A Doubly Recursive Stochastic Compositional Gradient Descent Method for Federated Multi-Level Compositional Optimization

Hongchang Gao [1]

Abstract

Federated compositional optimization has been actively studied in the past few years. However, existing methods mainly focus on the two-level compositional optimization problem, which cannot be directly applied to the multi-level counterparts. Moreover, the convergence rate of existing federated two-level compositional optimization learning algorithms fails to achieve linear speedup with respect to the number of workers under heterogeneous settings. After identifying the reason for this failure, we developed a novel federated stochastic multi-level compositional optimization algorithm by introducing a novel Jacobian-vector product estimator. This innovation mitigates both the heterogeneity issue and the communication efficiency issue simultaneously. We then theoretically proved that our algorithm can achieve the level-independent and linear speedup convergence rate for nonconvex problems. To our knowledge, this is the first time that a federated learning algorithm can achieve such a favorable convergence rate for multi-level compositional problems. Moreover, experimental results confirm the efficacy of our algorithm.

1. Introduction

In this paper, we consider the federated stochastic multi-level compositional optimization (FedSMCO) problem:

$$\min_{x\in\mathbb{R}^{d_0}} \Phi(x) = \frac{1}{N}\sum_{n=1}^{N} F_n^{(K)}\left(\cdots\left(\frac{1}{N}\sum_{n=1}^{N} F_n^{(1)}(x)\right)\cdots\right), \tag{1}$$

where $F_n^{(k)}(\cdot) = \mathbb{E}[F_n^{(k)}(\cdot;\xi_n^{(k)})] : \mathbb{R}^{d_{k-1}} \to \mathbb{R}^{d_k}$ is the k-th level function of the n-th worker for $k \in \{1, 2, \cdots, K\}$ and $n \in \{1, 2, \cdots, N\}$. Throughout this paper, it is assumed that $F_n^{(k)}(\cdot)$ is a *non-linear* function. The stochastic multi-level compositional optimization (SMCO) problem has a wide range of applications, such as multi-step model-agnostic meta-learning (Finn et al., 2017; Chen et al., 2020), the stochastic training of graph neural networks (Yu et al., 2022; Balasubramanian et al., 2022), and risk-averse portfolio optimization (Yang et al., 2019), *etc*.

In the past few years, there have been increasing efforts (Yang et al., 2019; Balasubramanian et al., 2022; Jiang et al., 2022; Chen et al., 2020; Zhang & Xiao, 2019b) to solve the stochastic multi-level compositional optimization problem. A significant challenge in solving SMCO problems lies in that each level function can introduce bias when computed on random samples, which could lead to a slow convergence rate. In particular, (Yang et al., 2019) shows that the standard stochastic multi-level compositional gradient descent algorithm's convergence rate has exponential dependence on the number of levels K, which is referred to as the *level-dependent* convergence rate. Later, (Balasubramanian et al., 2022) developed a multi-level nested linearized averaging algorithm, which improves the convergence rate from exponential to polynomial dependence on K, when certain constants are ignored. Such a convergence rate is referred to as the *level-independent* convergence rate in the literature (Balasubramanian et al., 2022). Recently, (Zhang & Xiao, 2019b) and (Jiang et al., 2022) leveraged the variance reduction techniques (Fang et al., 2018; Nguyen et al., 2017; Cutkosky & Orabona, 2019) for estimating both inner-level functions and gradients to accelerate the convergence rate, whose sample complexity can match traditional non-compositional optimization algorithms (Fang et al., 2018; Nguyen et al., 2017; Cutkosky & Orabona, 2019). However, these algorithms restrict their focus solely on the single-machine setting, which cannot be applied to the federated learning setting.

To solve the stochastic compositional optimization problem on distributed data, a series of distributed stochastic compositional optimization algorithms have been developed. In particular, (Gao & Huang, 2021) introduced the first distributed stochastic compositional gradient descent algorithm under the decentralized setting for stochastic *two-level* com-

[1]Department of Computer and Information Sciences, Temple University, Philadelphia, USA. Correspondence to: Hongchang Gao <hongchang.gao@temple.edu>.

Proceedings of the 41^st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

positional optimization problems and established its convergence rate for nonconvex problems. Subsequently, various federated optimization algorithms (Huang & Li, 2021; Gao et al., 2022; Tarzanagh et al., 2022; Wang et al., 2023; Guo et al., 2023; Wu et al., 2023; Huang, 2022) were developed for the two-level compositional optimization problem. However, all these existing algorithms' convergence rate **fail to achieve linear speedup** regarding the number of workers N under *the heterogeneous setting*, which cannot match federated optimization algorithms (Yu et al., 2019; Yang et al., 2021) for the standard non-compositional optimization problem. Moreover, the distributed *multi-level* compositional optimization has received less attention. Only a recent work (Gao, 2024) developed the decentralized stochastic multi-level compositional optimization algorithm under the homogeneous setting, which enjoys the level-independent convergence rate but fails to achieve linear speedup regarding the number of workers N. Thus, a natural question arises: **Is it possible to develop a federated optimization algorithm for the stochastic multi-level compositional optimization problem in Eq. (1), which can enjoy the level-independent convergence rate and achieve linear speedup regarding the number of workers simultaneously under the heterogeneous setting?**

In this paper, we provide an affirmative answer to this question. Specifically, we first identify the reason for the failure to achieve linear speedup under the heterogeneous setting. More specifically, our theoretical analysis reveals that *the **locally** computed Jacobian matrix is the primary obstacle to achieving a linear speedup convergence rate under the heterogeneous setting*. Even worse, it makes the algorithm converge only to the neighborhood of the optimal solution. Therefore, a natural approach to address this issue is to communicate Jacobian matrices to mitigate heterogeneity. However, communicating Jacobian matrices is practically prohibitive due to their large dimensionality. For instance, for the k-th level function $F_n^{(k)}(\cdot)$ where $k \in \{1, \cdots, K-1\}$, the dimensionality of its Jacobian matrix is $d_k \times d_{k-1}$, which can incur high communication costs when d_k and d_{k-1} are large. To address this issue, we developed a novel *federated stochastic doubly recursive multi-level compositional gradient descent* algorithm. Specifically, instead of communicating Jacobian matrices, our algorithm proposes to communicate the Jacobian-vector product, which is recursively computed across levels. As a result, our algorithm can reduce communication costs significantly. Meanwhile, our algorithm employs the variance reduction technique to estimate the *inner-level function* and the *Jacobian-vector product* to achieve the level-independent and linear speedup convergence rate. However, the Jacobian-vector product, particularly when coupled with the variance reduction technique, introduces significant challenges for convergence analysis since it introduces additional recursive dependence across levels for consensus errors and other estimation errors. In our paper, we addressed these theoretical challenges and established the theoretical convergence rate of our algorithm. Specifically, our algorithm can achieve $O(\frac{1}{N^{2/3}T^{2/3}})$ convergence rate for nonconvex problems, which indicates linear speedup with respect to the number of workers. To the best of our knowledge, this is the first work to achieve the linear speedup convergence rate for federated multi-level compositional optimization problems under the heterogeneous setting. Finally, we conducted extensive experiments to verify the performance of our algorithm and the results confirm the efficacy of our novel ideas. In summary, we made the following contributions in this paper:

- We developed a novel federated stochastic multi-level gradient descent algorithm under the heterogeneous setting by introducing a novel Jacobian-vector product estimator, which can significantly reduce communication costs by communicating the Jacobian-vector product rather than Jacobian matrices.
- We established the convergence rate of our algorithm by addressing the significant challenges caused by the recursively computed Jacobian-vector product, theoretically achieving the level-independent and linear speedup convergence rate.
- Extensive experimental results validate the efficacy of our algorithm.

2. Related Work

2.1. Stochastic Compositional Optimization

The stochastic compositional optimization problem exhibits a nested structure so that its stochastic gradient is a biased estimator of the full gradient, which could lead to a slow convergence rate. To address this issue with biased gradient estimators, (Wang et al., 2017a) developed the first stochastic compositional gradient descent (SCGD) algorithm for the two-level compositional optimization problem. However, its convergence rate is still worse than that of the standard stochastic gradient descent (SGD) algorithm for non-compositional optimization problems. Subsequently, numerous algorithms (Wang et al., 2017b; Ghadimi et al., 2020; Zhang & Xiao, 2019a;c; Yuan et al., 2019; Chen et al., 2020) have been developed to improve the convergence rate of SCGD. For instance, (Ghadimi et al., 2020) leverages the moving-average technique to estimate both the inner-level function and gradient, enabling the convergence rate of SCGD to match that of SGD for nonconvex problems. (Zhang & Xiao, 2019a;c; Yuan et al., 2019) employ the variance-reduction technique (Defazio et al., 2014; Fang et al., 2018; Nguyen et al., 2017) to estimate both inner-level function and Jacobian matrix, which further improve the convergence rate of (Ghadimi et al., 2020).

The aforementioned methods restrict their focus to the stochastic two-level compositional optimization problem. They cannot be directly applied to the multi-level case because the bias increases with the growth of the number of levels. (Yang et al., 2019) shows that the standard SCGD algorithm (Wang et al., 2017a) can only achieve a level-dependent convergence rate. In particular, the convergence rate becomes exponentially slow with increasing levels. To address this issue, (Balasubramanian et al., 2022) leverages the linearized averaging technique to achieve a level-independent convergence rate. (Zhang & Xiao, 2021; Jiang et al., 2022) employ the variance reduction technique (Fang et al., 2018; Nguyen et al., 2017; Cutkosky & Orabona, 2019) to both inner-level function and Jacobian matrices to further improve the convergence rate.

2.2. Federated Stochastic Compositional Optimization

With the recent advancements in federated learning, numerous distributed stochastic compositional optimization algorithms (Gao & Huang, 2021; Huang & Li, 2021; Tarzanagh et al., 2022; Wang et al., 2023; Gao et al., 2022; Guo et al., 2023; Wu et al., 2023; Huang, 2022; Gao, 2024; Zhang et al., 2023; Zhao & Liu, 2024; Wu et al., 2024) have been developed. In particular, for the stochastic *two-level* compositional optimization problem, the first distributed SCGD algorithm was proposed in (Gao & Huang, 2021) under the decentralized communication setting. Subsequently, (Zhao & Liu, 2024) improved the sample complexity of (Gao & Huang, 2021) with the variance-reduction technique (Cutkosky & Orabona, 2019). Under the standard federated learning setting, (Huang & Li, 2021) directly applied federated stochastic gradient descent method to solve the following two-level compositional optimization problem under the homogeneous setting:

$$\min_{x\in\mathbb{R}^{d_0}} \Phi(x) = \frac{1}{N}\sum_{n=1}^{N} F_n^{(2)}(F_n^{(1)}(x)) \ . \tag{2}$$

It is worth noting that the loss function in Eq. (2) is equivalent to $\frac{1}{N}\sum_{n=1}^{N} F_n^{(2)}\left(\frac{1}{N}\sum_{n=1}^{N} F_n^{(1)}(x)\right)$ when the data distribution is homogeneous. This equivalence arises from $F_n^{(1)}(x) = \frac{1}{N}\sum_{n'=1}^{N} F_{n'}^{(1)}(x)$ under this setting. However, due to the biased gradient estimator issue on each worker, the sample and communication complexities of (Huang & Li, 2021) are much worse than the standard federated SGD algorithm. Subsequently, (Gao et al., 2022) proposed a local SCGD with momentum (LocalSCGDM) algorithm, whose sample and communication complexities for nonconvex problems can match those of the standard federated SGD algorithm. (Tarzanagh et al., 2022) proposed to communicate the variable, gradient, and Jacobian matrix to handle the heterogeneous distribution. However, its communication complexity is the same as the iteration complexity, which is not practical for federated learning and fails to match federated SGD and LocalSCGDM. Recently, (Huang, 2022) further improved the convergence rate with the variance-reduction technique (Cutkosky & Orabona, 2019), where the variable and gradient estimator are communicated. More recently, (Wu et al., 2024) proposed a federated conditional stochastic optimization algorithm that can be used to optimize Eq. (2) under the heterogeneous setting. However, it cannot be applied to our Eq. (1) because (Wu et al., 2024) requires that the inner-level function should not be distributed on different workers. For the *multi-level* compositional optimization problem, the first distributed algorithm was proposed in (Gao, 2024) under the decentralized setting, where the variable and gradient estimator are communicated across workers. However, *these existing algorithms designed for both two-level and multi-level compositional optimization problems fail to achieve a linear speedup convergence rate under both homogeneous and heterogeneous settings*. Only two recent works (Gao, 2023; Zhang et al., 2023) can achieve the linear speedup convergence rate for the two-level composition minimax problem. However, they focus solely on the homogeneous setting. Thus, it remains unclear whether the linear speedup can be achieved under the heterogeneous setting for *multi-level* compositional optimization problems.

3. Federated Doubly Recursive Stochastic Compositional Gradient Descent

In this section, we propose a novel algorithm for the federated multi-level compositional problem in Eq. (1), which can achieve a level-independent convergence rate and the linear speedup regarding the number of workers simultaneously under the heterogeneous setting.

3.1. Problem Setup

At first, we introduce the following assumptions, which have been commonly used in existing multi-level compositional optimization algorithms (Yang et al., 2019; Balasubramanian et al., 2022; Jiang et al., 2022; Gao, 2024).

Assumption 3.1. For any $k \in \{1, 2, \cdots, K\}$, $n \in \{1, \cdots, N\}$, $h_1 \in \mathbb{R}^{d_{k-1}}$, and $h_2 \in \mathbb{R}^{d_{k-1}}$, the stochastic Jacobian matrix (or gradient) $\nabla F_n^{(k)}(\cdot; \xi^{(k)})$ satisfies $\mathbb{E}[\|\nabla F_n^{(k)}(h_1; \xi^{(k)})\|] \leq C$, $\mathbb{E}[\|\nabla F_n^{(k)}(h_1; \xi) - \nabla F_n^{(k)}(h_2; \xi^{(k)})\|] \leq L\|h_1 - h_2\|$, where $C > 0$ and $L > 0$ are constant values.

Throughout this paper, it is assumed that the full Jacobian matrix shares the same C and L as its stochastic counterpart.

Assumption 3.2. For any $k \in \{1, 2, \cdots, K\}$, $n \in \{1, \cdots, N\}$, and $h \in \mathbb{R}^{d_{k-1}}$, the stochastic function $F_n^{(k)}(\cdot; \xi^{(k)})$ satisfies $\mathbb{E}[\|F_n^{(k)}(h; \xi^{(k)}) - F_n^{(k)}(h)\|^2] \leq \sigma^2$, and stochastic Jacobian matrix (or gradient) $\nabla F_n^{(k)}(\cdot;\xi^{(k)})$ satisfies $\mathbb{E}[\|\nabla F^{(k)}(h;\xi^{(k)}) - \nabla F^{(k)}(h)\|^2] \leq \sigma^2$, where $\sigma > 0$ is a constant value.

Assumption 3.3. For any $k \in \{1, 2, \cdots, K\}$ and any $n \in \{1, \cdots, N\}$, there exists a constant $\delta > 0$ such that $\mathbb{E}[\|F_n^{(k)}(\cdot;\xi^{(k)}) - \frac{1}{N}\sum_{n'=1}^{N} F_n^{(k)}(\cdot;\xi^{(k)})\|^2] \leq \delta^2$ and $\mathbb{E}[\|\nabla F_n^{(k)}(\cdot;\xi^{(k)}) - \frac{1}{N}\sum_{n'=1}^{N} \nabla F_n^{(k)}(\cdot;\xi^{(k)})\|^2] \leq \delta^2$.

Throughout this paper, it is assumed that the full inner-level function and Jacobian matrix share the same δ as their stochastic counterparts.

Given these assumptions, we have the following lemma.

Lemma 3.4. *Suppose Assumptions 3.1-3.3 hold, then $\Phi(x)$ is L_Φ-smooth, where $L_\Phi = L\sum_{k=1}^{K} C^{K+k-2}$.*

Its proof is deferred to Appendix C.

