|^2]$, we have

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

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

$$\begin{aligned} & 400\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{4L^2}{n} (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240nc^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2) \\ &\leq 400\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{4L^2}{n} (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240n\alpha^2 s_A^2 s_B^2 M_B^2) \\ &= 1360\alpha^2 s_A^2 s_B^2 M_B^2 L^2 + \frac{8\alpha^2 s_A^2 L^2 \|n\pi_B - \mathbf{1}_n\|^2}{n} \end{aligned}$$

137 Since

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

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

138 So we can obtain the upper bound of α

$$\alpha \leq \sqrt{\frac{n}{1360ns_A^2 s_B^2 M_B^2 L^2 + 8s_A^2 L^2 \|n\pi_B - \mathbf{1}_n\|^2}}$$

139 and we obtain the lemma.

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

140

□

141 2.7 Descent Lemma: Basic 1

Lemma 10.

$$\begin{aligned} &\frac{1}{T+1} \sum_{k=0}^T \mathbb{E} [\|\nabla f(w^{(k)})\|^2] \\ &\leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + \frac{4}{T+1} \left(2c\alpha L - \frac{1}{2}\right) \sum_{k=0}^T \mathbb{E} [\|\overline{\nabla f}^{(k)}\|^2] + 8c\alpha L\sigma^2 \\ &\quad + \frac{4L^2}{n(T+1)} \sum_{k=0}^T \mathbb{E} [\|\Delta_x^{(k)}\|^2] + \frac{4\|\pi_A\|^2}{c\alpha(T+1)} \left(c^2\alpha^2 L + \frac{\alpha}{\textcolor{red}{c}}\right) \sum_{k=0}^T \mathbb{E} [\|\Delta_y^{(k)}\|^2] \end{aligned}$$

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

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

143 Taking conditional expectation, we have

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

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

$$\begin{aligned} &\mathbb{E} [f(w^{(k+1)})] - \mathbb{E} [f(w^{(k)})] \\ &\leq -c\alpha \mathbb{E} [\left\langle \bar{g}^{(k)}, \nabla f(w^{(k)}) \right\rangle] - \alpha \mathbb{E} [\left\langle \pi_A^T \Delta_y^{(k)}, \nabla f(w^{(k)}) \right\rangle] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \\ &= -c\alpha \mathbb{E} [\left\langle \overline{\nabla f}^{(k)}, \nabla f(w^{(k)}) \right\rangle] - \alpha \mathbb{E} [\left\langle \pi_A^T \Delta_y^{(k)}, \nabla f(w^{(k)}) \right\rangle] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \\ &\leq -\frac{c\alpha}{2} \mathbb{E} [\|\overline{\nabla f}^{(k)}\|^2] - \frac{c\alpha}{2} \mathbb{E} [\|\nabla f(w^{(k)})\|^2] + \frac{c\alpha}{2} \mathbb{E} [\|\overline{\nabla f}^{(k)} - \nabla f(w^{(k)})\|^2] \\ &\quad + \frac{\alpha}{\textcolor{red}{c}} \mathbb{E} [\|\pi_A^T \Delta_y^{(k)}\|^2] + \frac{\textcolor{red}{c}\alpha}{4} \mathbb{E} [\|\nabla f(w^{(k)})\|^2] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \\ &= -\frac{c\alpha}{2} \mathbb{E} [\|\overline{\nabla f}^{(k)}\|^2] - \frac{c\alpha}{4} \mathbb{E} [\|\nabla f(w^{(k)})\|^2] + \frac{c\alpha}{2} \mathbb{E} [\|\overline{\nabla f}^{(k)} - \nabla f(w^{(k)})\|^2] \\ &\quad + \frac{\alpha}{\textcolor{red}{c}} \mathbb{E} [\|\pi_A^T \Delta_y^{(k)}\|^2] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \end{aligned}$$

145 Notice that

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

146 we can obtain that

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

147 Further noticing that $\|\pi_A^T \mathbf{y}^{(k)}\|^2 \leq 2c^2 \|\bar{g}^{(k)}\|^2 + 2\|\pi_A^T \Delta_y^{(k)}\|^2$, we have

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

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

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

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

$$\begin{aligned} & \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} \left[\|\nabla f(w^{(k)})\|^2 \right] \\ & \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + \frac{4}{T+1} \left(2c\alpha L - \frac{1}{2} \right) \sum_{k=0}^T \mathbb{E} \left[\|\nabla f^{(k)}\|^2 \right] + 8c\alpha L \sigma^2 \\ & \quad + \frac{4L^2}{n(T+1)} \sum_{k=0}^T \mathbb{E} \left[\|\Delta_x^{(k)}\|^2 \right] + \frac{4\|\pi_A\|^2}{c\alpha(T+1)} \left(c^2 \alpha^2 L + \frac{\alpha}{c} \right) \sum_{k=0}^T \mathbb{E} \left[\|\Delta_y^{(k)}\|^2 \right] \end{aligned}$$

150 We finish the proof of this lemma. \square

151 2.8 Main Theorem: Basic 1

152 **Theorem 1.** By setting $\alpha \leq \min\left\{ \frac{1}{cL}, \frac{1}{25M_B s_B L \|A_\infty\|}, \frac{1}{20s_A s_B M_B L}, \frac{-4cL + \sqrt{16c^2 L^2 + 4\tilde{\mathbf{C}}_2}}{2\tilde{\mathbf{C}}_2} \right\}$, we
153 have

$$\begin{aligned} \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} \left[\|\nabla f(w^{(k)})\|^2 \right] & \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + (\mathbf{C}_1(1) + 2\mathbf{C}_2(\alpha^2)) \sigma^2 \\ & \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + (\tilde{\mathbf{C}}_1 + 2\alpha^2 \tilde{\mathbf{C}}_2) \sigma^2 \end{aligned}$$

154 Where

$$\begin{aligned} \mathbf{C}_1(1) & = \left(8c\alpha L + 80nM_B s_B \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) + \left(4L^2 + 800ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) \cdot 80\alpha^2 s_A^2 M_B s_B \\ \mathbf{C}_2(\alpha^2) & = \left(4L^2 + 800ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) \cdot 240c^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2 + 480ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c^3 \alpha^3 L + \alpha^2) \end{aligned}$$