In this paper, we denote $F^{(k)}(\cdot) = \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(\cdot)$. Additionally, we use $\bar{w}_t = \frac{1}{N}\sum_{n=1}^{N} w_{n,t}$ to denote the mean value across workers for any variable w in the t-th iteration. Moreover, x^* denotes the optimal value.

3.2. Heterogeneity Prevents Convergence

Based on the aforementioned assumptions, we first demonstrate that the federated compositional gradient descent algorithm designed for the homogeneous setting cannot be directly applied to the heterogeneous setting. In particular, we focus on the two-level case, i.e., $K = 2$ in Eq. (1). Then, we employ the state-of-the-art algorithm, LocalSCGDM (Gao et al., 2022), to solve it. The detailed steps can be found in Algorithm 2 in Appendix B.

Under the heterogeneous setting, LocalSCGDM has the following convergence rate.

Theorem 3.5. *Suppose Assumptions 3.1-3.3 hold, by setting $\alpha > 0$, $\beta > 0$, $\eta \leq \min\{\frac{1}{2\gamma L_\Phi}, \frac{1}{\alpha}, \frac{1}{\beta}\}$ and $\gamma \leq \frac{1}{2}\sqrt{1\Big/\left(\frac{2L^2C^4}{\alpha\beta} + \frac{2L_\Phi^2}{\alpha^2} + \frac{24C^4L^2}{\beta^2}\right)}$, the convergence rate of LocalSCGDM is*

$$\begin{aligned} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] &\leq O\left(\frac{1}{\gamma\eta T}\right) + O\left(\frac{\sigma^2}{\alpha\eta T}\right) \\ &+ O\left(\frac{\sigma^2}{\beta\eta T}\right) + O\left(\frac{\gamma^2\alpha^2p^2\eta^3}{\beta}\right) + O\left(\frac{\alpha\eta\sigma^2}{N}\right) \\ &+ O\left(\frac{\gamma^2\alpha^2p^2\eta^2}{\beta^2}\right) + O\left(p^2\eta^2\gamma^2\right) + O\left(\beta^2\eta^2\sigma^2\right) \\ &+ O\left(\beta\eta\sigma^2\right) + O\left(\alpha^2\gamma^2p^4\eta^4\right) + O\left(p^2\beta^2\eta^2\sigma^2\right) \\ &+ O\left(p^2\beta^2\eta^2\delta^2\right) + O\left(\alpha^2\beta^2\gamma^2p^6\eta^6\right) + O\left(\delta^2\right) . \end{aligned} \tag{3}$$

Theorem 3.5 indicates that LocalSCGDM can only converge to the optimal solution's neighborhood under the heterogeneous setting. Specifically, by setting $\alpha = O(1)$, $\beta = O(1)$, $\eta = O(1/\sqrt{T})$, and $p = O(T^{1/4})$, we can obtain

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \leq O\left(\frac{1}{\sqrt{T}}\right) + O\left(\delta^2\right) . \tag{4}$$

It can be observed that there exists an additive term regarding the heterogeneity on the right-hand side, which does not approach zero. Consequently, the algorithm designed for the homogeneous setting cannot be directly applied to the heterogeneous setting.

3.3. Our Algorithm

As discussed in the previous subsection, heterogeneity can prevent convergence. To address this issue, in this paper, we develop a novel federated compositional algorithm for the generic multi-level compositional optimization problem.

Key Idea. By investigating the proof of Theorem 3.5, we found that the additive term $O(\delta)$ appears when bounding $\mathbb{E}\left[\left\|\nabla\Phi(\bar{x}_t) - \frac{1}{N}\sum_{n=1}^{N}\nabla_x f_n^{(2)}(f_n^{(1)}(x_{n,t}))\right\|^2\right]$, which is marked with the blue color in Eq. (24), i.e., the Jacobian heterogeneity: $\mathbb{E}\left[\left\|\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(\bar{x}_t) - \nabla F_n^{(1)}(\bar{x}_t)\right\|^2\right]$. On the other hand, in Eq. (24), the heterogeneity regarding the *inner-level function* $\mathbb{E}\left[\left\|\frac{1}{N}\sum_{n'=1}^{N} F_{n'}^{(1)}(\bar{x}_t) - F_n^{(1)}(\bar{x}_t)\right\|^2\right]$ is addressed by *introducing the inner-level function estimator $h_{n,t}^{(1)}$ and communicating it across workers*, which can be found in Algorithm 2 in Appendix B. Inspired by that, a straightforward approach to address the Jacobian heterogeneity issue is to introduce an estimator $J^{(k)} \in \mathbb{R}^{d_k \times d_{k-1}}$ for the Jacobian matrix $\nabla F^{(k)}(\cdot)$ where $k \in \{1, \cdots, K-1\}$, and communicate it across workers. Then, one can decompose the Jacobian heterogeneity as follows:

$$\begin{aligned} &\mathbb{E}\left[\left\|\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(x) - \nabla F_n^{(1)}(x)\right\|^2\right] \\ &\leq 3\mathbb{E}\left[\left\|\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(x) - \frac{1}{N}\sum_{n'=1}^{N} J_{n'}^{(k)}\right\|^2\right] \\ &+ 3\mathbb{E}\left[\left\|\bar{J}^{(k)} - J_n^{(k)}\right\|^2\right] + 3\mathbb{E}\left[\left\|J_n^{(k)} - \nabla F_n^{(1)}(x)\right\|^2\right] . \end{aligned} \tag{5}$$

Because the Jacobian matrix estimator $J_n^{(k)}$ is communicated across workers, the consensus error of the Jacobian matrix $\mathbb{E}[\|\bar{J}^{(k)} - J_n^{(k)}\|^2]$ can be much smaller than keeping Jacobian locally as Algorithm 2. Moreover, by introducing the estimator $J_n^{(k)}$, which is associated with hyperameters like the inner-level function estimator $h_{n,t+1}^{(1)}$, one can further control the consensus error by adjusting those hyperparameters, thus avoiding the additive constant term $O(\delta^2)$ in the upper bound of the convergence rate.

Algorithm 1 Federated Doubly-Recursive Stochastic Compositional Gradient Descent Algorithm (Fed-DR-SCGD)

Input: $x_{n,0} = x_0$, $p > 1$, $\alpha > 0$, $\gamma > 0$, $\eta > 0$. All workers perform the following steps.

Initialization at $t = 0$:
Compute $h_{n,t}^{(k)}$ and $v_{n,t}^{(k)}$ with a mini-batch of samples with the batch size being S:

$$h_{n,t}^{(k)} = \begin{cases} x_{n,t} \ , & k = 0 \\ F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t}^{(k)}) \ , & k \in \{1, \cdots, K-1\} \end{cases} ,$$

$$v_{n,t}^{(k)} = \begin{cases} \nabla F_n^{(K)}(h_{n,t}^{(K-1)}; \xi_{n,t}^{(K)}) \ , & k = K \\ \nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} \ , & k \in \{K-1, \cdots, 1\} \end{cases} ,$$

for $t = 0, \cdots, T-1$ each worker n **do**
    $x_{n,t+1} = x_{n,t} - \gamma\eta v_{n,t}^{(1)}$,
    Recursively update the inner-level function estimator:
    $h_{n,t+1}^{(0)} = x_{n,t+1}$,
    for $k = 1, \cdots, K-1$ **do**
        $h_{n,t+1}^{(k)} = (1 - \alpha\eta^2)(h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})) + F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})$,
    end for
    Recursively update the Jacobian-vector product estimator:
    $v_{n,t+1}^{(K)} = \Pi_{C_{\Psi^{(K)}}}((1 - \alpha\eta^2)(v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)}; \xi_{n,t+1}^{(K)})) + \nabla F_n^{(K)}(h_{n,t+1}^{(K-1)}; \xi_{n,t+1}^{(K)}))$,
    for $k = K-1, \cdots, 1$ **do**
        $v_{n,t+1}^{(k)} = \Pi_{C_{\Psi^{(k)}}}((1 - \alpha\eta^2)(v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t}^{(k+1)}) + \nabla F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t+1}^{(k+1)})$,
    end for
    Perform communication:
    if $\mathrm{mod}(t+1, p) == 0$ **then**
        $x_{n,t+1} = \bar{x}_{t+1} \triangleq \frac{1}{N}\sum_{n'=1}^{N} x_{n',t+1}$,
        $h_{n,t+1}^{(k)} = \bar{h}_{t+1}^{(k)} \triangleq \frac{1}{N}\sum_{n'=1}^{N} h_{n',t+1}^{(k)}$, $k \in \{1, \cdots, K-1\}$,
        $v_{n,t+1}^{(k)} = \bar{v}_{t+1} \triangleq \frac{1}{N}\sum_{n'=1}^{N} v_{n',t+1}^{(k)}$, $k \in \{1, \cdots, K\}$,
    end if
end for

Algorithm. The aforementioned straightforward approach is not practical since it can *incur high communication costs when the dimensionality of Jacobian matrix is large*. To address this issue, we develop a novel federated compositional optimization algorithm in Algorithm 1, which **introduces the Jacobian-vector product estimator to address the heterogeneity and communication challenges simultaneously**. Specifically, for any $k \in \{K, \cdots, 1\}$, we recursively define the Jacobian-vector product $\Psi^{(k)}(x) \in \mathbb{R}^{d_{k-1}}$ as follows:

$$\Psi^{(k)}(x) = \begin{cases} \nabla F^{(K)}(h^{(K-1)}), & k = K \\ \nabla F^{(k)}(h^{(k-1)})^T \Psi^{(k+1)}(x), & \text{otherwise.} \end{cases} \tag{6}$$

Note that $\Psi^{(k)}(x)$ is computed *from the last level to the first level*. Additionally, $h^{(k)} \in \mathbb{R}^{d_k}$ is the estimator of the k-th level function $F^{(k)}(\cdot)$. If $h^{(k)}$ is exactly the k-th level function $F^{(k)}$, $\Psi^{(1)}(x)$ is the gradient of the loss function $\Phi(x)$, i.e., $\Psi^{(1)}(x) = \nabla\Phi(x)$. From the definition of $\Psi^{(k)}(x)$, it can be observed that the dimensionality of $\Psi^{(k)}(x)$ is much smaller than the Jacobian matrix $J^{(k)}$. Thus, it can reduce the communication cost significantly.

When using random samples to compute $\Psi^{(k)}(x)$ for $k \in \{K-1, \cdots, 1\}$, we introduce the following variance-reduced estimator:

$$\begin{aligned} v_{n,t}^{(k)} &= \Pi_{C_{\Psi^{(k)}}}(\nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} \\ &\quad + (1 - \alpha\eta^2)(v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)}; \xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)})) \ , \end{aligned} \tag{7}$$

where n is the worker index, t is the iteration index, k is the level index, $\alpha > 0$, $\eta > 0$, and $\alpha\eta^2 < 1$. Here, because of $\mathbb{E}[\|\Psi^{(k)}(x)\|] \leq C_{\Psi^{(k)}}$, whose proof can be found in Lemma C.1, we use a projection operator $\Pi_{C_{\Psi^{(k)}}}(\cdot)$ to guarantee its estimator $v_{n,t}^{(k)}$ to satisfy this condition. A linear projection operator in Lemma C.21 can be used in this step. In fact, $v_{n,t}^{(1)}$ is the stochastic variance-reduced compositional gradient, which is used to update x as follows:

$$x_{n,t+1} = x_{n,t} - \gamma\eta v_{n,t}^{(1)} \ , \tag{8}$$

where $\gamma > 0$ and $\eta > 0$ are two hyperparameters.

As for the inner-level function estimator $h_{n,t}^{(k)} \in \mathbb{R}^{d_k}$ where $k \in \{1, \cdots, K-1\}$, similar to existing multi-level compositional optimization methods (Zhang & Xiao, 2021; Jiang et al., 2022; Gao, 2024), we update it as follows:

$$h_{n,t}^{(k)} = (1-\alpha\eta^2)(h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})) + F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)}) \ . \tag{9}$$

Different from the update of $v_{n,t}^{(k)}$, $h_{n,t}^{(k)}$ is updated from the first level to the last level. Then, at every p iterations, where $p > 1$, each worker communicates its local variable $x_{n,t}$, inner-level function estimators $\{h_{n,t}^{(k)}\}_{k=1}^{K-1}$, and Jacobian-vector product estimators $\{v_{n,t}^{(k)}\}_{k=1}^{K}$ with the central server to get the global counterparts.

One can observe that both the inner-level function estimator $h_{n,t}^{(k)}$ and the Jacobian-vector estimator $v_{n,t}^{(k)}$ are *recursively computed across levels*. Therefore, our algorithm is referred to as the **Federated Doubly Recursive Stochastic Compositional Gradient Descent** (Fed-DR-SCGD) algorithm.

Discussions. Different from existing stochastic multi-level compositional gradient descent methods (Zhang & Xiao, 2021; Jiang et al., 2022), we apply the variance reduction technique to the Jacobian-vector product $\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)}$, rather than the Jacobian matrix $\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})$. This novel strategy can effectively address the *practical* communication issue. But it brings new challenges for *theoretical* analysis. In particular, in existing methods (Zhang & Xiao, 2021; Jiang et al., 2022), the estimation error of the Jacobian matrix is decoupled in different levels, e.g., Lemma 6 of (Jiang et al., 2022). On the contrary, the estimation error of the Jacobian-vector product depends on the estimator in upper levels, e.g., Lemma C.12 in Appendix C, which makes the convergence analysis much more challenging than existing methods (Zhang & Xiao, 2021; Jiang et al., 2022). On the other hand, compared with existing federated learning algorithms (Huang & Li, 2021; Gao et al., 2022; Tsaknakis et al., 2020) for two-level compositional problems, the convergence analysis is much more challenging. First, the multi-level structure leads to more unique challenges. For example, unlike the consensus error about the compositional gradient of two-level problems in Eq. (16) , which can be bounded by a constant, $\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2]$ of Algorithm 1 depends on the estimation error of all Jacobian-vector product and inner-level functions as shown in Lemma C.7, which is much more difficult to bound. Second, the additional Jacobian-vector product estimator introduces more challenges for convergence analysis. For example, beyond the challenge for bounding its own estimation error in Lemma C.11, it also brings significant challenges for other terms, such as $\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2]$ in Lemma C.16 and $\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2]$ in Lemma C.17. In our paper, we successfully addressed these unique and significant challenges, establishing the convergence rate of our algorithm.

4. Convergence Analysis

4.1. Convergence Rate

Theorem 4.1. *Suppose Assumptions 3.1-3.3 hold, by setting* $\alpha \le \frac{C}{N}$, $\gamma \le \min\Big\{\frac{1}{4\sqrt{J_3(C,L,\frac{1}{\alpha N})K}}, \frac{1}{2(J_4(C,L,\frac{1}{\alpha N})K)^{1/4}}, \frac{1}{(16J_5(C,L,\frac{1}{\alpha N})K)^{1/6}}\Big\}$, $\eta \le \min\Big\{\frac{\sqrt{C\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\sqrt{D_1(C,L)}}, \frac{\sqrt{C\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma\sqrt{D_2(C,L,\frac{1}{\alpha N})}}, \frac{\sqrt{C^3\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma\sqrt{D_3(C,L)}}, \frac{\sqrt{C^3\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma^2\sqrt{D_4(C,L,\frac{1}{\alpha N})}}, \frac{\sqrt{C\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\sqrt{E_1(C)}}, \frac{\sqrt{C\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma\sqrt{E_2(C,L,\frac{1}{\alpha N})}}, , \frac{1}{20\gamma pC^{K-1}L}, \frac{\sqrt{C^3\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma\sqrt{E_3(C,L)}}, \frac{\sqrt{C^3\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma^2\sqrt{E_4(C,L,\frac{1}{\alpha N})}}, \frac{\sqrt{C\omega_{a_k}(C)}}{3\gamma\sqrt{B_1(C,L,\frac{1}{\alpha N})}}, \frac{\sqrt{C^3\omega_{a_k}(C)}}{3\gamma\sqrt{B_2(C,L)}}, \frac{\sqrt{C^3\omega_{a_k}(C)}}{3\gamma^2\sqrt{B_3(C,L,\frac{1}{\alpha N})}}, \frac{1}{6\gamma p(3C^2)^{K-1}} \frac{1}{2\gamma L_\Phi}, \frac{\sqrt{\omega_{a_k}(C)}}{\sqrt{6(36\sum_{j=1}^{k-1}C^{2j}(3C^2)^{k-1-j}+\sum_{j=1}^{k-1}C^{2j-1}(3C^2)^{k-1-j})}}, \frac{\sqrt{C\omega_{a_k}(C)}}{\gamma\sqrt{3\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{k-1}}}, \frac{C_Q(C,L,\frac{1}{\alpha N})}{96\gamma p(3C)^{K-1}L}, \frac{1}{20pC}, \frac{1}{\sqrt{\alpha}}, 1\Big\}$, *where* $\{B_i\}_{i=1}^3$ *in Eq. (128),* $\{D_i\}_{i=1}^4$ *in Eq. (140),* $\{E_i\}_{i=1}^4$ *in Eq. (133),* $\{J_i\}_{i=1}^5$ *in Eqs. (145, 147),* $\{\omega_{a_k}\}_{k=1}^K$ *in Lemma C.18,* $\{\nu_{b_k}\}_{k=1}^{K-1}$ *in Lemma C.19,* C_Q *in Eq. (153), and* $\hat{P}_{v_1^c}$ *in Lemma C.20 are constant values with respect to* C, L, *or* $\frac{1}{\alpha N}$, *then we set* $\alpha = O(\frac{1}{N})$, *Fed-DR-SCGD has the following convergence rate:*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \le O\left(\frac{1}{\alpha\eta^2 SNT}\right) + O\left(\frac{1}{\eta^2 ST}\right) + O\left(\gamma^2\alpha^2p^2\eta^4\right) + O\left(\gamma^2\alpha^2p^4\eta^6\right) + O\left(\gamma^4\alpha^2p^4\eta^6\right) + O\left(\alpha^2\eta^2\right) + O\left(\alpha^2p^2\eta^4\right) + O\left(\frac{\alpha\eta^2}{N}\right) \ . \tag{10}$$

Remark 4.2. In Theorem 4.1, we set $\alpha = O(\frac{1}{N})$. Consequently, the upper bounds of γ and η can be considered independent of $\frac{1}{\alpha N}$, which actually are constants with respect to the Lipschitz constants C and L.

Remark 4.3. For a sufficiently large T, by setting the hyperparameter $\alpha = O(\frac{1}{N})$, the learning rate $\eta = O(\frac{N^{2/3}}{T^{1/3}})$, the communication period $p = O(\frac{T^{1/3}}{N^{2/3}})$, the batch size in the initial step $S = O(\frac{T^{1/3}}{N^{2/3}})$, Fed-DR-SCGD has the following convergence upper bound:

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \le O\left(\frac{1}{N^{2/3}T^{2/3}}\right) \ . \tag{11}$$

This convergence rate indicates the linear speedup with respect to the number of workers N, while the *decentralized* stochastic multi-level gradient descent method in (Gao, 2024) cannot achieve linear speedup. When $N = 1$, our convergence rate can match the state-of-the-art algorithms

(Zhang & Xiao, 2021; Jiang et al., 2022) under the single-machine setting.

Remark 4.4. To achieve the ϵ-accuracy solution such that $\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \leq \epsilon^2$, we can set the hyperparameter $\alpha = O(\frac{1}{N})$, the learning rate $\eta = O(N\epsilon)$, the communication period $p = O(\frac{1}{N\epsilon})$, the batch size in the initial step $S = O(\frac{1}{N\epsilon})$, and the number of iterations $T = O(\frac{1}{N\epsilon^3})$. Then, the sample complexity is $O(\frac{1}{N\epsilon^3})$ and the communication complexity is $O(\frac{1}{\epsilon^2})$.

4.2. Proof Sketch

To prove Theorem 4.1, we developed the following potential function:

$$
\begin{aligned}
P_t &= \mathbb{E}[\Phi(\bar{x}_t)] + a_{2,K}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&+ \sum_{k=1}^{K-1} a_{1,k}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&+ \sum_{k=1}^{K-1} b_{1,k}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&+ \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1} a_{2,k}\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&+ \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1} b_{2,k}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] ,
\end{aligned} \tag{12}
$$

where the coefficients $\{a_{1,k}\}_{k=1}^{K}$, $\{a_{2,k}\}_{k=1}^{K}$, $\{b_{1,k}\}_{k=1}^{K-1}$, and $\{b_{2,k}\}_{k=1}^{K-1}$ are positive.

As discussed previously, the doubly recursive property introduces significant challenges for convergence analysis. In particular, there are two main challenges when bounding the potential function. First, *the upper bounds of those terms in the potential function are interdependent to each other*. Second, *the coefficients are interdependent across levels*. To address the first challenge, we established the upper bound for each term of the potential function and then expanded it such that the final upper bound depends on $\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2]$. As shown in Lemma C.16, the upper bound of $\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2]$ primarily consists of the terms in the potential function. Then, they can be combined together and canceled out by proper coefficients. For the second challenge, we explicitly established the dependence between different levels for $\{a_{2,k}\}_{k=1}^{K}$ and $\{b_{2,k}\}_{k=1}^{K-1}$ in Lemmas C.18, C.19, respectively. Based on those coefficients, we can cancel out most terms in the potential function to complete the proof.

5. Experiment

In this section, we conduct experiments to verify the performance of our proposed algorithm.

5.1. Experiment Setup

In our experiments, we focus on the Risk-Averse Portfolio Optimization problem, which has been commonly used to verify the performance of multi-level compositional optimization algorithms (Yang et al., 2019; Zhang & Xiao, 2021; Jiang et al., 2022; Balasubramanian et al., 2022). Specifically, it is assumed that there are d assets to invest in T time steps. $r_t \in \mathbb{R}^d$ is the payoff of those d assets in each time step $t \in \{1, \cdots, T\}$. Then, to maximize the total return with the minimal risk, one can optimize the following multi-level compositional problem:

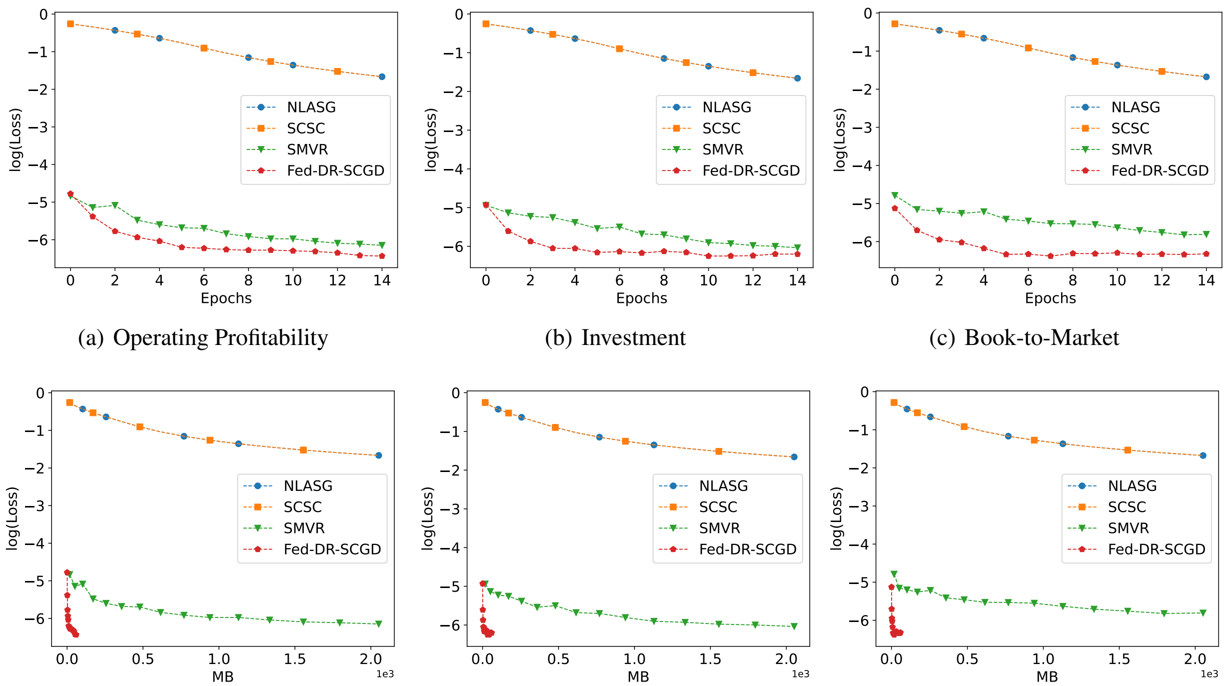
$$
\min_{x\in\mathbb{R}^d} \lambda\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_t^T x - \frac{1}{T}\sum_{t=1}^{T} r_t^T x\right)^2} - \frac{1}{T}\sum_{t=1}^{T} r_t^T x \ , \tag{13}
$$

where $\lambda > 0$ is a hyperparameter, which is set to 1 in our experiments. Such an optimization problem can be represented as a three-level compositional optimization problem. Specifically, each level function can be defined as follows:

$$
\begin{aligned}
f^{(1)}(x) &= \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} r_t^T x \\ x \end{bmatrix} \triangleq \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^{1+d}, \\
f^{(2)}(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) &= \begin{bmatrix} y_1 \\ \frac{1}{T}\sum_{t=1}^{T}\left(r_t^T y_2 - y_1\right)^2 \end{bmatrix} \triangleq \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in \mathbb{R}^2 , \\
f^{(3)}(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}) &= -z_1 + \lambda\sqrt{z_2} \in \mathbb{R} .
\end{aligned} \tag{14}
$$

Then, it is easy to know that $\nabla f^{(3)}(\cdot) \in \mathbb{R}^2$, $\nabla f^{(2)}(\cdot) \in \mathbb{R}^{2\times(d+1)}$, and $\nabla f^{(1)}(\cdot) \in \mathbb{R}^{(d+1)\times d}$. Obviously, if directly communicating these Jacobian matrices, the communication cost will be high when the number of assets d is large. Following (Lin et al., 2020), we use three datasets in our experiments: Book-to-Market, Operating Profitability, and Investment. Each dataset has 13,781 time steps and 100 assets, i.e., $T = 13,781$, $d = 100$.

In our experiments, we compare our algorithm with three state-of-the-art stochastic multi-level compositional gradient descent algorithms: multi-level nested linearized averaging stochastic Gradient (NLASG) algorithm (Balasubramanian et al., 2022), stochastically corrected stochastic compositional gradient (SCSC) algorithm (Chen et al., 2020), and stochastic multi-level variance reduction (SMVR) algorithm (Jiang et al., 2022). In our experiments, we set the solution accuracy ϵ to $1e-6$. Then, according to the theoretical results in (Balasubramanian et al., 2022; Chen et al., 2020; Jiang et al., 2022), the learning rate of NLASG and SCSC is set to ϵ, while that of SMVR and our algorithm is set to $\epsilon^{1/2}$, i.e., $\gamma = 1$ and $\eta = \epsilon^{1/2}$ for our algorithm. In addition, we set the coefficient of the momentum in all methods to 0.95, i.e., $\alpha\eta^2 = 0.95$ for our algorithm. In addition, we use eight workers in our experiments, where the batch size on each worker is 1 for all methods. Moreover, we parallelize baseline methods by communicating their inner-level



(a) Operating Profitability (b) Investment (c) Book-to-Market

(d) Operating Profitability (e) Investment (f) Book-to-Market

Figure 1. First row: The loss function versus epochs for different methods. Second row: The loss function versus the communication cost (megabytes) for different methods. The communication period is $p = 4$.

function estimators, Jacobian matrices, and variable in the same communication frequency as our method.

5.2. Experimental Results

In Figure 1, we show the logarithm of the loss function value versus the number of epochs in the first row and the communication cost in the second row, where the communication period is set to 4. From Figure 1, it can be observed that our Fed-DR-SCGD algorithm is much more communication-efficient than baseline algorithms, which confirms the efficacy of the proposed Jacobian-vector product estimator. Moreover, one can observe that our algorithm and SMVR converge much faster than the other two algorithms due to the variance-reduced gradient, which confirms the correctness of our theoretical convergence rate.

To verify the linear speedup effect, we compare the convergence rate when using 8 workers and 16 workers. Each A5000 GPU card accommodates 4 workers. Moreover, we use Operating Profitability dataset, and the communication period is set to 4. For these two configurations, we use the same hyperparameters. In Figure 2, we plot the loss function value versus the used time (seconds). It can be observed that our algorithm takes less time when using more workers, confirming the speedup effect.

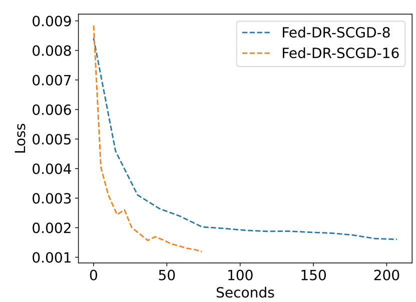


Figure 2. The loss function values versus the used time (seconds) when using 8 workers and 16 workers. The dataset is Operating Profitability and communication period is 4.

6. Conclusion

In this paper, we developed a novel federated learning algorithm for the stochastic multi-level compositional optimization problem. In particular, we developed a novel Jacobian-vector product estimator to address the heterogeneity and communication issues. Then, we proposed novel strategies to address the theoretical challenges caused by the recursively computed Jacobian-vector product estimator, establishing the convergence rate of our algorithm. Our algorithm is the first one achieving the linear speedup convergence rate for multi-level compositional optimization problems under the heterogeneous setting. The extensive experiments confirm the effectiveness of our new algorithm.

Acknowledgements

We thank anonymous reviewers for constructive comments.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Balasubramanian, K., Ghadimi, S., and Nguyen, A. Stochastic multilevel composition optimization algorithms with level-independent convergence rates. *SIAM Journal on Optimization*, 32(2):519–544, 2022.

Chen, T., Sun, Y., and Yin, W. Solving stochastic compositional optimization is nearly as easy as solving stochastic optimization. *arXiv preprint arXiv:2008.10847*, 2020.

Cutkosky, A. and Orabona, F. Momentum-based variance reduction in non-convex sgd. In *Advances in Neural Information Processing Systems*, pp. 15236–15245, 2019.

Defazio, A., Bach, F., and Lacoste-Julien, S. Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. *arXiv preprint arXiv:1407.0202*, 2014.

Fang, C., Li, C. J., Lin, Z., and Zhang, T. Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator. *Advances in Neural Information Processing Systems*, 31, 2018.

Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pp. 1126–1135. PMLR, 2017.

Gao, H. Achieving linear speedup in decentralized stochastic compositional minimax optimization. *arXiv preprint arXiv:2307.13430*, 2023.

Gao, H. Decentralized multi-level compositional optimization algorithms with level-independent convergence rate. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

Gao, H. and Huang, H. Fast training method for stochastic compositional optimization problems. *Advances in Neural Information Processing Systems*, 34, 2021.

Gao, H., Li, J., and Huang, H. On the convergence of local stochastic compositional gradient descent with momentum. In *International Conference on Machine Learning*, pp. 7017–7035. PMLR, 2022.

Ghadimi, S., Ruszczynski, A., and Wang, M. A single timescale stochastic approximation method for nested stochastic optimization. *SIAM Journal on Optimization*, 30(1):960–979, 2020.

Guo, Z., Jin, R., Luo, J., and Yang, T. Fedxl: provable federated learning for deep x-risk optimization. In *International Conference on Machine Learning*, pp. 11934–11966. PMLR, 2023.

Huang, F. Faster adaptive momentum-based federated methods for distributed composition optimization. *arXiv preprint arXiv:2211.01883*, 2022.

Huang, F. and Li, J. Compositional federated learning for distributionally robust and meta learning. *arXiv preprint arXiv:2106.11264*, 2021.

Jiang, W., Wang, B., Wang, Y., Zhang, L., and Yang, T. Optimal algorithms for stochastic multi-level compositional optimization. *arXiv preprint arXiv:2202.07530*, 2022.