155 and

$$\tilde{\mathbf{C}}_1 := \left(8 + 80nM_B s_B \|\pi_A\|^2 (1 + \frac{1}{c^2}) \right) + \left(4\frac{1}{c^2} + 800ns_B^2 M_B^2 \|\pi_A\|^2 \frac{c^2 + 1}{c^4} \right) \cdot 80s_A^2 M_B s_B$$

$$\tilde{\mathbf{C}}_2 := \left(4L^2 + 800ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (cL + \frac{1}{c^2}) \right) \cdot 240s_A^2 s_B^2 M_B^2 + 480ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c^2 + 1)$$

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

$$\begin{aligned} & \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} [\|\nabla f(w^{(k)})\|^2] \\ & \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + \frac{4}{T+1} \left(2c\alpha L - \frac{1}{2} \right) \sum_{k=0}^T \mathbb{E} [\|\bar{\nabla} f^{(k)}\|^2] \\ & \quad + \left(8c\alpha L + 80nM_B s_B \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) \sigma^2 \\ & \quad + \left(\frac{4L^2}{n(T+1)} + \frac{800s_B^2 M_B^2 L^2 \|\pi_A\|^2}{T+1} (c\alpha L + \frac{1}{c^2}) \right) \sum_{k=0}^T \mathbb{E} [\|\Delta_x^{(k)}\|^2] \\ & \quad + \frac{480ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c^3 \alpha^3 L + \alpha^2)}{T+1} \sum_{k=0}^T \mathbb{E} [\|\bar{g}^{(k)}\|^2] \end{aligned}$$

157 Substitute $\sum_{k=0}^T \mathbb{E} [\|\Delta_x^{(k)}\|^2]$ by Lemma 8, we have

$$\begin{aligned} & \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} [\|\nabla f(w^{(k)})\|^2] \\ & \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + \frac{4}{T+1} \left(2c\alpha L - \frac{1}{2} \right) \sum_{k=0}^T \mathbb{E} [\|\bar{\nabla} f^{(k)}\|^2] \\ & \quad + \mathbf{C}_1 \sigma^2 + \frac{\mathbf{C}_2}{T+1} \sum_{k=0}^T \mathbb{E} [\|\bar{g}^{(k)}\|^2] \end{aligned}$$

158 Where

$$\begin{aligned} \mathbf{C}_1(1) &= \left(8c\alpha L + 80nM_B s_B \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) \\ & \quad + \left(4L^2 + 800ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) \cdot 80\alpha^2 s_A^2 M_B s_B \end{aligned}$$

159 and

$$\begin{aligned} \mathbf{C}_2(\alpha^2) &= \left(4L^2 + 800ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c\alpha L + \frac{1}{c^2}) \right) \cdot 240c^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2 \\ & \quad + 480ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c^3 \alpha^3 L + \alpha^2) \end{aligned}$$

160 Since $\mathbb{E} [\|\bar{g}^{(k)}\|^2] \leq 2\sigma^2 + 2\mathbb{E} [\|\bar{\nabla} f^{(k)}\|^2]$, we have

$$\begin{aligned} & \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} [\|\nabla f(w^{(k)})\|^2] \\ & \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + \frac{4}{T+1} \left(2c\alpha L - \frac{1}{2} + \frac{\mathbf{C}_2(\alpha^2)}{2} \right) \sum_{k=0}^T \mathbb{E} [\|\bar{\nabla} f^{(k)}\|^2] \\ & \quad + (\mathbf{C}_1(1) + 2\mathbf{C}_2(\alpha^2)) \sigma^2 \end{aligned}$$

161 Since $2c\alpha L - \frac{1}{2} + \frac{\mathbf{C}_2(\alpha^2)}{2} \leq 0 \iff 2c\alpha L + \frac{\mathbf{C}_2(\alpha^2)}{2} \leq \frac{1}{2}$, and use the upper bound $\alpha \leq \frac{1}{cL}$

$$\mathbf{C}_2(\alpha^2) \leq \alpha^2 \tilde{\mathbf{C}}_2$$

$$\tilde{\mathbf{C}}_2 := \left(4L^2 + 800ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (cL + \frac{1}{c^2})\right) \cdot 240s_A^2 s_B^2 M_B^2 + 480ns_B^2 M_B^2 L^2 \|\pi_A\|^2 (c^2 + 1)$$

162 Then we can obtain the upper bound of α

$$2c\alpha L + \frac{1}{2}\alpha^2 \tilde{\mathbf{C}}_2 \leq \frac{1}{2} \iff \alpha \leq \frac{-4cL + \sqrt{16c^2 L^2 + 4\tilde{\mathbf{C}}_2}}{2\tilde{\mathbf{C}}_2}$$

163 and obtain the lemma

$$\frac{1}{T+1} \sum_{k=0}^T \mathbb{E} [\|\nabla f(w^{(k)})\|^2] \leq \frac{4(f(w^{(0)}) - f(w^*))}{c\alpha(T+1)} + (\mathbf{C}_1(1) + 2\mathbf{C}_2(\alpha^2)) \sigma^2$$

164 We can also use the bound $c\alpha L \leq 1$ to simplify the upper bound of $\mathbf{C}_1(1), \mathbf{C}_2(\alpha^2)$

$$\begin{aligned} \mathbf{C}_1(1) &\leq \tilde{\mathbf{C}}_1 := \left(8 + 80nM_B s_B \|\pi_A\|^2 (1 + \frac{1}{c^2})\right) \\ &\quad + \left(4\frac{1}{c^2} + 800ns_B^2 M_B^2 \|\pi_A\|^2 \frac{c^2 + 1}{c^4}\right) \cdot 80s_A^2 M_B s_B \\ \mathbf{C}_2(\alpha^2) &\leq \alpha^2 \tilde{\mathbf{C}}_2 \end{aligned}$$

165

□

166 2.9 Descent Lemma: Basic 2

Lemma 11.