Lin, T., Fan, C., Wang, M., and Jordan, M. I. Improved sample complexity for stochastic compositional variance reduced gradient. In *2020 American Control Conference (ACC)*, pp. 126–131. IEEE, 2020.

Nguyen, L. M., Liu, J., Scheinberg, K., and Takáč, M. Sarah: A novel method for machine learning problems using stochastic recursive gradient. In *International Conference on Machine Learning*, pp. 2613–2621. PMLR, 2017.

Tarzanagh, D. A., Li, M., Thrampoulidis, C., and Oymak, S. Fednest: Federated bilevel, minimax, and compositional optimization. In *International Conference on Machine Learning*, pp. 21146–21179. PMLR, 2022.

Tsaknakis, I., Hong, M., and Liu, S. Decentralized min-max optimization: Formulations, algorithms and applications in network poisoning attack. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5755–5759. IEEE, 2020.

Wang, B., Yuan, Z., Ying, Y., and Yang, T. Memory-based optimization methods for model-agnostic meta-learning and personalized federated learning. *Journal of Machine Learning Research*, 24:1–46, 2023.

Wang, M., Fang, E. X., and Liu, H. Stochastic compositional gradient descent: algorithms for minimizing compositions of expected-value functions. *Mathematical Programming*, 161(1-2):419–449, 2017a.

Wang, M., Liu, J., and Fang, E. X. Accelerating stochastic composition optimization. *The Journal of Machine Learning Research*, 18(1):3721–3743, 2017b.

Wu, X., Sun, J., Hu, Z., Li, J., Zhang, A., and Huang, H. Federated conditional stochastic optimization. *arXiv preprint arXiv:2310.02524*, 2023.

Wu, X., Sun, J., Hu, Z., Li, J., Zhang, A., and Huang, H. Federated conditional stochastic optimization. *Advances in Neural Information Processing Systems*, 36, 2024.

Yang, H., Fang, M., and Liu, J. Achieving linear speedup with partial worker participation in non-iid federated learning. *arXiv preprint arXiv:2101.11203*, 2021.

Yang, S., Wang, M., and Fang, E. X. Multilevel stochastic gradient methods for nested composition optimization. *SIAM Journal on Optimization*, 29(1):616–659, 2019.

Yu, H., Jin, R., and Yang, S. On the linear speedup analysis of communication efficient momentum sgd for distributed non-convex optimization. *arXiv preprint arXiv:1905.03817*, 2019.

Yu, H., Wang, L., Wang, B., Liu, M., Yang, T., and Ji, S. Graphfm: Improving large-scale gnn training via feature momentum. In *International Conference on Machine Learning*, pp. 25684–25701. PMLR, 2022.

Yuan, H., Lian, X., and Liu, J. Stochastic recursive variance reduction for efficient smooth non-convex compositional optimization. *arXiv preprint arXiv:1912.13515*, 2019.

Zhang, J. and Xiao, L. A composite randomized incremental gradient method. In *International Conference on Machine Learning*, pp. 7454–7462, 2019a.

Zhang, J. and Xiao, L. Multi-level composite stochastic optimization via nested variance reduction. *arXiv preprint arXiv:1908.11468*, 2019b.

Zhang, J. and Xiao, L. A stochastic composite gradient method with incremental variance reduction. In *Advances in Neural Information Processing Systems*, pp. 9078–9088, 2019c.

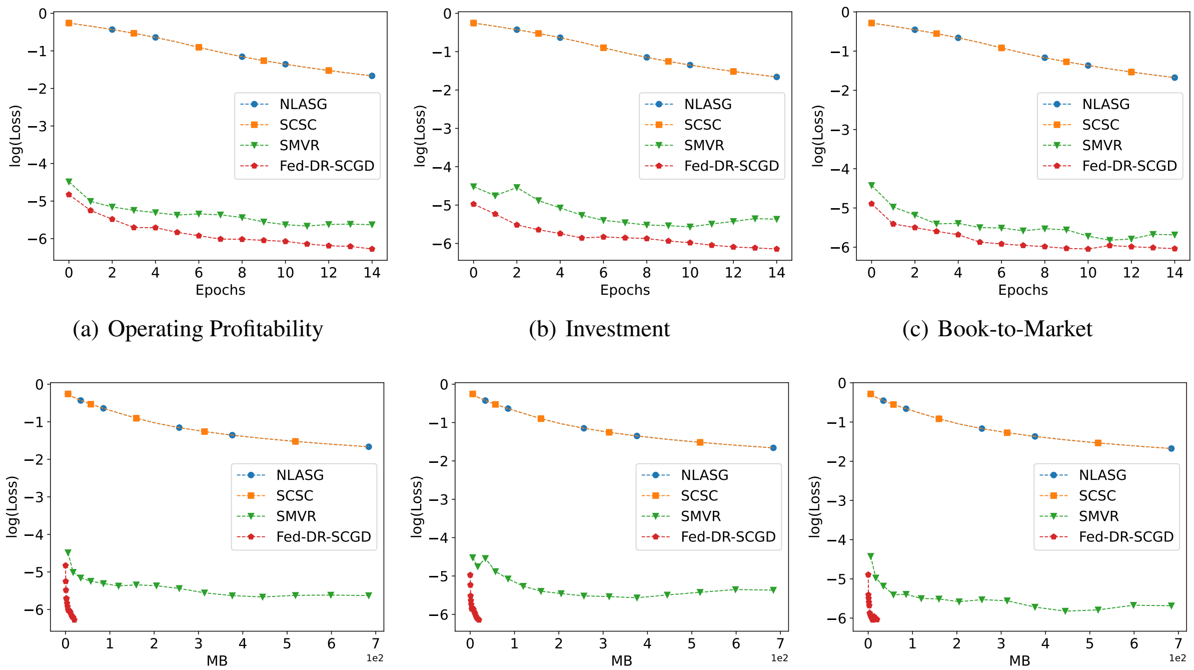
Zhang, J. and Xiao, L. Multilevel composite stochastic optimization via nested variance reduction. *SIAM Journal on Optimization*, 31(2):1131–1157, 2021.

Zhang, X., Zhang, Y., Yang, T., Souvenir, R., and Gao, H. Federated compositional deep auc maximization. *Advances in Neural Information Processing Systems*, 36, 2023.

Zhao, S. and Liu, Y. Distributed stochastic compositional optimization problems over directed networks. *Computational Optimization and Applications*, 87(1):249–288, 2024.

A. More Experimental Results

In Figure 3, we set the communication period p to 12. We can still find that our algorithm is more communication-efficient than all baseline algorithms.



(a) Operating Profitability (b) Investment (c) Book-to-Market

(d) Operating Profitability (e) Investment (f) Book-to-Market

Figure 3. First row: The loss function versus epochs for different methods. Second row: The loss function versus the communication cost (megabytes) for different methods. The communication period is $p = 12$.

B. Proof of Theorem 3.5

Algorithm 2 LocalSCGDM (Gao et al., 2022)

Input: $\eta > 0$, $\beta > 0$, $\gamma > 0$, $\alpha > 0$, $p > 1$, $x_{n,0} = x_0$.

Initialization at $t = 0$:
$h_{n,t}^{(1)} = F^{(1)}(x_{n,t}; \xi_{n,t}^{(1)})$, $v_{n,t} = \nabla F_n^{(1)}(x_{n,t}; \xi_{n,t}^{(1)})^T \nabla F_n^{(2)}(h_{n,t}^{(1)}; \xi_{n,t}^{(2)})$,
for $t = 0, \cdots, T-1$, each worker n **do**
  $x_{n,t+1} = x_{n,t} - \gamma\eta v_{n,t}$,
  $h_{n,t+1}^{(1)} = (1-\beta\eta)h_{n,t}^{(1)} + \beta\eta F_n^{(1)}(x_{n,t+1}; \xi_{n,t+1}^{(1)})$,
  $v_{n,t+1} = (1-\alpha\eta)v_{n,t} + \alpha\eta\nabla F_n^{(1)}(x_{n,t+1}; \xi_{n,t+1}^{(1)})^T \nabla F_n^{(2)}(h_{n,t+1}^{(1)}; \xi_{n,t+1}^{(2)})$,
  if mod$(t+1, p) == 0$ **then**
    $h_{n,t+1}^{(1)} = \bar{h}_{t+1}^{(1)} \triangleq \frac{1}{N}\sum_{n'=1}^{N} h_{n',t+1}^{(1)}$,
    $v_{n,t+1} = \bar{v}_{t+1} \triangleq \frac{1}{N}\sum_{n'=1}^{N} v_{n',t+1}$,
    $x_{n,t+1} = \bar{x}_{t+1} \triangleq \frac{1}{N}\sum_{n'=1}^{N} x_{n',t+1}$,
  end if
end for

Lemma B.1. *Suppose Assumptions 3.1-3.3 hold and $\eta \le \frac{1}{\alpha}$, we have*

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] \le 6\alpha^2\gamma^2p^4\eta^4C^4 \ . \tag{15}$$

Proof. At first, we have

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t+1} - \bar{v}_{t+1}\|^2] \\
&= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|(1-\alpha\eta)v_{n,t} + \alpha\eta\nabla F_n^{(1)}(x_{n,t+1}; \xi_{n,t+1}^{(1)})^T \nabla F_n^{(2)}(h_{n,t+1}^{(1)}; \xi_{n,t+1}^{(2)}) \\
&\quad - \frac{1}{N}\sum_{n'=1}^{N}((1-\alpha\eta)v_{n',t} + \alpha\eta\nabla F_{n'}^{(1)}(x_{n',t+1}; \xi_{n',t+1}^{(1)})^T \nabla F_{n'}^{(2)}(h_{n',t+1}^{(1)}; \xi_{n',t+1}^{(2)}))\|^2] \\
&\le (1+1/p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|(1-\alpha\eta)v_{n,t} - \frac{1}{N}\sum_{n'=1}^{N}(1-\alpha\eta)v_{n',t}\|^2] \\
&\quad + (1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\alpha\eta\nabla F_n^{(1)}(x_{n,t+1}; \xi_{n,t+1}^{(1)})^T \nabla F_n^{(2)}(h_{n,t+1}^{(1)}; \xi_{n,t+1}^{(2)}) \\
&\qquad - \alpha\eta\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(x_{n',t+1}; \xi_{n',t+1}^{(1)})^T \nabla F_{n'}^{(2)}(h_{n',t+1}^{(1)}; \xi_{n',t+1}^{(2)})\|^2] \\
&\le (1+1/p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t} - \frac{1}{N}\sum_{n'=1}^{N}v_{n',t}\|^2] + 2p\alpha^2\eta^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(1)}(x_{n,t+1}; \xi_{n,t+1}^{(1)})^T \nabla F_n^{(2)}(h_{n,t+1}^{(1)}; \xi_{n,t+1}^{(2)})\|^2] \\
&\le 2p\alpha^2\eta^2\sum_{t'=s_tp}^{t}(1+1/p)^{t-t'}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(1)}(x_{n,t'+1}; \xi_{n,t'+1}^{(1)})^T \nabla F_n^{(2)}(h_{n,t'+1}^{(1)}; \xi_{n,t'+1}^{(2)})\|^2] \\
&\le 2p\alpha^2\eta^2C^4\sum_{t'=s_tp}^{t}(1+1/p)^{t-t'} \\
&\le 6p^2\alpha^2\eta^2C^4 \ ,
\end{aligned} \tag{16}$$

where s_t denotes the communication round that happened before the $(t+1)$-th iteration, the third step holds due to $\alpha\eta \in (0,1)$, the second to last step holds due to Assumption 3.1, the last step holds due to $(1+1/p)^p < 3$.

Then, we can obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_{s_tp} - \gamma\eta\sum_{t'=s_tp}^{t-1}\bar{v}_{t'} - x_{n,s_tp} + \gamma\eta\sum_{t'=s_tp}^{t-1}v_{n,t'}\|^2] \\
&\leq p\gamma^2\eta^2\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t'} - \bar{v}_{t'}\|^2] \\
&\leq 6\alpha^2\gamma^2p^4\eta^4C^4 \ .
\end{aligned} \tag{17}$$

□

Lemma B.2. *Suppose Assumptions 3.1-3.3 hold and $\eta \leq \frac{1}{\beta}$, then we have*

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_{t+1}^{(1)} - h_{n,t+1}^{(1)}\|^2] \leq 60p^2\beta^2\eta^2\sigma^2 + 30p^2\beta^2\eta^2\delta^2 + 360\alpha^2\beta^2\gamma^2p^6\eta^6C^6 \ . \tag{18}$$

Proof. Similar to Eq. (16), we can obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_{t+1}^{(1)} - h_{n,t+1}^{(1)}\|^2] \\
&\leq (1+1/p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|(1-\beta\eta)h_{n,t}^{(1)} - \frac{1}{N}\sum_{n'=1}^{N}(1-\beta\eta)h_{n',t}^{(1)}\|^2] \\
&\quad + (1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\beta\eta F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)}) - \beta\eta\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t+1};\xi_{n',t+1}^{(1)})\|^2] \ .
\end{aligned} \tag{19}$$

Additionally, we have

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)}) - \frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t+1};\xi_{n',t+1}^{(1)})\|^2] \\
&= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)}) - F_n^{(1)}(x_{n,t+1}) + F_n^{(1)}(x_{n,t+1}) - F_n^{(1)}(\bar{x}_{t+1}) + F_n^{(1)}(\bar{x}_{t+1}) - F^{(1)}(\bar{x}_{t+1}) \\
&\quad + F^{(1)}(\bar{x}_{t+1}) - \frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t+1}) + \frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t+1}) - \frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t+1};\xi_{n',t+1}^{(1)})\|^2] \\
&\leq 5\sigma^2 + 5C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t+1} - \bar{x}_{t+1}\|^2] + 5\delta^2 + 5C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t+1} - \bar{x}_{t+1}\|^2] + 5\sigma^2 \\
&\leq 10\sigma^2 + 5\delta^2 + 60\alpha^2\gamma^2p^4\eta^4C^6 \ ,
\end{aligned} \tag{20}$$

where the last step holds due to Lemma B.1.

By combining them together, we can obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_{t+1}^{(1)} - h_{n,t+1}^{(1)}\|^2] \\
&\leq (1+1/p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_t^{(1)} - h_{n,t}^{(1)}\|^2] + 2p\beta^2\eta^2(10\sigma^2 + 5\delta^2 + 60\alpha^2\gamma^2p^4\eta^4C^6)
\end{aligned}$$

$$
\begin{aligned}
&\leq 2p\beta^2\eta^2(10\sigma^2+5\delta^2+60\alpha^2\gamma^2p^4\eta^4C^6)\sum_{t'=s_tp}^{t}(1+1/p)^{t-t'} \\
&\leq 60p^2\beta^2\eta^2\sigma^2+30p^2\beta^2\eta^2\delta^2+360\alpha^2\beta^2\gamma^2p^6\eta^6C^6\ . \qquad (21)
\end{aligned}
$$

□

Lemma B.3. *Suppose Assumptions 3.1-3.3 hold, and $\eta \leq \frac{1}{2\gamma L_\Phi}$, then we have*

$$
\begin{aligned}
\mathbb{E}[\Phi(\bar{x}_{t+1})] &\leq \mathbb{E}[\Phi(\bar{x}_t)] - \frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] - \frac{\gamma\eta}{4}\mathbb{E}[\|\bar{v}_t\|^2] + \gamma\eta\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))-\bar{v}_t\Big\|^2\Big] \\
&+ 12\gamma\eta C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|h_{n,t}^{(1)}-F_n^{(1)}(x_{n,t})\Big\|^2\Big] + 6\gamma\eta C^2\delta^2 + 36\gamma\eta\alpha^2\gamma^2p^4\eta^4C^6L^2(1+C^2) \\
&+ 360\gamma\eta p^2\beta^2\eta^2\sigma^2C^2L^2 + 180\gamma\eta p^2\beta^2\eta^2\delta^2C^2L^2 + 1080\gamma\eta\alpha^2\beta^2\gamma^2p^6\eta^6C^8L^2\ . \qquad (22)
\end{aligned}
$$

Proof. At first, due to the smoothness of $\Phi(\bar{x}_t)$, we have

$$
\begin{aligned}
&\mathbb{E}[\Phi(\bar{x}_{t+1})] \leq \mathbb{E}[\Phi(\bar{x}_t)] + \mathbb{E}[\langle\nabla\Phi(\bar{x}_t), \bar{x}_{t+1}-\bar{x}_t\rangle] + \frac{L_\Phi}{2}\mathbb{E}[\|\bar{x}_{t+1}-\bar{x}_t\|^2] \\
&= \mathbb{E}[\Phi(\bar{x}_t)] - \gamma\eta\mathbb{E}[\langle\nabla\Phi(\bar{x}_t), \bar{v}_t\rangle] + \frac{\gamma^2\eta^2L_\Phi}{2}\mathbb{E}[\|\bar{v}_t\|^2] \\
&= \mathbb{E}[\Phi(\bar{x}_t)] - \frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] + \Big(\frac{\gamma^2\eta^2L_\Phi}{2}-\frac{\gamma\eta}{2}\Big)\mathbb{E}[\|\bar{v}_t\|^2] + \frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)-\bar{v}_t\|^2] \\
&\leq \mathbb{E}[\Phi(\bar{x}_t)] - \frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] - \frac{\gamma\eta}{4}\mathbb{E}[\|\bar{v}_t\|^2] \\
&\quad + \gamma\eta\mathbb{E}\Big[\Big\|\nabla\Phi(\bar{x}_t)-\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))\Big\|^2\Big] + \gamma\eta\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))-\bar{v}_t\Big\|^2\Big] \\
&\leq \mathbb{E}[\Phi(\bar{x}_t)] - \frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] - \frac{\gamma\eta}{4}\mathbb{E}[\|\bar{v}_t\|^2] + \gamma\eta\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))-\bar{v}_t\Big\|^2\Big] \\
&\quad + 12\gamma\eta C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|h_{n,t}^{(1)}-F_n^{(1)}(x_{n,t})\Big\|^2\Big] + 6\gamma\eta C^2\delta^2 + 36\gamma\eta\alpha^2\gamma^2p^4\eta^4C^6L^2(1+C^2) \\
&\quad + 360\gamma\eta p^2\beta^2\eta^2\sigma^2C^2L^2 + 180\gamma\eta p^2\beta^2\eta^2\delta^2C^2L^2 + 2160\gamma\eta\alpha^2\beta^2\gamma^2p^6\eta^6C^8L^2\ , \qquad (23)
\end{aligned}
$$

where the second to last step follows from $\eta \leq \frac{1}{2\gamma L_\Phi}$, and the last step follows from the following inequality.

$$
\begin{aligned}
&\mathbb{E}\Big[\Big\|\nabla\Phi(\bar{x}_t)-\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))\Big\|^2\Big] \\
&= \mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(\bar{x}_t)\Big)\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t)\Big) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(F_n^{(1)}(x_{n,t})\Big)\Big\|^2\Big] \\
&\leq 6\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(\bar{x}_t)\Big)\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t)\Big) \\
&\quad - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(\bar{x}_t)\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t)\Big)\Big\|^2\Big] \\
&+ 6\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(\bar{x}_t)\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t)\Big) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t)\Big)\Big\|^2\Big]
\end{aligned}
$$

$$
\begin{aligned}
&+ 6\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t)\Big) \\
&\quad - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t})\Big)\Big\|^2\Big] \\
&+ 6\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t})\Big) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}h_{n',t}^{(1)}\Big)\Big\|^2\Big] \\
&+ 6\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(\frac{1}{N}\sum_{n'=1}^{N}h_{n',t}^{(1)}\Big) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(h_{n,t}^{(1)}\Big)\Big\|^2\Big] \\
&+ 6\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(h_{n,t}^{(1)}\Big) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t})\nabla_{F^{(1)}}F_n^{(2)}\Big(F_n^{(1)}(x_{n,t})\Big)\Big\|^2\Big] \\
&\leq 6C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n'=1}^{N}\nabla F_{n'}^{(1)}(\bar{x}_t) - \nabla F_n^{(1)}(\bar{x}_t)\Big\|^2\Big] + 6C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\nabla F_n^{(1)}(\bar{x}_t) - \nabla F_n^{(1)}(x_{n,t})\Big\|^2\Big] \\
&+ 6C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(\bar{x}_t) - \frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t})\Big\|^2\Big] \\
&+ 6C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n'=1}^{N}F_{n'}^{(1)}(x_{n',t}) - \frac{1}{N}\sum_{n'=1}^{N}h_{n',t}\Big\|^2\Big] \\
&+ 6C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n'=1}^{N}h_{n',t}^{(1)} - h_{n,t}^{(1)}\Big\|^2\Big] + 6C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|h_{n,t}^{(1)} - F_n^{(1)}(x_{n,t})\Big\|^2\Big] \\
&\leq 6C^2\delta^2 + 6C^2L^2(C^2+1)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\bar{x}_t - x_{n,t}\Big\|^2\Big] + 6C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|\bar{h}_t^{(1)} - h_{n,t}^{(1)}\Big\|^2\Big] \\
&+ 12C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|h_{n,t}^{(1)} - F_n^{(1)}(x_{n,t})\Big\|^2\Big] \\
&\leq 6C^2\delta^2 + 36\alpha^2\gamma^2p^4\eta^4C^6L^2(1+C^2) + 360p^2\beta^2\eta^2\sigma^2C^2L^2 + 180p^2\beta^2\eta^2\delta^2C^2L^2 + 2160\alpha^2\beta^2\gamma^2p^6\eta^6C^8L^2 \\
&+ 12C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|h_{n,t}^{(1)} - F_n^{(1)}(x_{n,t})\Big\|^2\Big]\ , \qquad (24)
\end{aligned}
$$

where the last step holds due to Lemma B.1 and Lemma B.2.

□

Lemma B.4. *Suppose Assumptions 3.1-3.3 hold and $\eta \leq \frac{1}{\beta}$, then*

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t+1}) - h_{n,t+1}^{(1)}\Big\|^2\Big] \\
&\leq (1-\beta\eta)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t}) - h_{n,t}^{(1)}\Big\|^2\Big] + \frac{2\eta\gamma^2C^2}{\beta}\mathbb{E}[\|\bar{v}_t\|^2] + \frac{12\gamma^2\alpha^2p^2\eta^3C^6}{\beta} + \beta^2\eta^2\sigma^2\ . \qquad (25)
\end{aligned}
$$

Proof.

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t+1}) - h_{n,t+1}^{(1)}\Big\|^2\Big] \\
&\leq (1-\beta\eta)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t}) - h_{n,t}^{(1)}\Big\|^2\Big] + \frac{\eta\gamma^2C^2}{\beta}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}\|^2] + \beta^2\eta^2\sigma^2
\end{aligned}
$$

$$\leq (1-\beta\eta)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t})-h_{n,t}^{(1)}\Big\|^2\Big]+\frac{2\eta\gamma^2C^2}{\beta}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}-\bar{v}_t\|^2]+\frac{2\eta\gamma^2C^2}{\beta}\mathbb{E}[\|\bar{v}_t\|^2]+\beta^2\eta^2\sigma^2$$

$$\leq (1-\beta\eta)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t})-h_{n,t}^{(1)}\Big\|^2\Big]+\frac{2\eta\gamma^2C^2}{\beta}\mathbb{E}[\|\bar{v}_t\|^2]+\frac{12\gamma^2\alpha^2p^2\eta^3C^6}{\beta}+\beta^2\eta^2\sigma^2\ , \tag{26}$$

where the first step holds due to Lemma 4.2 in (Gao et al., 2022), and the last step holds due to Eq. (16).

□

Lemma B.5. *Suppose Assumptions 3.1-3.3 hold and $\eta\leq\frac{1}{\alpha}$, then*

$$\begin{aligned}
&\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\bar{v}_{t+1}\Big\|^2\Big]\\
&\leq (1-\alpha\eta)\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}\|^2\Big]+2\alpha\eta C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t})-h_{n,t}^{(1)}\Big\|^2\Big]\\
&\quad+\Big(\frac{2\alpha\eta^2\gamma^2L^2C^4}{\beta}+\frac{2\eta\gamma^2L_\Phi^2}{\alpha}\Big)\mathbb{E}\Big[\Big\|\bar{v}_t\Big\|^2\Big]+\Big(\frac{2\alpha\eta^2\gamma^2L^2C^4}{\beta}+\frac{2\eta\gamma^2L_\Phi^2}{\alpha}\Big)6p^2\alpha^2\eta^2C^4\\
&\quad+2\alpha\beta^2\eta^3C^2L^2\sigma^2+4\alpha^2\eta^2\frac{C^2\sigma^2}{N}\ .
\end{aligned}\tag{27}$$

Proof.

$$\begin{aligned}
&\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\bar{v}_{t+1}\Big\|^2\Big]\\
&\leq \mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta)v_{n,t}\\
&\quad-\alpha\eta\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)};\xi_{n,t+1}^{(2)})\Big\|^2\Big]\\
&=\mathbb{E}\Big[\Big\|(1-\alpha\eta)\Big(\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}\Big)\\
&\quad+(1-\alpha\eta)\Big(\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))\Big)\\
&\quad+\alpha\eta\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla F_n^{(1)}(x_{n,t+1})^T\nabla F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\nabla F_n^{(1)}(x_{n,t+1})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)})\Big)\Big\|^2\Big]\\
&\quad+\alpha^2\eta^2\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla F_n^{(1)}(x_{n,t+1})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)})-\nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)})\\
&\qquad+\nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)})-\nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)};\xi_{n,t+1}^{(2)})\Big)\Big\|^2\Big]\\
&\leq (1-\alpha\eta)^2(1+a)\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}\|^2\Big]\\
&\quad+2(1-\alpha\eta)^2(1+1/a)\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t}))\Big\|^2\Big]\\
&\quad+2\alpha^2\eta^2(1+1/a)\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla F_n^{(1)}(x_{n,t+1})^T\nabla F_n^{(2)}(F_n^{(1)}(x_{n,t+1}))-\nabla F_n^{(1)}(x_{n,t+1})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)})\Big)\Big\|^2\Big]
\end{aligned}$$

$$
\begin{aligned}
&+ 2\alpha^2\eta^2\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla F_n^{(1)}(x_{n,t+1})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)}) - \nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)})\Big)\Big\|^2\Big] \\
&+ 2\alpha^2\eta^2\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)}) - \nabla F_n^{(1)}(x_{n,t+1};\xi_{n,t+1}^{(1)})^T\nabla F_n^{(2)}(h_{n,t+1}^{(1)};\xi_{n,t+1}^{(2)})\Big)\Big\|^2\Big] \\
&\leq (1-\alpha\eta)\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t})) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}\|^2\Big] + \frac{L_\Phi^2}{\alpha\eta}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|x_{n,t+1}-x_{n,t}\Big\|^2\Big] \\
&+ 2\alpha\eta C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t+1}) - h_{n,t+1}^{(1)}\Big\|^2\Big] + 4\alpha^2\eta^2\frac{C^2\sigma^2}{N} \\
&\leq (1-\alpha\eta)\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t})) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}\|^2\Big] + 2\alpha\eta C^2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t}) - h_{n,t}^{(1)}\Big\|^2\Big] \\
&+ \Big(\frac{2\alpha\eta^2\gamma^2L^2C^4}{\beta} + \frac{2\eta\gamma^2L_\Phi^2}{\alpha}\Big)\mathbb{E}\Big[\Big\|\bar{v}_t\Big\|^2\Big] + \Big(\frac{2\alpha\eta^2\gamma^2L^2C^4}{\beta} + \frac{2\eta\gamma^2L_\Phi^2}{\alpha}\Big)6p^2\alpha^2\eta^2C^4 \\
&+ 2\alpha\beta^2\eta^3C^2L^2\sigma^2 + 4\alpha^2\eta^2\frac{C^2\sigma^2}{N}\ ,
\end{aligned}
\tag{28}
$$

where the second step holds due to the independence between $\xi_{n,t+1}^{(1)}$ and $\xi_{n,t+1}^{(2)}$, the second to last step holds due to $a = \frac{\alpha\eta}{1-\alpha\eta}$ and the independence between $\xi_{n,t+1}^{(1)}$ and $\xi_{n,t+1}^{(2)}$, the last step holds due to Eq. (16). □

Proof. At first, we define the following potential function:

$$
\begin{aligned}
P_{t+1} = \mathbb{E}[\Phi(\bar{x}_{t+1})] &+ \frac{\gamma}{\alpha}\mathbb{E}\Big[\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla_x F_n^{(2)}(F_n^{(1)}(x_{n,t+1})) - \bar{v}_{t+1}\Big\|^2\Big] \\
&+ \frac{12\gamma C^2L^2}{\beta}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\Big[\Big\|F_n^{(1)}(x_{n,t+1}) - h_{n,t+1}^{(1)}\Big\|^2\Big]\ .
\end{aligned}
\tag{29}
$$

Then, based on Lemmas B.3, B.5, B.4, we can obtain

$$
\begin{aligned}
& P_{t+1} - P_t \\
&\leq -\frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \\
&\quad + \Big(\frac{\gamma}{\alpha}\big(\frac{2\alpha\eta^2\gamma^2L^2C^4}{\beta} + \frac{2\eta\gamma^2L_\Phi^2}{\alpha}\big) + \frac{12\gamma C^2L^2}{\beta}\frac{2\eta\gamma^2C^2}{\beta} - \frac{\gamma\eta}{4}\Big)\mathbb{E}[\|\bar{v}_t\|^2] \\
&\quad + 36\alpha^2\gamma^3p^4\eta^5C^6L^2(1+C^2) + 6\gamma\eta C^2\delta^2 + \frac{\gamma}{\alpha}\Big(\frac{2\alpha\eta^2\gamma^2L^2C^4}{\beta} + \frac{2\eta\gamma^2L_\Phi^2}{\alpha}\Big)6p^2\alpha^2\eta^2C^4 \\
&\quad + 2\frac{\gamma}{\alpha}\alpha\beta^2\eta^3C^2L^2\sigma^2 + 4\frac{\gamma}{\alpha}\alpha^2\eta^2\frac{C^2\sigma^2}{N} + \frac{12\gamma C^2L^2}{\beta}\frac{12\gamma^2\alpha^2p^2\eta^3C^6}{\beta} + \frac{12\gamma C^2L^2}{\beta}\beta^2\eta^2\sigma^2\ . \\
&\quad + 360\gamma\eta p^2\beta^2\eta^2\sigma^2C^2L^2 + 180\gamma\eta p^2\beta^2\eta^2\delta^2C^2L^2 + 2160\gamma\eta\alpha^2\beta^2\gamma^2p^6\eta^6C^8L^2\ .
\end{aligned}
\tag{30}
$$

By setting $\gamma \leq \frac{1}{2}\sqrt{1\Big/\Big(\frac{2L^2C^4}{\alpha\beta} + \frac{2L_\Phi^2}{\alpha^2} + \frac{24C^4L^2}{\beta^2}\Big)}$ and summing t from 0 to $T-1$, we can obtain

$$
\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] &\leq \frac{2(P_0-P_T)}{\gamma\eta T} + 72\alpha^2\gamma^2p^4\eta^4C^6L^2(1+C^2) + 12C^2\delta^2 + 12p^2\alpha\eta^2C^4\Big(\frac{2\alpha\eta\gamma^2L^2C^4}{\beta} + \frac{2\gamma^2L_\Phi^2}{\alpha}\Big) \\
&+ 4\beta^2\eta^2C^2L^2\sigma^2 + 8\alpha\eta\frac{C^2\sigma^2}{N} + \frac{288C^2L^2\gamma^2\alpha^2p^2\eta^2C^6}{\beta^2} + 24C^2L^2\beta\eta\sigma^2 \\
&+ 720p^2\beta^2\eta^2\sigma^2C^2L^2 + 360p^2\beta^2\eta^2\delta^2C^2L^2 + 4320\alpha^2\beta^2\gamma^2p^6\eta^6C^8L^2\ .
\end{aligned}
\tag{31}
$$

According to initialization step, it is easy to know that $P_0 = O(\frac{1}{\gamma\eta T}) + O(\frac{\gamma}{\alpha}\sigma^2) + O(\frac{\gamma}{\beta}\sigma^2)$. Therefore, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \leq O(\frac{1}{\gamma\eta T}) + O(\frac{\sigma^2}{\alpha\eta T}) + O(\frac{\sigma^2}{\beta\eta T}) + O(\alpha^2\gamma^2p^4\eta^4) + O(\delta^2) + O(\frac{\gamma^2\alpha^2p^2\eta^3}{\beta}) + O(p^2\eta^2\gamma^2)$$
$$+ O(\beta^2\eta^2\sigma^2) + O(\frac{\alpha\eta\sigma^2}{N}) + O(\frac{\gamma^2\alpha^2p^2\eta^2}{\beta^2}) + O(\beta\eta\sigma^2) + O(p^2\beta^2\eta^2\sigma^2) + O(p^2\beta^2\eta^2\delta^2) + O(\alpha^2\beta^2\gamma^2p^6\eta^6)\ . \tag{32}$$

□

C. Proof of Theorem 4.1

Lemma C.1. *Suppose Assumptions 3.1-3.3 hold, then* $\|\nabla\Phi^{(k)}(x)\| \leq C^k$ *and* $\|\Psi^{(k)}(x)\| \leq C_{\Psi^{(k)}} \triangleq C^{K-k+1}$, *where* $k \in \{1, \cdots, K\}$.