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

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

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

168 Taking conditional expectation, we have

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

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

$$\begin{aligned} &\mathbb{E} [f(w^{(k+1)})] - \mathbb{E} [f(w^{(k)})] \\ &\leq -c\alpha \mathbb{E} [\left\langle \bar{g}^{(k)}, \nabla f(w^{(k)}) \right\rangle] - \alpha \mathbb{E} [\left\langle \pi_A^T \Delta_y^{(k)}, \nabla f(w^{(k)}) \right\rangle] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \\ &= -c\alpha \mathbb{E} [\left\langle \bar{\nabla} f^{(k)}, \nabla f(w^{(k)}) \right\rangle] - \alpha \mathbb{E} [\left\langle \pi_A^T \Delta_y^{(k)}, \nabla f(w^{(k)}) \right\rangle] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \\ &\leq -\frac{c\alpha}{2} \mathbb{E} [\|\bar{\nabla} f^{(k)}\|^2] - \frac{c\alpha}{2} \mathbb{E} [\|\nabla f(w^{(k)})\|^2] + \frac{c\alpha}{2} \mathbb{E} [\|\bar{\nabla} f^{(k)} - \nabla f(w^{(k)})\|^2] \\ &\quad + \frac{\alpha}{c} \mathbb{E} [\|\pi_A^T \Delta_y^{(k)}\|^2] + \frac{c\alpha}{4} \mathbb{E} [\|\nabla f(w^{(k)})\|^2] + \frac{\alpha^2 L}{2} \mathbb{E} [\|\pi_A^T \mathbf{y}^{(k)}\|^2] \\ &= -\frac{c\alpha}{2} \mathbb{E} [\|\bar{\nabla} f^{(k)}\|^2] - \frac{c\alpha}{4} \mathbb{E} [\|\nabla f(w^{(k)})\|^2] + \frac{c\alpha}{2} \mathbb{E} [\|\bar{\nabla} f^{(k)} - \nabla f(w^{(k)})\|^2] \end{aligned}$$

$$+ \frac{\alpha}{\textcolor{red}{c}} \mathbb{E} \left[\|\pi_A^T \Delta_y^{(k)}\|^2 \right] + \frac{\alpha^2 L}{2} \mathbb{E} \left[\|\pi_A^T \mathbf{y}^{(k)}\|^2 \right]$$

170 Notice that

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

171 we can obtain that

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

172 Further noticing that $\|\pi_A^T \mathbf{y}^{(k)}\|^2 \leq 2c^2 \|\bar{g}^{(k)}\|^2 + 2\|\pi_A^T \Delta_y^{(k)}\|^2$, we have

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

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

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

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

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

175 We finish the proof of this lemma. \square

176 2.10 Main Theorem: Basic 2

177 **Theorem 2.** By setting $\alpha \leq \min \left\{ \frac{1}{cL}, \frac{1}{25M_B s_B L \|A_\infty\|}, \sqrt{\frac{n}{1360n s_A^2 s_B^2 M_B^2 L^2 + 8s_A^2 L^2 \|n\pi_B - \mathbf{1}_n\|^2}}, \right.$
 178 $\left. \frac{-8cL + \sqrt{16c^2 L^2 + 2(960n \|\pi_A\|^2 M_B^2 s_B^2 L^2 (1+c^2) + \tilde{\mathbf{D}}_2)}}{2(960n \|\pi_A\|^2 M_B^2 s_B^2 L^2 (1+c^2) + \tilde{\mathbf{D}}_2)} \right\}$, we have

$$\begin{aligned} \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} \left[\|\overline{\nabla f}^{(k)}\|^2 \right] & \leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \mathbf{D}_1(1)\sigma^2 \\ & \leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \tilde{\mathbf{D}}_1 \sigma^2 \end{aligned}$$

179 *Where*

$$\begin{aligned}
\tilde{\mathbf{D}}_1 := & 4 + 2n\|\pi_A\|^2 M_B s_B \left(\frac{1}{c^2} + 1 \right) + 480n\|\pi_A\|^2 M_B^2 s_B^2 \left(\frac{1}{c^2} + 1 \right) \\
& + 160ns_A^2 M_B s_B \left(\frac{2}{nc^2} + 200\|\pi_A\|^2 M_B^2 s_B^2 \frac{c^2 + 1}{c^4} \right) + 7680n\|\pi_A\|^2 s_A^2 M_B^3 s_B^3 \frac{c^2 + 1}{c^4} \\
& + 8 \left(\frac{5}{nc^2} + 200\|\pi_A\|^2 M_B^2 s_B^2 \frac{c^2 + 1}{c^4} \right) (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \\
& + 3840\|\pi_A\|^2 M_B^2 s_B^2 \frac{c^2 + 1}{c^4} (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \\
\tilde{\mathbf{D}}_2 = & 16 \left(\frac{5L^2}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2 \left(\frac{1}{c^2} + 1 \right) \right) (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \\
& + 7680\|\pi_A\|^2 M_B^2 s_B^2 \left(\frac{L^2}{c^2} + L^2 \right) (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2)
\end{aligned}$$

180 *Proof.* Substitute $\sum_{k=0}^T \mathbb{E} [\|\Delta_y^{(k)}\|^2]$ in Lemma 11 by Lemma 7, we have

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

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

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

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

$$\frac{1}{T+1} \sum_{k=0}^T \mathbb{E} [\|\nabla f^{(k)}\|^2] \leq \frac{2}{T+1} \left(480n\|\pi_A\|^2 M_B^2 s_B^2 (\alpha^2 L^2 + c^3 \alpha^3 L^3) + \frac{\mathbf{D}_2(\alpha^2)}{2} + 4c\alpha L - \frac{1}{4} \right) \sum_{k=0}^T \mathbb{E} [\|\nabla f(w^{(k)})\|^2]$$

$$+ \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \mathbf{D}_1(1)\sigma^2$$

183 Where

$$\begin{aligned} \mathbf{D}_1(1) = & 4c\alpha L + 2n\|\pi_A\|^2 M_B s_B \left(\frac{1}{c^2} + c\alpha L \right) + 480n\|\pi_A\|^2 M_B^2 s_B^2 (\alpha^2 L^2 + c^3 \alpha^3 L^3) \\ & + 160n\alpha^2 s_A^2 M_B s_B \left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2 \left(\frac{1}{c^2} + c\alpha L \right) \right) \\ & + 8 \left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2 \left(\frac{1}{c^2} + c\alpha L \right) \right) (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240nc^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2) \\ & + 7680n\|\pi_A\|^2 s_A^2 M_B^3 s_B^3 (\alpha^4 L^4 + c^3 \alpha^5 L^5) \\ & + 3840\|\pi_A\|^2 M_B^2 s_B^2 (\alpha^2 L^4 + c^3 \alpha^3 L^5) (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240nc^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2) \end{aligned}$$