Proof. For $k \in \{1, \cdots, K\}$, we have

$$\begin{aligned}
&\|\nabla\Phi^{(k)}(x)\| \\
&= \|\nabla F^{(1)}(x)\nabla F^{(2)}(\Phi^{(1)}(x))\nabla F^{(3)}(\Phi^{(2)}(x))\cdots\nabla F^{(k)}(\Phi^{(k-1)}(x))\| \\
&\leq \|\nabla F^{(1)}(x)\|\|\nabla F^{(2)}(\Phi^{(1)}(x))\|\|\nabla F^{(3)}(\Phi^{(2)}(x))\|\cdots\|\nabla F^{(k)}(\Phi^{(k-1)}(x))\| \\
&\leq C^k \ .
\end{aligned} \tag{33}$$

For $k = K$, we have

$$\|\Psi^{(K)}(x)\| = \|\nabla F^{(K)}(h^{(K-1)}))\| \leq C \ . \tag{34}$$

Then, for $k \in \{K-1, \cdots, 1\}$, according to the definition of $\|\Psi^{(k)}(x)\|$, it is easy to know

$$\|\Psi^{(k)}(x)\| = \|\nabla F^{(k)}(h^{(k-1)})^T\Psi^{(k+1)}(x)\| \leq C\|\Psi^{(k+1)}(x)\| \leq C^{K-k+1} \ . \tag{35}$$

□

Lemma C.2. *Suppose Assumptions 3.1-3.3 hold, then* $\Phi(x)$ *is* L_Φ *smooth, where* $L_\Phi = L\sum_{k=1}^{K} C^{K+k-2}$.

Proof. For any $x \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_1}$, one can get

$$\begin{aligned}
&\|\nabla\Phi(x) - \nabla\Phi(y)\| \\
&= \|\nabla F^{(1)}(x)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(y))^T\nabla F^{(K)}(\Phi^{(K-1)}(y))\| \\
&\leq \|\nabla F^{(1)}(x)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad + \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad + \cdots \\
&\quad + \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(y))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(y))^T\nabla F^{(K)}(\Phi^{(K-1)}(y))\| \\
&\leq \|\nabla F^{(1)}(x)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x))\| \\
&\quad + \|\nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(x))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(x))^T\nabla F^{(K)}(\Phi^{(K-1)}(x))\| \\
&\quad + \cdots \\
&\quad + \|\nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(y))^T\nabla F^{(K)}(\Phi^{(K-1)}(x)) \\
&\quad - \nabla F^{(1)}(y)^T\nabla F^{(2)}(\Phi^{(1)}(y))^T\cdots\nabla F^{(K-1)}(\Phi^{(K-2)}(y))^T\nabla F^{(K)}(\Phi^{(K-1)}(y))\| \\
&\leq C^{K-1}\|\nabla F^{(1)}(x) - \nabla F^{(1)}(y)\| + C^{K-1}\|\nabla F^{(2)}(\Phi^{(1)}(x)) - \nabla F^{(2)}(\Phi^{(1)}(y))\| \\
&\quad + \cdots + C^{K-1}\|\nabla F^{(K)}(\Phi^{(K-1)}(x)) - \nabla F^{(K)}(\Phi^{(K-1)}(y))\| \\
&\leq C^{K-1}L\|x-y\| + C^{K-1}L\|\Phi^{(1)}(x) - \Phi^{(1)}(y)\| + \cdots + C^{K-1}L\|\Phi^{(K-1)}(x) - \Phi^{(K-1)}(y)\| \\
&\leq C^{K-1}L\|x-y\| + C^{K-1}LC\|x-y\| + \cdots + C^{K-1}LC^{K-1}\|x-y\|
\end{aligned}$$

$$= \sum_{k=1}^{K} C^{K+k-2} L\|x-y\| \,, \tag{36}$$

where the fourth and fifth steps hold due to Assumption 3.1, the second to last step holds due to Lemma C.1.

□

Lemma C.3. *Suppose Assumptions 3.1-3.3 hold, then*

$$\|\Psi(\bar{x}_t) - \nabla\Phi(\bar{x}_t)\|^2 \leq 2AK \sum_{k=1}^{K-1} \Big\| \frac{1}{N}\sum_{n=1}^{N} h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(h_{n,t}^{(k-1)}) \Big\|^2 + 2AKC^2 \sum_{k=1}^{K-1} \frac{1}{N}\sum_{n=1}^{N} \Big\| h_{n,t}^{(k-1)} - \bar{h}_t^{(k-1)} \Big\|^2 \,, \tag{37}$$

where $A = C^{2(K-1)}L^2\Big(\sum_{j=0}^{K-2} C^j\Big)^2$.

Proof.

$$\begin{aligned}
&\|\Psi(\bar{x}_t) - \nabla\Phi(\bar{x}_t)\| \\
&= \|\nabla F^{(1)}(\bar{x}_t)^T \nabla F^{(2)}(\bar{h}_t^{(1)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(1)}(\bar{x}_t)^T \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\Phi^{(K-2)}(\bar{x}_t))^T \nabla F^{(K)}(\Phi^{(K-1)}(\bar{x}_t))\| \\
&\leq C\|\nabla F^{(2)}(\bar{h}_t^{(1)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\Phi^{(K-2)}(\bar{x}_t))^T \nabla F^{(K)}(\Phi^{(K-1)}(\bar{x}_t))\| \\
&\leq C\|\nabla F^{(2)}(\bar{h}_t^{(1)})^T \nabla F^{(3)}(\bar{h}_t^{(2)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\bar{h}_t^{(2)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad + \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\bar{h}_t^{(2)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\Phi^{(2)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad + \cdots \\
&\quad + \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\Phi^{(2)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\Phi^{(K-2)}(\bar{x}_t))^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\Phi^{(2)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\Phi^{(K-2)}(\bar{x}_t))^T \nabla F^{(K)}(\Phi^{(K-1)}(\bar{x}_t))\| \\
&\leq C\|\nabla F^{(2)}(\bar{h}_t^{(1)})^T \nabla F^{(3)}(\bar{h}_t^{(2)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\bar{h}_t^{(2)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)})\| \\
&\quad + C\|\nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\bar{h}_t^{(2)})^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\Phi^{(2)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\bar{h}_t^{(K-2)})^T \nabla F^{(K)}(\bar{h}_t^{(K-1)})\| \\
&\quad + \cdots \\
&\quad + C\|\nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\Phi^{(2)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\Phi^{(K-2)}(\bar{x}_t))^T \nabla F^{(K)}(\bar{h}_t^{(K-1)}) \\
&\quad - \nabla F^{(2)}(\Phi^{(1)}(\bar{x}_t))^T \nabla F^{(3)}(\Phi^{(2)}(\bar{x}_t))^T \cdots \nabla F^{(K-1)}(\Phi^{(K-2)}(\bar{x}_t))^T \nabla F^{(K)}(\Phi^{(K-1)}(\bar{x}_t))\| \\
&\leq C^{K-1}L\|\bar{h}_t^{(1)} - \Phi^{(1)}(\bar{x}_t)\| + C^{K-1}L\|\bar{h}_t^{(2)} - \Phi^{(2)}(\bar{x}_t)\| + \cdots + C^{K-1}L\|\bar{h}_t^{(K-1)} - \Phi^{(K-1)}(\bar{x}_t)\| \\
&\leq C^{K-1}L \sum_{k=1}^{K-1}\sum_{i=1}^{k} C^{k-i}\|\bar{h}_t^{(i)} - F^{(i)}(\bar{h}_t^{(i-1)})\| \\
&= \sum_{k=1}^{K-1}\Big(C^{K-1}L\sum_{j=k}^{K-1} C^{j-k}\Big)\|\bar{h}_t^{(k)} - F^{(k)}(\bar{h}_t^{(k-1)})\| \,,
\end{aligned} \tag{38}$$

where the second step and fifth step hold due to Assumption 3.1, the second to last step follows from Lemma C.4. Then, due to

$$\Big(C^{K-1}L\sum_{j=k}^{K-1}C^{j-k}\Big)^2=C^{2(K-1)}L^2\Big(\sum_{j=k}^{K-1}C^{j-k}\Big)^2\leq C^{2(K-1)}L^2\Big(\sum_{j=0}^{K-2}C^{j}\Big)^2\,, \tag{39}$$

one can get

$$\|\Psi(\bar{x}_t)-\nabla\Phi(\bar{x}_t)\|^2\leq K\sum_{k=1}^{K-1}\Big(C^{K-1}L\sum_{j=k}^{K-1}C^{j-k}\Big)^2\|\bar{h}_t^{(k)}-F^{(k)}(\bar{h}_t^{(k-1)})\|^2\leq AK\sum_{k=1}^{K-1}\|\bar{h}_t^{(k)}-F^{(k)}(\bar{h}_t^{(k-1)})\|^2\,, \tag{40}$$

where $A=C^{2(K-1)}L^2\Big(\sum_{j=0}^{K-2}C^j\Big)^2$.

Additionally, one can get

$$\begin{aligned}
&\|\bar{h}_t^{(k)}-F^{(k)}(\bar{h}_t^{(k-1)})\|^2\\
&=\Big\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)}-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})+\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(\bar{h}_t^{(k-1)})\Big\|^2\\
&\leq 2\Big\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)}-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\Big\|^2+2\Big\|\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(\bar{h}_t^{(k-1)})\Big\|^2\\
&\leq 2\Big\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)}-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\Big\|^2+2C^2\frac{1}{N}\sum_{n=1}^{N}\Big\|h_{n,t}^{(k-1)}-\bar{h}_t^{(k-1)}\Big\|^2\,,
\end{aligned} \tag{41}$$

where the last step holds due to Assumption 3.1.

By combining above two inequalities, one can get

$$\|\Psi(\bar{x}_t)-\nabla\Phi(\bar{x}_t)\|^2\leq 2AK\sum_{k=1}^{K-1}\Big\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)}-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\Big\|^2+2AKC^2\sum_{k=1}^{K-1}\frac{1}{N}\sum_{n=1}^{N}\Big\|h_{n,t}^{(k-1)}-\bar{h}_t^{(k-1)}\Big\|^2\,. \tag{42}$$

□

Lemma C.4. *Suppose Assumptions 3.1-3.3 hold, then for $k\in\{1,\cdots,K-1\}$, one can get*

$$\|\bar{h}_t^{(k)}-\Phi^{(k)}(\bar{x}_t)\|\leq\sum_{i=1}^{k}C^{k-i}\|\bar{h}_t^{(i)}-F^{(i)}(\bar{h}_t^{(i-1)})\|\,. \tag{43}$$

Proof. When $k=1$, one can get

$$\|\bar{h}_t^{(1)}-\Phi^{(1)}(\bar{x}_t)\|=\|\bar{h}_t^{(1)}-F^{(1)}(\bar{h}_t^{(0)})\|\,. \tag{44}$$

When $k\in\{2,\cdots,K-1\}$, one can get

$$\begin{aligned}
&\|\bar{h}_t^{(k)}-\Phi^{(k)}(\bar{x}_t)\|\\
&=\|\bar{h}_t^{(k)}-F^{(k)}(\Phi^{(k-1)}(\bar{x}_t))\|\\
&=\|\bar{h}_t^{(k)}-F^{(k)}(\bar{h}_t^{(k-1)})+F^{(k)}(\bar{h}_t^{(k-1)})-F^{(k)}(\Phi^{(k-1)}(\bar{x}_t))\|\\
&\leq\|\bar{h}_t^{(k)}-F^{(k)}(\bar{h}_t^{(k-1)})\|+\|F^{(k)}(\bar{h}_t^{(k-1)})-F^{(k)}(\Phi^{(k-1)}(\bar{x}_t))\|
\end{aligned}$$

$$\leq \|\bar{h}_t^{(k)} - F^{(k)}(\bar{h}_t^{(k-1)})\| + C\|\bar{h}_t^{(k-1)} - \Phi^{(k-1)}(\bar{x}_t)\| \,, \tag{45}$$

where the last step holds due to Assumption 3.1.

Then, for $k \in \{1, \cdots, K-1\}$, one can get

$$\|\bar{h}_t^{(k)} - \Phi^{(k)}(\bar{x}_t)\| \leq \sum_{i=1}^{k} C^{k-i}\|\bar{h}_t^{(i)} - F^{(i)}(\bar{h}_t^{(i-1)})\| \,. \tag{46}$$

□

Lemma C.5. *Suppose Assumptions 3.1-3.3 hold, one can get*

$$\begin{aligned}
&\Big\|\frac{1}{N}\sum_{n=1}^{N}\Psi_n^{(1)}(\bar{x}_t) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(1)}\Big\|^2 \\
&\leq 2KC^{2K}\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{h}_t^{(K-1)} - h_{n,t}^{(K-1)}\Big\|^2 + 2KC^{2(K-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)}) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(K)}\Big\|^2 \\
&\quad + 3K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}C^{2k}\Big\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\Big\|^2 + 3KL^2C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big\|\bar{h}_t^{(k-1)} - h_{n,t}^{(k-1)}\Big\|^2 \\
&\quad + 3K\sum_{k=1}^{K-1}C^{2(k-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\|^2 \,.
\end{aligned} \tag{47}$$

Proof. When $k \in \{1, \cdots, K-1\}$, one can obtain

$$\begin{aligned}
&\Big\|\frac{1}{N}\sum_{n=1}^{N}\Psi_n^{(k)}(\bar{x}_t) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\| \\
&\leq \Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}\Psi_m^{(k+1)}(\bar{x}_t) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)}\Big\| \\
&\quad + \Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\| \\
&\leq C\Big\|\frac{1}{N}\sum_{m=1}^{N}\Psi_m^{(k+1)}(\bar{x}_t) - \frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)}\Big\| + \Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\| \,,
\end{aligned} \tag{48}$$

where the last step holds due to Assumption 3.1.

Then, by recursive expansion, we can obtain

$$\begin{aligned}
\Big\|\frac{1}{N}\sum_{n=1}^{N}\Psi_n^{(1)}(\bar{x}_t) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(1)}\Big\| &\leq C^{K-1}\Big\|\frac{1}{N}\sum_{m=1}^{N}\Psi_m^{(K)}(\bar{x}_t) - \frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(K)}\Big\| \\
&\quad + \sum_{k=1}^{K-1}C^{k-1}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\| \,.
\end{aligned} \tag{49}$$

Therefore, one can get

$$\begin{aligned}
\Big\|\frac{1}{N}\sum_{n=1}^{N}\Psi_n^{(1)}(\bar{x}_t) - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(1)}\Big\|^2 &\leq KC^{2(K-1)}\Big\|\frac{1}{N}\sum_{m=1}^{N}\Psi_m^{(K)}(\bar{x}_t) - \frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(K)}\Big\|^2 \\
&\quad + K\sum_{k=1}^{K-1}C^{2(k-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\|^2 \,.
\end{aligned} \tag{50}$$

Additionally, one can get

$$
\begin{aligned}
&\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)}-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\|^2\\
&\leq 3\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)}-\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T v_{n,t}^{(k+1)}\Big\|^2\\
&\quad+3\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(\bar{h}_t^{(k-1)})^T v_{n,t}^{(k+1)}-\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\Big\|^2\\
&\quad+3\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\|^2\\
&\leq 3C^2\frac{1}{N}\sum_{n=1}^{N}\Big\|\frac{1}{N}\sum_{m=1}^{N}v_{m,t}^{(k+1)}-v_{n,t}^{(k+1)}\Big\|^2+3L^2C^{2(K-k)}\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{h}_t^{(k-1)}-h_{n,t}^{(k-1)}\Big\|^2\\
&\quad+3\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\|^2, \qquad (51)
\end{aligned}
$$

where the last step holds due to Assumption 3.1.

Furthermore, when $k=K$, one can get

$$
\begin{aligned}
&\Big\|\frac{1}{N}\sum_{n=1}^{N}\Psi_n^{(K)}(\bar{x}_t)-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(K)}\Big\|^2\\
&\leq 2\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(\bar{h}_t^{(K-1)})-\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\Big\|^2+2\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(K)}\Big\|^2\\
&\leq 2C^2\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{h}_t^{(K-1)}-h_{n,t}^{(K-1)}\Big\|^2+2\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(K)}\Big\|^2, \qquad (52)
\end{aligned}
$$

where the last step holds due to Assumption 3.1.

Finally, by combining above three inequalities, one can get

$$
\begin{aligned}
&\Big\|\frac{1}{N}\sum_{n=1}^{N}\Psi_n^{(1)}(\bar{x}_t)-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(1)}\Big\|^2\\
&\leq 2KC^{2K}\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{h}_t^{(K-1)}-h_{n,t}^{(K-1)}\Big\|^2+2KC^{2(K-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(K)}\Big\|^2\\
&\quad+3K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}C^{2k}\Big\|\bar{v}_t^{(k+1)}-v_{n,t}^{(k+1)}\Big\|^2+3KL^2C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big\|\bar{h}_t^{(k-1)}-h_{n,t}^{(k-1)}\Big\|^2\\
&\quad+3K\sum_{k=1}^{K-1}C^{2(k-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}-\frac{1}{N}\sum_{n=1}^{N}v_{n,t}^{(k)}\Big\|^2. \qquad (53)
\end{aligned}
$$

□

Lemma C.6. *Suppose Assumptions 3.1-3.3 hold and* $\eta\leq\frac{1}{2\gamma L_\Phi}$*, one can get*

$$
\Phi\left(\bar{x}_{t+1}\right)\leq\Phi(\bar{x}_t)-\frac{\gamma\eta}{2}\Big\|\nabla\Phi(\bar{x}_t)\Big\|^2-\frac{\gamma\eta}{4}\Big\|\bar{v}_t^{(1)}\Big\|^2+2\gamma\eta AK\sum_{k=1}^{K-1}\Big\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)}-\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\Big\|^2
$$

$$+ 2\gamma\eta K C^{2(K-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)}) - \frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)}\Big\|^2 + 3\gamma\eta K C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\Big\|^2$$
$$+ 3\gamma\eta K\sum_{k=1}^{K-1} C^{2(k-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)}\Big\|^2$$
$$+ 3\gamma\eta K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}\Big\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\Big\|^2 + 3\gamma\eta K(L^2C^{2(K-1)} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{x}_t - x_{n,t}\Big\|^2$$
$$+ 3\gamma\eta K(L^2C^{2(K-1)} + C^{2K} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\Big\|^2 . \tag{54}$$

Proof. Due to the smoothness of $\Phi(\bar{x}_t)$, one can get

$$\Phi(\bar{x}_{t+1}) \le \Phi(\bar{x}_t) + \langle\nabla\Phi(\bar{x}_t), \bar{x}_{t+1} - \bar{x}_t\rangle + \frac{L_\Phi}{2}\|\bar{x}_{t+1} - \bar{x}_t\|^2$$
$$= \Phi(\bar{x}_t) - \gamma\eta\langle\nabla\Phi(\bar{x}_t), \bar{v}_t^{(1)}\rangle + \frac{\gamma^2\eta^2 L_\Phi}{2}\|\bar{v}_t^{(1)}\|^2$$
$$= \Phi(\bar{x}_t) - \frac{\gamma\eta}{2}\|\nabla\Phi(\bar{x}_t)\|^2 - (\frac{\gamma\eta}{2} - \frac{\gamma^2\eta^2 L_\Phi}{2})\|\bar{v}_t^{(1)}\|^2 + \frac{\gamma\eta}{2}\|\bar{v}_t^{(1)} - \nabla\Phi(\bar{x}_t)\|^2$$
$$\le \Phi(\bar{x}_t) - \frac{\gamma\eta}{2}\|\nabla\Phi(\bar{x}_t)\|^2 - \frac{\gamma\eta}{4}\|\bar{v}_t^{(1)}\|^2 + \frac{\gamma\eta}{2}\|\bar{v}_t^{(1)} - \nabla\Phi(\bar{x}_t)\|^2$$
$$\le \Phi(\bar{x}_t) - \frac{\gamma\eta}{2}\|\nabla\Phi(\bar{x}_t)\|^2 - \frac{\gamma\eta}{4}\|\bar{v}_t^{(1)}\|^2 + \gamma\eta\|\nabla\Phi(\bar{x}_t) - \Psi(\bar{x}_t)\|^2 + \gamma\eta\|\Psi(\bar{x}_t) - \bar{v}_t^{(1)}\|^2$$
$$\le \Phi(\bar{x}_t) - \frac{\gamma\eta}{2}\|\nabla\Phi(\bar{x}_t)\|^2 - \frac{\gamma\eta}{4}\|\bar{v}_t^{(1)}\|^2$$
$$+ 2\gamma\eta A K\sum_{k=1}^{K-1}\Big\|\frac{1}{N}\sum_{n=1}^{N} h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(h_{n,t}^{(k-1)})\Big\|^2 + 2\gamma\eta A K C^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big\|h_{n,t}^{(k-1)} - \bar{h}_t^{(k-1)}\Big\|^2$$
$$+ 2\gamma\eta K C^{2K}\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{h}_t^{(K-1)} - h_{n,t}^{(K-1)}\Big\|^2 + 2\gamma\eta K C^{2(K-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)}) - \frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)}\Big\|^2$$
$$+ 3\gamma\eta K\sum_{k=1}^{K-1} C^{2(k-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)}\Big\|^2$$
$$+ 3\gamma\eta K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1} C^{2k}\Big\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\Big\|^2 + 3\gamma\eta K L^2 C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big\|\bar{h}_t^{(k-1)} - h_{n,t}^{(k-1)}\Big\|^2$$
$$\le \Phi(\bar{x}_t) - \frac{\gamma\eta}{2}\Big\|\nabla\Phi(\bar{x}_t)\Big\|^2 - \frac{\gamma\eta}{4}\Big\|\bar{v}_t^{(1)}\Big\|^2 + 2\gamma\eta A K\sum_{k=1}^{K-1}\Big\|\frac{1}{N}\sum_{n=1}^{N} h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(h_{n,t}^{(k-1)})\Big\|^2$$
$$+ 2\gamma\eta K C^{2(K-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)}) - \frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)}\Big\|^2 + 3\gamma\eta K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}\Big\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\Big\|^2$$
$$+ 3\gamma\eta K\sum_{k=1}^{K-1} C^{2(k-1)}\Big\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)} - \frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)}\Big\|^2 + 3\gamma\eta K C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\Big\|^2$$
$$+ 3\gamma\eta K(L^2C^{2(K-1)} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\Big\|\bar{x}_t - x_{n,t}\Big\|^2$$
$$+ 3\gamma\eta K(L^2C^{2(K-1)} + C^{2K} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\Big\|^2 , \tag{55}$$

where the second inequality holds due to $\eta \le \frac{1}{2\gamma L_\Phi}$, the second to last inequality holds due to Lemma C.3 and Lemma C.5.

□

Lemma C.7. *Suppose Assumptions 3.1-3.3 hold,* $\eta \leq \frac{1}{20\gamma p C^{K-1} L}$, *and* $\eta \leq \frac{1}{6\gamma p (3C^2)^{K-1}}$, *one can get*

$$
\begin{aligned}
&\sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \leq \Big(96p^2\gamma^2\eta^2 C^{2(K-1)} L^2 + 32p^2\gamma^2\eta^2 (3C^2)^{2(K-1)}\Big) \sum_{t=0}^{T-1} \mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
&\quad + 48p^2 (3C^2)^{K-1} L^2 \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=2}^{K-1} \mathbb{E}[\|h_{n,t+1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2] \\
&\quad + 32p^2\alpha^2\eta^4 (3C^2)^{K-1} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\quad + 48\alpha^2 p^2 \eta^4 \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=2}^{K-1} (3C^2)^{k-1} \mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\quad + 64p^2\alpha^2\eta^4 (3C^2)^{K-1} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-1} (2C^2)^{K-1-k} \mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\quad + 64p^2\alpha^2\eta^4\sigma^2 (3C^2)^{K-1} T \sum_{k=1}^{K-1} (2C^2)^{K-1-k} + 32p^2\alpha^2\eta^4\sigma^2 (3C^2)^{K-1} T \\
&\quad + 24\alpha^2 p^2 \eta^4 C^{2K} T + 48\alpha^2 p^2 \eta^4 C^{2(K-1)} \sigma^2 T \sum_{k=1}^{K-2} 3^k \, .
\end{aligned}
\tag{56}
$$

Proof.

$$
\begin{aligned}
&\frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \\
&= \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|\Pi_{C^K}((1-\alpha\eta^2)(v_{n,t-1}^{(1)} - \nabla F_n^{(1)}(h_{n,t-1}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)}) + \nabla F_n^{(1)}(h_{n,t}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t}^{(2)}) \\
&\quad - \frac{1}{N} \sum_{m=1}^{N} \Pi_{C^K}((1-\alpha\eta^2)(v_{m,t-1}^{(1)} - \nabla F_m^{(1)}(h_{m,t-1}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t-1}^{(2)}) + \nabla F_m^{(1)}(h_{m,t}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t}^{(2)})\|^2] \\
&\leq \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|(1-\alpha\eta^2)(v_{n,t-1}^{(1)} - \nabla F_n^{(1)}(h_{n,t-1}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)}) + \nabla F_n^{(1)}(h_{n,t}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t}^{(2)} \\
&\quad - \frac{1}{N} \sum_{m=1}^{N} (1-\alpha\eta^2)(v_{m,t-1}^{(1)} - \nabla F_m^{(1)}(h_{m,t-1}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t-1}^{(2)}) - \nabla F_m^{(1)}(h_{m,t}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t}^{(2)}\|^2] \\
&\leq (1+\frac{1}{p})(1-\alpha\eta^2)^2 \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t-1}^{(1)} - \frac{1}{N} \sum_{m=1}^{N} v_{m,t-1}^{(1)}\|^2] \\
&\quad + 2(1+p) \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\| - \nabla F_n^{(1)}(h_{n,t-1}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)} + \nabla F_n^{(1)}(h_{n,t}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t}^{(2)} \\
&\quad + \frac{1}{N} \sum_{m=1}^{N} \nabla F_m^{(1)}(h_{m,t-1}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t-1}^{(2)} - \frac{1}{N} \sum_{m=1}^{N} \nabla F_m^{(1)}(h_{m,t}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t}^{(2)}\|^2] \\
&\quad + 2(1+p)\alpha^2\eta^4 \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|\nabla F_n^{(1)}(h_{n,t-1}^{(0)}; \xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)} - \frac{1}{N} \sum_{m=1}^{N} \nabla F_m^{(1)}(h_{m,t-1}^{(0)}; \xi_{m,t}^{(1)})^T v_{m,t-1}^{(2)}\|^2]
\end{aligned}
$$

$$
\begin{aligned}
&\leq (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(1)}-\bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 2(1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|-\nabla F_n^{(1)}(h_{n,t-1}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)} + \nabla F_n^{(1)}(h_{n,t}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t}^{(2)}\|^2] \\
&\quad + 2(1+p)\alpha^2\eta^4\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(1)}(h_{n,t-1}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)} - \frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(1)}(h_{m,t-1}^{(0)};\xi_{m,t}^{(1)})^T v_{m,t-1}^{(2)}\|^2] \ ,
\end{aligned} \tag{57}
$$

where the second step holds because $\Pi_{C^K}(\cdot)$ is a linear operator. The second term is bounded as follows:

$$
\begin{aligned}
&\mathbb{E}[\|\nabla F_n^{(1)}(h_{n,t}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t}^{(2)} - \nabla F_n^{(1)}(h_{n,t-1}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)}\|^2] \\
&\leq 2\mathbb{E}[\|\nabla F_n^{(1)}(h_{n,t}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t}^{(2)} - \nabla F_n^{(1)}(h_{n,t}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)}\|^2] \\
&\quad + 2\mathbb{E}[\|\nabla F_n^{(1)}(h_{n,t}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)} - \nabla F_n^{(1)}(h_{n,t-1}^{(0)};\xi_{n,t}^{(1)})^T v_{n,t-1}^{(2)}\|^2] \\
&\leq 2C^2\mathbb{E}[\|v_{n,t}^{(2)} - v_{n,t-1}^{(2)}\|^2] + 2\gamma^2\eta^2 C^{2(K-1)}L^2\mathbb{E}[\|v_{n,t-1}^{(1)}\|^2] \\
&\leq 2C^2\mathbb{E}[\|v_{n,t}^{(2)} - v_{n,t-1}^{(2)}\|^2] + 4\gamma^2\eta^2 C^{2(K-1)}L^2\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 4\gamma^2\eta^2 C^{2(K-1)}L^2\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \ ,
\end{aligned} \tag{58}
$$

where the last two steps hold due to Assumption 3.1. By combining them together, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \\
&\leq (1+\frac{1}{p}+16p\gamma^2\eta^2C^{2(K-1)}L^2)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 8pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(2)} - v_{n,t-1}^{(2)}\|^2] + 16p\gamma^2\eta^2C^{2(K-1)}L^2\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] + 4p\alpha^2\eta^4C^{2K} \ .
\end{aligned} \tag{59}
$$

By setting $\eta \leq \frac{1}{20\gamma p C^{K-1} L}$, one can get

$$
\begin{aligned}
&1+\frac{1}{p}+16p\gamma^2\eta^2C^{2(K-1)}L^2 \\
&\leq 1+\frac{1}{p}+16p\gamma^2C^{2(K-1)}L^2\frac{1}{400\gamma^2p^2C^{2(K-1)}L^2} \\
&\leq 1+\frac{26}{25p} \ .
\end{aligned} \tag{60}
$$

Thus, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \\
&\leq (1+\frac{26}{25p})\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 8pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(2)} - v_{n,t-1}^{(2)}\|^2] + 16p\gamma^2\eta^2C^{2(K-1)}L^2\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] + 4p\alpha^2\eta^4C^{2K} \\
&\leq 8pC^2\sum_{t'=s_tp}^{t-1}(1+\frac{26}{25p})^{t-1-t'}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t'+1}^{(2)} - v_{n,t'}^{(2)}\|^2] + 16p\gamma^2\eta^2C^{2(K-1)}L^2\sum_{t'=s_tp}^{t-1}(1+\frac{26}{25p})^{t-1-t'}\mathbb{E}[\|\bar{v}_{t'}^{(1)}\|^2]
\end{aligned}
$$

$$
\begin{aligned}
&+ 4p\alpha^2\eta^4 C^{2K} \sum_{t'=s_t p}^{t-1} (1+\frac{26}{25p})^{t-1-t'} \\
&\leq 24pC^2 \sum_{t'=s_t p}^{t-1} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[\|v_{n,t'+1}^{(2)} - v_{n,t'}^{(2)}\|^2] + 48p\gamma^2\eta^2 C^{2(K-1)} L^2 \sum_{t'=s_t p}^{t-1} \mathbb{E}[\|\bar{v}_{t'}^{(1)}\|^2] + 12p^2\alpha^2\eta^4 C^{2K} \ . \qquad (61)
\end{aligned}
$$