184 and

$$\begin{aligned} \mathbf{D}_2(\alpha^2) = & 16 \left(\frac{L^2}{n} + \frac{4c\alpha L^3}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2 \left(\frac{1}{c^2} + c\alpha L \right) \right) (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240nc^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2) \\ & + 7680\|\pi_A\|^2 M_B^2 s_B^2 (\alpha^2 L^4 + c^3 \alpha^3 L^5) (2\alpha^2 s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240nc^2 \alpha^4 s_A^2 s_B^2 M_B^2 L^2) \end{aligned}$$

185 Use the bound $c\alpha L \leq 1$, we have that

$$\begin{aligned} \mathbf{D}_1(1) \leq & \tilde{\mathbf{D}}_1 := 4 + 2n\|\pi_A\|^2 M_B s_B \left(\frac{1}{c^2} + 1 \right) + 480n\|\pi_A\|^2 M_B^2 s_B^2 \left(\frac{1}{c^2} + 1 \right) \\ & + 160ns_A^2 M_B s_B \left(\frac{2}{nc^2} + 200\|\pi_A\|^2 M_B^2 s_B^2 \frac{c^2 + 1}{c^4} \right) + 7680n\|\pi_A\|^2 s_A^2 M_B^3 s_B^3 \frac{c^2 + 1}{c^4} \\ & + 8 \left(\frac{5}{nc^2} + 200\|\pi_A\|^2 M_B^2 s_B^2 \frac{c^2 + 1}{c^4} \right) (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \\ & + 3840\|\pi_A\|^2 M_B^2 s_B^2 \frac{c^2 + 1}{c^4} (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \\ \mathbf{D}_2(\alpha^2) \leq & \alpha^2 \tilde{\mathbf{D}}_2 \end{aligned}$$

186 and

$$\begin{aligned} \tilde{\mathbf{D}}_2 = & 16 \left(\frac{5L^2}{n} + 200\|\pi_A\|^2 M_B^2 s_B^2 L^2 \left(\frac{1}{c^2} + 1 \right) \right) (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \\ & + 7680\|\pi_A\|^2 M_B^2 s_B^2 \left(\frac{L^2}{c^2} + L^2 \right) (2s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 240ns_A^2 s_B^2 M_B^2) \end{aligned}$$

187 Since

$$480n\|\pi_A\|^2 M_B^2 s_B^2 (\alpha^2 L^2 + c^3 \alpha^3 L^3) + \frac{\mathbf{D}_2(\alpha^2)}{2} \leq 480n\|\pi_A\|^2 M_B^2 s_B^2 \alpha^2 L^2 (1 + c^2) + \frac{\alpha^2 \tilde{\mathbf{D}}_2}{2}$$

188 So we have that

$$\begin{aligned} & 960n\|\pi_A\|^2 M_B^2 s_B^2 \alpha^2 L^2 (1 + c^2) + \alpha^2 \tilde{\mathbf{D}}_2 + 8c\alpha L - \frac{1}{2} \leq 0 \\ \iff \alpha \leq & \frac{-8cL + \sqrt{16c^2 L^2 + 2 \left(960n\|\pi_A\|^2 M_B^2 s_B^2 L^2 (1 + c^2) + \tilde{\mathbf{D}}_2 \right)}}{2 \left(960n\|\pi_A\|^2 M_B^2 s_B^2 L^2 (1 + c^2) + \tilde{\mathbf{D}}_2 \right)} \end{aligned}$$

189 So apply the upper bound we obtain the lemma.

$$\begin{aligned} \frac{1}{T+1} \sum_{k=0}^T \mathbb{E} \left[\|\nabla f^{(k)}\|^2 \right] & \leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \mathbf{D}_1(1)\sigma^2 \\ & \leq \frac{2(\nabla f(w^{(0)}) - \nabla f(w^{(*)}))}{c\alpha(T+1)} + \tilde{\mathbf{D}}_1\sigma^2 \end{aligned}$$

190

□

191 3 Convergence Analysis: Quadratic Term

192 3.1 Decomposition

Lemma 12.

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

193 *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
194 term can be decomposed as follows.

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

195 Since $\sum_{i=0}^{m-1} \pi_A^T (I - B_\infty) \mathbf{y}^{(k+i)} = \pi_A^T \left(\sum_{i=0}^{m-1} B^i - mB_\infty \right) \mathbf{y}^{(k)} + \pi_A^T \sum_{i=0}^{m-1} (B^{m-1-i} - B_\infty)$
196 $B_\infty) (\mathbf{g}^{(k+i)} - \mathbf{g}^{(k)})$, we have

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

197 We finish the proof of the lemma. □

198 3.2 Technical Lemmas

Lemma 13.

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

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

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

200 Based on the independence in the expectation calculation, we have

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

$$\leq \frac{2c^2\alpha^2mL}{n} \cdot \sigma^2$$

201 So we have

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

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

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

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

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

204 Noting that

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

205 then we obtain the lemma.

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

206 We finish the proof of the lemma. \square

Lemma 14.

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

207 *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

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

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

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

209 We finish the proof of the lemma. \square

Lemma 15.

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

210 *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
211 denoted as $\mathbf{G}^{(k)} = \mathbf{g}^{(k)} - \nabla f(\mathbf{x}^{(k)})$, taking conditional expectation, we have

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

212 Based on the independence in the expectation calculation, we have

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

213 And we have

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

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

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

215 Noticing that

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

216 Since

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

217 Then we have

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

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

218 We finish the proof of the lemma. \square

219 3.3 Main Theorem

Theorem 3.

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

220 *Proof.* Substitute Lemma ?,? and ? to Lemma ?, we obtain that

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

221 We finish the proof of the theorem. \square

222 4 Convergence Analysis: Inner Product Term

223 4.1 Decomposition

Lemma 16.

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

224 *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

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

225 taking conditional expectation, we obtain the lemma. \square

226 4.2 Technical Lemmas

Lemma 17.

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

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

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

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

$$-\frac{\alpha}{m(K+1)} \sum_{k=0, m, \dots, mK} \mathbb{E} \left[\left\langle \pi_A^T \left(\sum_{i=0}^{m-1} B^i - B_\infty \right) \mathbf{y}^{(k)}, \nabla f(w^{(k)}) \right\rangle \right]$$

$$\leq \frac{c\alpha}{4(K+1)} \sum_{k=0,m,\dots,mK} \mathbb{E} \left[\|\nabla f(w^{(k)})\|^2 \right] + \frac{\alpha \|\pi_A\|^2 s_B^2}{cm^2(K+1)} \sum_{k=0,m,\dots,mK} \mathbb{E} \left[\|\Delta_y^{(k)}\|^2 \right]$$

229 We finish the proof of the lemma. \square

Lemma 18.

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

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

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

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

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

232 We finish the proof of the lemma. \square

Lemma 19.

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

233 *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

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

234 Noting that

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

235 and that

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

236 and that

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

237 Then we obtain the lemma.

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

238 We finish the proof of the lemma. □

239 4.3 Main Theorem

Theorem 4.

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

$$\begin{aligned}
& + \frac{54\alpha L^2 \|\pi_A\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_x^{(t)}\|^2 \right] \\
& + \frac{144\alpha^3 L^2 \|\pi_A\|^2 \|A_\infty\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_y^{(t)}\|^2 \right] \\
& + \frac{144n^2 \alpha^3 L^2 \|\pi_A\|^2 \|\pi_B\|^2 \|A_\infty\|^2 (s_B + m\|B_\infty\|)^2}{cm^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right]
\end{aligned}$$

240 *Proof.* Substitute Lemma ?,? and ? to Lemma ?, we obtain that

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

241 Then we finish the proof of the theorem. \square

242 5 Convergence Analysis and Linear Speedup

243 5.1 Analysis

Lemma 20.

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

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

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

246 Taking conditional expectation, we have

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

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

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

248 Then we finish the proof of the lemma. \square

249 5.2 Substitution

250 **Lemma 21.** *With many const upper bound for α , we have*

$$\begin{aligned} & \frac{c\alpha}{2(K + 1)} \sum_{k=0, m, \dots, mK} \mathbb{E} \left[\left\| \nabla f^{(k)} \right\|^2 \right] \\ & \leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K + 1)} + \frac{4\alpha L \mathbf{I}_2(1)(f(w^{(0)}) - f(w^{(*)}))}{cm^2(K + 1)} + \frac{4\alpha^2 L^2 \mathbf{I}_1(1)(f(w^{(0)}) - f(w^{(*)}))}{c(K + 1)} \\ & \quad + \frac{8\alpha \|\pi_A\|^2 s_B^4}{cm^2} \sigma^2 + \frac{2c^2 \alpha^2 L}{n} \sigma^2 + \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m} \sigma^2 + \frac{8\alpha^2 s_B^4 L \|\pi_A\|^2}{m} \sigma^2 \\ & \quad + 16c^2 m^2 \alpha^4 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \alpha^2 mn L \mathbf{H}_2\left(\frac{1}{m^2}\right) (20s_B^2 + 8M_B s_B) \sigma^2 \\ & \quad + \frac{2\alpha^3 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2\left(\frac{1}{m^2}\right) + 1) (40s_B^2 + 16M_B s_B)}{cm} \sigma^2 \\ & \quad + 2m\alpha^3 L^2 \mathbf{I}_1(1) \sigma^2 + \frac{2\alpha^2 L \mathbf{I}_2(1)}{m} \sigma^2 + 2m\alpha^3 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1) \sigma^2 + \frac{2\alpha^2 L \mathbf{I}_2(1) \mathbf{D}_1(1)}{m} \sigma^2 \end{aligned}$$

251 *Proof.* Substitute Theorem ? and ? to Lemma ?, we have

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

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

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

253 $\left(\frac{4c^2\alpha^2mL}{K+1} - \frac{c\alpha}{32(K+1)} \right) \sum_{k=0,m,\dots,mK} \mathbb{E} \left[\|\nabla f(w^{(k)})\|^2 \right] \leq 0.$

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

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

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

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

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

257 And

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

258 So we have that

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

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

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

$$\begin{aligned}
\frac{\alpha L^2 \mathbf{H}_1(1)}{K+1} &= \frac{13\|\pi_A\|^2s_B^4L^2}{cm^2(K+1)} + \frac{cL^2}{2n(K+1)} + \frac{54L^2\|\pi_A\|^2(s_B+m\|B_\infty\|)^2}{cm^2(K+1)} \\
&+ \frac{2cL^2}{3mn(K+1)} + \frac{3L^2\|\pi_A\|^2s_B^2}{2cm(K+1)}
\end{aligned}$$

260 By setting $\alpha \leq \frac{1}{2cmL}$, the coefficient of $\sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_y^{(t)}\|^2 \right]$ can be simplified to:

$$\begin{aligned}
\frac{\alpha^2 L \mathbf{H}_2(\frac{1}{m^2})}{K+1} + \frac{16c^2m\alpha^4L^3}{n(K+1)} &= \frac{6\|\pi_A\|^2s_B^4L}{c^2m^3(K+1)} + \frac{3s_B^2L\|\pi_A\|^2\|A_\infty\|^2}{c^2m^3(K+1)} + \frac{16c^2m\alpha^4L^3}{n(K+1)} \\
&+ \frac{9L\|\pi_A\|^2s_B^2\|A_\infty\|^2}{2c^2m^2(K+1)} + \frac{72L\|\pi_A\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{c^2m^3(K+1)}
\end{aligned}$$

261 By setting $\alpha \leq \frac{1}{2cmL}$, the coefficient of $\sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right]$ can be simplified to:

$$\begin{aligned}
\frac{\alpha^2 L \mathbf{H}_3(\frac{1}{m^2})}{K+1} + \frac{16c^4m\alpha^4L^3}{K+1} &= \frac{6n^2\|\pi_A\|^2\|\pi_B\|^2s_B^4L}{c^2m^3(K+1)} + \frac{3n^2s_B^2L\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{c^2m^3(K+1)} \\
&+ \frac{72n^2L\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2(s_B+m\|B_\infty\|)^2}{c^2m^3(K+1)} \\
&+ \frac{9n^2Ls_B^2\|\pi_A\|^2\|\pi_B\|^2\|A_\infty\|^2}{2c^2m^2(K+1)} + \frac{16c^4m\alpha^4L^3}{K+1}
\end{aligned}$$

262 Where the expression inside the parentheses denote the order of m . Then we have that

$$\begin{aligned}
& \frac{c\alpha}{2(K+1)} \sum_{k=0,m,\dots,mK} \mathbb{E} \left[\|\nabla f^{(k)}\|^2 \right] \\
& \leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{8\alpha\|\pi_A\|^2s_B^4}{cm^2} \sigma^2 + \frac{2c^2\alpha^2L}{n} \sigma^2 + \frac{6\alpha^2L\|\pi_A\|^2s_B^2}{m} \sigma^2 + \frac{8\alpha^2s_B^4L\|\pi_A\|^2}{m} \sigma^2 \\
& + \frac{\alpha L^2 \mathbf{H}_1(1)}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_x^{(t)}\|^2 \right] + \frac{\alpha^2 L \mathbf{H}_2(\frac{1}{m^2})}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_y^{(t)}\|^2 \right] \\
& + \frac{16c^2m\alpha^4L^3}{n(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_y^{(t)}\|^2 \right] \\
& + \frac{\alpha^2 L \mathbf{H}_3(\frac{1}{m^2})}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right] + \frac{16c^4m\alpha^4L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right]
\end{aligned}$$

263 Then we substitute $\sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_y^{(t)}\|^2 \right]$ by Gradient Consensus Lemma. And we set $\alpha \leq$

264 $\min \left\{ \frac{1}{2s_BcmL\sqrt{5+20M_B^2}}, \frac{1}{cmLs_B^2(20+160M_B^2)}, \frac{1}{\sqrt[3]{16s_B^2(20+160M_B^2)}}, \frac{1}{\sqrt[4]{64s_B^2(5+20M_B^2)}} \right\}$, we have that

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

$$\begin{aligned}
&\leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2} \sigma^2 + \frac{2c^2\alpha^2 L}{n} \sigma^2 + \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m} \sigma^2 + \frac{8\alpha^2 s_B^4 L \|\pi_A\|^2}{m} \sigma^2 \\
&\quad + 16c^2 m^2 \alpha^4 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \alpha^2 mn L \mathbf{H}_2\left(\frac{1}{m^2}\right) (20s_B^2 + 8M_B s_B) \sigma^2 \\
&\quad + \frac{\alpha L^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1)}{cm^2 n (K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\Delta_x^{(t)}\|^2 \right] \\
&\quad + \frac{\alpha^2 L (m^3 \mathbf{H}_3(\frac{1}{m^2}) + mn \mathbf{H}_2(\frac{1}{m^2}) + 1)}{m^3 (K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right] \\
&\quad + \frac{16c^4 m \alpha^4 L^3}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right]
\end{aligned}$$

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

$$\begin{aligned}
&\frac{c\alpha}{2(K+1)} \sum_{k=0, m, \dots, mK} \mathbb{E} \left[\|\nabla f^{(k)}\|^2 \right] \\
&\leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2} \sigma^2 + \frac{2c^2\alpha^2 L}{n} \sigma^2 + \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m} \sigma^2 + \frac{8\alpha^2 s_B^4 L \|\pi_A\|^2}{m} \sigma^2 \\
&\quad + 16c^2 m^2 \alpha^4 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \alpha^2 mn L \mathbf{H}_2\left(\frac{1}{m^2}\right) (20s_B^2 + 8M_B s_B) \sigma^2 \\
&\quad + \frac{2\alpha^3 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{cm} \sigma^2 \\
&\quad + \frac{\alpha^3 L^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1)}{cm^2 n (K+1)} (4s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 16n\alpha^2 c^2 s_B^4 L^2 (5 + 20M_B^2)) \sum_{t=0}^{m(T+1)} \|\bar{g}^{(t)}\|^2 \\
&\quad + \frac{\alpha^2 L (m^3 \mathbf{H}_3(\frac{1}{m^2}) + mn \mathbf{H}_2(\frac{1}{m^2}) + 1)}{m^3 (K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right] \\
&\quad + \frac{c^3 \alpha^3 L^2}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right]
\end{aligned}$$

267 We simplify the coefficient of $\sum_{t=0}^{m(K+1)} \mathbb{E} \left[\|\bar{g}^{(t)}\|^2 \right]$ as follows.

$$\begin{aligned}
&\frac{\alpha^3 L^2 \mathbf{I}_1(1)}{K+1} + \frac{\alpha^2 L \mathbf{I}_2(1)}{m^2 (K+1)} \\
&= \frac{\alpha^3 L^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1)}{cm^2 n (K+1)} (4s_A^2 \|n\pi_B - \mathbf{1}_n\|^2 + 16n\alpha^2 c^2 s_B^4 L^2 (5 + 20M_B^2)) \\
&\quad + \frac{\alpha^2 L (m^3 \mathbf{H}_3(\frac{1}{m^2}) + mn \mathbf{H}_2(\frac{1}{m^2}) + 1)}{m^3 (K+1)} + \frac{c^3 \alpha^3 L^2}{K+1}
\end{aligned}$$

268 And we have

$$\begin{aligned}
&\frac{c\alpha}{2(K+1)} \sum_{k=0, m, \dots, mK} \mathbb{E} \left[\|\nabla f^{(k)}\|^2 \right] \\
&\leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2} \sigma^2 + \frac{2c^2\alpha^2 L}{n} \sigma^2 + \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m} \sigma^2 + \frac{8\alpha^2 s_B^4 L \|\pi_A\|^2}{m} \sigma^2 \\
&\quad + 16c^2 m^2 \alpha^4 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \alpha^2 mn L \mathbf{H}_2\left(\frac{1}{m^2}\right) (20s_B^2 + 8M_B s_B) \sigma^2 \\
&\quad + \frac{2\alpha^3 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{cm} \sigma^2
\end{aligned}$$

$$+ \frac{\alpha^3 L^2 \mathbf{I}_1(1)}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} [\|\bar{g}^{(t)}\|^2] + \frac{\alpha^2 L \mathbf{I}_2(1)}{m^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} [\|\bar{g}^{(t)}\|^2]$$

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

$$\begin{aligned} & \frac{c\alpha}{2(K+1)} \sum_{k=0, m, \dots, mK} \mathbb{E} [\|\nabla f^{(k)}\|^2] \\ & \leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2} \sigma^2 + \frac{2c^2 \alpha^2 L}{n} \sigma^2 + \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m} \sigma^2 + \frac{8\alpha^2 s_B^4 L \|\pi_A\|^2}{m} \sigma^2 \\ & \quad + 16c^2 m^2 \alpha^4 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \alpha^2 mn L \mathbf{H}_2\left(\frac{1}{m^2}\right) (20s_B^2 + 8M_B s_B) \sigma^2 \\ & \quad + \frac{2\alpha^3 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{cm} \sigma^2 \\ & \quad + 2m\alpha^3 L^2 \mathbf{I}_1(1) \sigma^2 + \frac{2\alpha^2 L \mathbf{I}_2(1)}{m} \sigma^2 \\ & \quad + \frac{2\alpha^3 L^2 \mathbf{I}_1(1)}{K+1} \sum_{t=0}^{m(K+1)} \mathbb{E} [\|\nabla f^{(t)}\|^2] + \frac{2\alpha^2 L \mathbf{I}_2(1)}{m^2(K+1)} \sum_{t=0}^{m(K+1)} \mathbb{E} [\|\nabla f^{(t)}\|^2] \end{aligned}$$

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

$$\begin{aligned} & \frac{c\alpha}{2(K+1)} \sum_{k=0, m, \dots, mK} \mathbb{E} [\|\nabla f^{(k)}\|^2] \\ & \leq \frac{f(w^{(0)}) - f(w^{(*)})}{m(K+1)} + \frac{4\alpha L \mathbf{I}_2(1) (f(w^{(0)}) - f(w^{(*)}))}{cm^2(K+1)} + \frac{4\alpha^2 L^2 \mathbf{I}_1(1) (f(w^{(0)}) - f(w^{(*)}))}{c(K+1)} \\ & \quad + \frac{8\alpha\|\pi_A\|^2 s_B^4}{cm^2} \sigma^2 + \frac{2c^2 \alpha^2 L}{n} \sigma^2 + \frac{6\alpha^2 L \|\pi_A\|^2 s_B^2}{m} \sigma^2 + \frac{8\alpha^2 s_B^4 L \|\pi_A\|^2}{m} \sigma^2 \\ & \quad + 16c^2 m^2 \alpha^4 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \alpha^2 mn L \mathbf{H}_2\left(\frac{1}{m^2}\right) (20s_B^2 + 8M_B s_B) \sigma^2 \\ & \quad + \frac{2\alpha^3 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{cm} \sigma^2 \\ & \quad + 2m\alpha^3 L^2 \mathbf{I}_1(1) \sigma^2 + \frac{2\alpha^2 L \mathbf{I}_2(1)}{m} \sigma^2 + 2m\alpha^3 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1) \sigma^2 + \frac{2\alpha^2 L \mathbf{I}_2(1) \mathbf{D}_1(1)}{m} \sigma^2 \end{aligned}$$

271

□

272 5.3 Main Theorem

Theorem 5.

$$\begin{aligned} & \frac{1}{K+1} \sum_{k=0, m, \dots, mK} \mathbb{E} [\|\nabla f^{(k)}\|^2] \\ & \leq \frac{2(f(w^{(0)}) - f(w^{(*)}))}{c\alpha m(K+1)} + \frac{8L \mathbf{I}_2(1) (f(w^{(0)}) - f(w^{(*)}))}{c^2 m^2(K+1)} + \frac{8\alpha L^2 \mathbf{I}_1(1) (f(w^{(0)}) - f(w^{(*)}))}{c^2(K+1)} \\ & \quad + \frac{16\|\pi_A\|^2 s_B^4}{c^2 m^2} \sigma^2 + \frac{4c\alpha L}{n} \sigma^2 + \frac{12\alpha L \|\pi_A\|^2 s_B^2}{cm} \sigma^2 + \frac{16\alpha s_B^4 L \|\pi_A\|^2}{cm} \sigma^2 \\ & \quad + 32cm^2 \alpha^3 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \frac{2\alpha mn L \mathbf{H}_2(\frac{1}{m^2}) (20s_B^2 + 8M_B s_B)}{c} \sigma^2 \\ & \quad + \frac{4\alpha^2 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{c^2 m} \sigma^2 \\ & \quad + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1)}{cm} \sigma^2 + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1) \mathbf{D}_1(1)}{cm} \sigma^2 \end{aligned}$$

$$\sim \frac{f(w^{(0)}) - f(w^{(*)})}{\sqrt{nT}} + \frac{\sigma^2}{\sqrt{nT}} + \mathbf{J}_2 \left(\frac{1}{T^{\frac{3}{4}}} \right) \sigma^2$$

273 *Proof.* Multiple $\frac{2}{c\alpha}$ on both side of ?, and we have

$$\begin{aligned} & \frac{1}{K+1} \sum_{k=0, m, \dots, mK} \mathbb{E} \left[\|\nabla f^{(k)}\|^2 \right] \\ & \leq \frac{2(f(w^{(0)}) - f(w^{(*)}))}{c\alpha m(K+1)} + \frac{8L\mathbf{I}_2(1)(f(w^{(0)}) - f(w^{(*)}))}{c^2 m^2(K+1)} + \frac{8\alpha L^2 \mathbf{I}_1(1)(f(w^{(0)}) - f(w^{(*)}))}{c^2(K+1)} \\ & \quad + \frac{16\|\pi_A\|^2 s_B^4}{c^2 m^2} \sigma^2 + \frac{4c\alpha L}{n} \sigma^2 + \frac{12\alpha L \|\pi_A\|^2 s_B^2}{cm} \sigma^2 + \frac{16\alpha s_B^4 L \|\pi_A\|^2}{cm} \sigma^2 \\ & \quad + 32cm^2 \alpha^3 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \frac{2\alpha mn L \mathbf{H}_2(\frac{1}{m^2})(20s_B^2 + 8M_B s_B)}{c} \sigma^2 \\ & \quad + \frac{4\alpha^2 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{c^2 m} \sigma^2 \\ & \quad + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1)}{cm} \sigma^2 + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1) \mathbf{D}_1(1)}{cm} \sigma^2 \end{aligned}$$

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

$$\mathbf{J}_1 = 2 + \frac{8\alpha L \mathbf{I}_2(1)}{cm} + \frac{8m\alpha^2 L^2 \mathbf{I}_1(1)}{c}$$

275 Consider the coefficient of the non-red term σ^2

$$\begin{aligned} \mathbf{J}_2 &= \frac{12\alpha L \|\pi_A\|^2 s_B^2}{cm} \sigma^2 + \frac{16\alpha s_B^4 L \|\pi_A\|^2}{cm} \sigma^2 \\ & \quad + 32cm^2 \alpha^3 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \frac{2\alpha mn L \mathbf{H}_2(\frac{1}{m^2})(20s_B^2 + 8M_B s_B)}{c} \sigma^2 \\ & \quad + \frac{4\alpha^2 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{c^2 m} \sigma^2 \\ & \quad + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1)}{cm} \sigma^2 + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1) \mathbf{D}_1(1)}{cm} \sigma^2 \end{aligned}$$

276 So when $m \geq \frac{4\sqrt{2}\|\pi_A\| s_B n^{\frac{1}{4}} T^{\frac{1}{4}}}{c}$, we have that $\frac{16\|\pi_A\|^2 s_B^4}{c^2 m^2} \sigma^2 \leq \frac{\sigma^2}{2\sqrt{nT}}$. When $\alpha \leq \frac{\sqrt{n}}{8cL\sqrt{T}}$, we have
 277 $\frac{4c\alpha L}{n} \sigma^2 \leq \frac{\sigma^2}{2\sqrt{nT}}$. Then we have that $\frac{16\|\pi_A\|^2 s_B^4}{c^2 m^2} \sigma^2 + \frac{4c\alpha L}{n} \sigma^2 \leq \frac{\sigma^2}{\sqrt{nT}}$, this is the linear speedup
 278 term.

279 Furthermore, by setting $\frac{4\sqrt{2}\|\pi_A\| s_B n^{\frac{1}{4}} T^{\frac{1}{4}}}{c} \leq m \leq \frac{8\sqrt{2}\|\pi_A\| s_B n^{\frac{1}{4}} T^{\frac{1}{4}}}{c}$, and $0.5 \min\{\text{many terms}\} \leq$
 280 $\alpha \leq \min\{\text{many terms}\}$. Since T can be sufficiently large to make $\frac{\sqrt{n}}{8cL\sqrt{T}}$ be the minimum, we have
 281 that $\alpha \sim O(\frac{1}{T^{\frac{1}{2}}})$, $m \sim O(T^{\frac{1}{4}})$. With help of this, we have that

$$\mathbf{J}_1 = 2 + \frac{8\alpha L \mathbf{I}_2(1)}{cm} + \frac{8m\alpha^2 L^2 \mathbf{I}_1(1)}{c} \sim 2 + O\left(\frac{1}{T^{\frac{1}{4}}}\right)$$

282 so \mathbf{J}_1 have a const upper bound 3. And we have that

$$\begin{aligned} \mathbf{J}_2 &= \frac{12\alpha L \|\pi_A\|^2 s_B^2}{cm} \sigma^2 + \frac{16\alpha s_B^4 L \|\pi_A\|^2}{cm} \sigma^2 \\ & \quad + 32cm^2 \alpha^3 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \frac{2\alpha mn L \mathbf{H}_2(\frac{1}{m^2})(20s_B^2 + 8M_B s_B)}{c} \sigma^2 \\ & \quad + \frac{4\alpha^2 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{c^2 m} \sigma^2 \\ & \quad + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1)}{cm} \sigma^2 + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1) \mathbf{D}_1(1)}{cm} \sigma^2 \end{aligned}$$

$$\sim O\left(\frac{1}{T^{\frac{3}{4}}}\right)\sigma^2$$

283 So we obtain the main theorem

$$\begin{aligned} & \frac{1}{K+1} \sum_{k=0, m, \dots, mK} \mathbb{E} \left[\|\bar{\nabla} f^{(k)}\|^2 \right] \\ & \leq \frac{2(f(w^{(0)}) - f(w^{(*)}))}{c\alpha m(K+1)} + \frac{8L\mathbf{I}_2(1)(f(w^{(0)}) - f(w^{(*)}))}{c^2 m^2(K+1)} + \frac{8\alpha L^2 \mathbf{I}_1(1)(f(w^{(0)}) - f(w^{(*)}))}{c^2(K+1)} \\ & \quad \frac{16\|\pi_A\|^2 s_B^4}{c^2 m^2} \sigma^2 + \frac{4c\alpha L}{n} \sigma^2 + \frac{12\alpha L \|\pi_A\|^2 s_B^2}{cm} \sigma^2 + \frac{16\alpha s_B^4 L \|\pi_A\|^2}{cm} \sigma^2 \\ & \quad + 32cm^2 \alpha^3 L^3 (20s_B^2 + 8M_B s_B) \sigma^2 + \frac{2\alpha mn L \mathbf{H}_2(\frac{1}{m^2})(20s_B^2 + 8M_B s_B)}{c} \sigma^2 \\ & \quad + \frac{4\alpha^2 L^2 s_A^2 (cm^2 n \mathbf{H}_1(1) + nm \mathbf{H}_2(\frac{1}{m^2}) + 1) (40s_B^2 + 16M_B s_B)}{c^2 m} \sigma^2 \\ & \quad + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1)}{cm} \sigma^2 + \frac{4m\alpha^2 L^2 \mathbf{I}_1(1) \mathbf{D}_1(1)}{c} \sigma^2 + \frac{4\alpha L \mathbf{I}_2(1) \mathbf{D}_1(1)}{cm} \sigma^2 \\ & \sim \frac{f(w^{(0)}) - f(w^{(*)})}{\sqrt{nT}} + \frac{\sigma^2}{\sqrt{nT}} + \mathbf{J}_2\left(\frac{1}{T^{\frac{3}{4}}}\right) \sigma^2 \end{aligned}$$

284

□