By summing over t from 0 to $T-1$, one can get

$$
\begin{aligned}
&\sum_{t=0}^{T-1} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \\
&\leq 24pC^2 \sum_{t=0}^{T-1}\sum_{t'=s_t p}^{t-1} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[\|v_{n,t'+1}^{(2)} - v_{n,t'}^{(2)}\|^2] + 48p\gamma^2\eta^2 C^{2(K-1)} L^2 \sum_{t=0}^{T-1}\sum_{t'=s_t p}^{t-1} \mathbb{E}[\|\bar{v}_{t'}^{(1)}\|^2] + 12p^2\alpha^2\eta^4 C^{2K} T \\
&\leq 24p^2C^2 \sum_{t=0}^{T-1} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[\|v_{n,t+1}^{(2)} - v_{n,t}^{(2)}\|^2] + 48p^2\gamma^2\eta^2 C^{2(K-1)} L^2 \sum_{t=0}^{T-1} \mathbb{E}[\|\bar{v}_t^{(1)}\|^2] + 12p^2\alpha^2\eta^4 C^{2K} T \\
&\leq 48p^2\gamma^2\eta^2 C^{2(K-1)} L^2 \sum_{t=0}^{T-1} \mathbb{E}[\|\bar{v}_t^{(1)}\|^2] + 12\alpha^2p^2\eta^4C^{2K}T + 24\alpha^2p^2\eta^4C^{2(K-1)}\sigma^2 T\sum_{k=1}^{K-2} 3^k \\
&\quad + 24p^2C^2(3C^2)^{K-2}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t+1}^{(K)} - v_{n,t}^{(K)}\|^2] \\
&\quad + 24p^2C^{2(K-1)}L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1} 3^{k-1}\mathbb{E}[\|h_{n,t+1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2] \\
&\quad + 72\alpha^2p^2\eta^4C^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}(3C^2)^{k-2}\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\leq \Big(48p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 16p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
&\quad + 12\alpha^2p^2\eta^4C^{2K}T + 24\alpha^2p^2\eta^4C^{2(K-1)}\sigma^2T\sum_{k=1}^{K-2}3^k \\
&\quad + 24p^2(3C^2)^{K-1}L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t+1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2] \\
&\quad + 16p^2\alpha^2\eta^4(3C^2)^{K-1}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\quad + 24\alpha^2p^2\eta^4\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}(3C^2)^{k-1}\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\quad + 16p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \\
&\quad + 32p^2\alpha^2\eta^4(3C^2)^{K-1}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\quad + 32p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}T\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 16p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}T \ , \qquad (62)
\end{aligned}
$$

where the last two inequalities hold due to Lemma C.17.

By setting $\eta \leq \frac{1}{6\gamma p(3C^2)^{K-1}}$, one can get $1 - 16p^2\gamma^2\eta^2(3C^2)^{2K-2} \geq \frac{1}{2}$. Then,

$$
\begin{aligned}
&\sum_{t=0}^{T-1} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \\
&\leq \Big(96p^2\gamma^2\eta^2 C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
&\quad + 24\alpha^2p^2\eta^4C^{2K}T + 48\alpha^2p^2\eta^4C^{2(K-1)}\sigma^2T\sum_{k=1}^{K-2}3^k \\
&\quad + 48p^2(3C^2)^{K-1}L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t+1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2] \\
&\quad + 32p^2\alpha^2\eta^4(3C^2)^{K-1}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\quad + 48\alpha^2p^2\eta^4\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}(3C^2)^{k-1}\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^Tv_{n,t}^{(k+1)}\|^2] \\
&\quad + 64p^2\alpha^2\eta^4(3C^2)^{K-1}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\quad + 64p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}T\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 32p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}T\,. \qquad (63)
\end{aligned}
$$

□

Lemma C.8. *Suppose Assumptions 3.1-3.3 hold, one can get*

$$
\begin{aligned}
&\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\|^2] \leq 24\alpha^2p^2\eta^4L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\quad + 24\gamma^2p^2\eta^2(2C^2)^{K-1}L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] + 24\gamma^2p^2\eta^2(2C^2)^{K-1}L^2\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
&\quad + 48p^2\alpha^2\eta^4C^2T + 24\alpha^2p^2\eta^4\sigma^2L^2T\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\,. \qquad (64)
\end{aligned}
$$

Proof. When $k = K-1$, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\|^2] \\
&= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}\Pi_C((1-\alpha\eta^2)(v_{m,t-1}^{(K)} - \nabla F_m^{(K)}(h_{m,t-1}^{(K-1)};\xi_{m,t}^{(K)}) + \nabla F_m^{(K)}(h_{m,t}^{(K-1)};\xi_{m,t}^{(K)}) \\
&\quad - \Pi_C((1-\alpha\eta^2)(v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&\leq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}(1-\alpha\eta^2)(v_{m,t-1}^{(K)} - \nabla F_m^{(K)}(h_{m,t-1}^{(K-1)};\xi_{m,t}^{(K)})) + \frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(K)}(h_{m,t}^{(K-1)};\xi_{m,t}^{(K)}) \\
&\quad - (1-\alpha\eta^2)(v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})) - \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&\leq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}(1-\alpha\eta^2)(v_{m,t-1}^{(K)} - v_{n,t-1}^{(K)})
\end{aligned}
$$

$$
\begin{aligned}
& - \frac{1}{N}\sum_{m=1}^{N} \nabla F_m^{(K)}(h_{m,t-1}^{(K-1)};\xi_{m,t}^{(K)}) + \frac{1}{N}\sum_{m=1}^{N} \nabla F_m^{(K)}(h_{m,t}^{(K-1)};\xi_{m,t}^{(K)}) \\
& + \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) - \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)}) \\
& - \alpha\eta^2 \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) + \frac{1}{N}\sum_{m=1}^{N} \alpha\eta^2 \nabla F_m^{(K)}(h_{m,t-1}^{(K-1)};\xi_{m,t}^{(K)})\|^2] \\
\leq\ & (1+\frac{1}{p})(1-\alpha\eta^2)^2 \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N} v_{m,t-1}^{(K)} - v_{n,t-1}^{(K)}\|^2] \\
& + 2(1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) - \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
& + 2(1+p)\alpha^2\eta^4 \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\| - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) + \frac{1}{N}\sum_{m=1}^{N} \nabla F_m^{(K)}(h_{m,t-1}^{(K-1)};\xi_{m,t}^{(K)})\|^2] \\
\leq\ & (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_{t-1}^{(K)} - v_{n,t-1}^{(K)}\|^2] + 4pL^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(K-1)} - h_{n,t}^{(K-1)}\|^2] + 16p\alpha^2\eta^4 C^2 \ ,
\end{aligned} \tag{65}
$$

where the last step holds due to Assumption 3.1. Similarly, one can get

$$
\begin{aligned}
& \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\|^2] \\
& \leq 4pL^2 \sum_{t'=s_t p}^{t-1}(1+\frac{1}{p})^{t-1-t'}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t'}^{(K-1)} - h_{n,t'+1}^{(K-1)}\|^2] + 16p\alpha^2\eta^4C^2\sum_{t'=s_t p}^{t-1}(1+\frac{1}{p})^{t-1-t'} \\
& \leq 12pL^2 \sum_{t'=s_t p}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t'}^{(K-1)} - h_{n,t'+1}^{(K-1)}\|^2] + 48p^2\alpha^2\eta^4C^2 \ .
\end{aligned} \tag{66}
$$

By summing over t from 0 to $T-1$, one can get

$$
\begin{aligned}
& \sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\|^2] \\
\leq\ & 12pL^2\sum_{t=0}^{T-1}\sum_{t'=s_t p}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t'}^{(K-1)} - h_{n,t'+1}^{(K-1)}\|^2] + 48p^2\alpha^2\eta^4C^2T \\
\leq\ & 12p^2L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(K-1)} - h_{n,t+1}^{(K-1)}\|^2] + 48p^2\alpha^2\eta^4C^2T \\
\leq\ & 24\alpha^2p^2\eta^4L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
& + 24\gamma^2p^2\eta^2(2C^2)^{K-1}L^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] + 24\gamma^2p^2\eta^2(2C^2)^{K-1}L^2\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
& + 48p^2\alpha^2\eta^4C^2T + 24\alpha^2p^2\eta^4\sigma^2L^2T\sum_{k=1}^{K-1}(2C^2)^{K-1-k} \ ,
\end{aligned} \tag{67}
$$

where the second to last step holds due to Lemma C.16, the last step holds due to Lemma. (C.7). □

Lemma C.9. *Suppose Assumptions 3.1-3.3 hold, one can get*

$$
\begin{aligned}
&\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2}C^{2k}\mathbb{E}[\|\bar{v}_t^{(k+1)}-v_{n,t}^{(k+1)}\|^2]\\
&\leq 24p^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2}C^{2k}\mathbb{E}[\|v_{n,t}^{(k+1)}-v_{n,t+1}^{(k+1)}\|^2]+24p^2L^2C^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t}^{(k-1)}-h_{n,t+1}^{(k-1)}\|^2]\\
&\quad+2\alpha^2\eta^4 24p^2C^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(K)}-\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2]\\
&\quad+96\alpha^2p^2\eta^4C^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t}^{(k)}-F_n^{(k)}(h_{n,t}^{(k-1)})\|^2]\\
&\quad+48\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)}-\bar{v}_t^{(1)}\|^2]+48\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2]\\
&\quad+96\alpha^2p^2\eta^4\sigma^2C^{2(K-1)}\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}+48\alpha^2p^2\eta^4\sigma^2C^{2(K-1)}T+48p^2\alpha^2\eta^4C^{2K}KT\,. \qquad (68)
\end{aligned}
$$

Proof. When $k\in\{1,\cdots,K-1\}$, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(k+1)}-v_{n,t}^{(k+1)}\|^2]\\
&=\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}\Pi_{C^{K-k}}((1-\alpha\eta^2)(v_{m,t-1}^{(k+1)}-\nabla F_m^{(k+1)}(h_{m,t-1}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t-1}^{(k+2)})\\
&\qquad+\nabla F_m^{(k+1)}(h_{m,t}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t}^{(k+2)})\\
&\quad-\Pi_{C^{K-k}}((1-\alpha\eta^2)(v_{n,t-1}^{(k+1)}-\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t-1}^{(k+2)})+\nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t}^{(k+2)})\|^2]\\
&\leq\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}(1-\alpha\eta^2)(v_{m,t-1}^{(k+1)}-\nabla F_m^{(k+1)}(h_{m,t-1}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t-1}^{(k+2)})\\
&\qquad+\frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(k+1)}(h_{m,t}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t}^{(k+2)}\\
&\quad-(1-\alpha\eta^2)(v_{n,t-1}^{(k+1)}-\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t-1}^{(k+2)})-\nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t}^{(k+2)}\|^2]\\
&\leq\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}(1-\alpha\eta^2)(v_{m,t-1}^{(k+1)}-v_{n,t-1}^{(k+1)})\\
&\quad-\frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(k+1)}(h_{m,t-1}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t-1}^{(k+2)}+\frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(k+1)}(h_{m,t}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t}^{(k+2)}\\
&\quad+\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t-1}^{(k+2)}-\nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t}^{(k+2)}\\
&\quad-\alpha\eta^2\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^Tv_{n,t-1}^{(k+2)}+\alpha\eta^2\frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(k+1)}(h_{m,t-1}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t-1}^{(k+2)}\|^2]\\
&\leq(1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}(1-\alpha\eta^2)(v_{m,t-1}^{(k+1)}-v_{n,t-1}^{(k+1)})\|^2]\\
&\quad+2(1+p)\frac{1}{N}\sum_{n=1}^{N}\|-\frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(k+1)}(h_{m,t-1}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t-1}^{(k+2)}+\frac{1}{N}\sum_{m=1}^{N}\nabla F_m^{(k+1)}(h_{m,t}^{(k)};\xi_{m,t}^{(k+1)})^Tv_{m,t}^{(k+2)}
\end{aligned}
$$

$$
\begin{aligned}
&+ \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} - \nabla F_n^{(k+1)}(h_{n,t}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)} \|^2] \\
&+ 2(1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\| - \alpha\eta^2 \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} + \alpha\eta^2 \frac{1}{N}\sum_{m=1}^{N} \nabla F_m^{(k+1)}(h_{m,t-1}^{(k)}; \xi_{m,t}^{(k+1)})^T v_{m,t-1}^{(k+2)} \|^2] \\
\leq\ & (1+\frac{1}{p})(1-\alpha\eta^2)^2 \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N} v_{m,t-1}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] \\
&+ 2(1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} - \nabla F_n^{(k+1)}(h_{n,t}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&+ 2(1+p)\alpha^2\eta^4 \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\| - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} + \frac{1}{N}\sum_{m=1}^{N} \nabla F_m^{(k+1)}(h_{m,t-1}^{(k)}; \xi_{m,t}^{(k+1)})^T v_{m,t-1}^{(k+2)}\|^2] ,
\end{aligned} \tag{69}
$$

where the second step holds since $\Pi_{C^{K-k}}(\cdot)$ in Algorithm 1 is a linear operator.

When $k \in \{1, \cdots, K-1\}$, the second term can be bounded as follows:

$$
\begin{aligned}
&\mathbb{E}[\|\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} - \nabla F_n^{(k+1)}(h_{n,t}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&\leq 2\mathbb{E}[\|\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)})\|^2] \\
&\quad + 2\mathbb{E}[\|\nabla F_n^{(k+1)}(h_{n,t-1}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)} - \nabla F_n^{(k+1)}(h_{n,t}^{(k)}; \xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&\leq 2C^2\mathbb{E}[\|v_{n,t-1}^{(k+2)} - v_{n,t}^{(k+2)}\|^2] + 2C^{2(K-k-1)}L^2\mathbb{E}[\|h_{n,t-1}^{(k)} - h_{n,t}^{(k)}\|^2] ,
\end{aligned} \tag{70}
$$

where the last inequality holds due to Assumption 3.1.

By combining above two inequalities, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \\
&\leq (1+\frac{1}{p})(1-\alpha\eta^2)^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N} v_{m,t-1}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] + 8(1+p)\alpha^2\eta^4 C^{2(K-k)} \\
&\quad + 4(1+p)C^2\frac{1}{N}\sum_{n=1}^{N}\|v_{n,t-1}^{(k+2)} - v_{n,t}^{(k+2)}\|^2] + 4(1+p)C^{2(K-k-1)}L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k)} - h_{n,t}^{(k)}\|^2] \\
&\leq (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_{t-1}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] + 8pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(k+2)} - v_{n,t}^{(k+2)}\|^2] \\
&\quad + 8pC^{2(K-k-1)}L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k)} - h_{n,t}^{(k)}\|^2] + 16p\alpha^2\eta^4C^{2(K-k)} ,
\end{aligned} \tag{71}
$$

where the last step holds due to $1+p \leq 2p$.

By summing over k from 1 to $K-2$, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}\mathbb{E}[\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \\
&\leq (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}\mathbb{E}[\|\bar{v}_{t-1}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] + \sum_{k=1}^{K-2} C^{2k}16p\alpha^2\eta^4C^{2(K-k)} \\
&\quad + 8pC^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}\mathbb{E}[\|v_{n,t-1}^{(k+2)} - v_{n,t}^{(k+2)}\|^2] + 8pL^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}C^{2(K-k-1)}\mathbb{E}[\|h_{n,t-1}^{(k)} - h_{n,t}^{(k)}\|^2]
\end{aligned}
$$

$$\leq 8p \sum_{t'=s_t p}^{t-1} (1+\frac{1}{p})^{t-1-t'} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2(k+1)} \mathbb{E}[\|v_{n,t'}^{(k+2)} - v_{n,t'+1}^{(k+2)}\|^2]$$
$$+ 8pL^2 C^{2(K-1)} \sum_{t'=s_t p}^{t-1} (1+\frac{1}{p})^{t-1-t'} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} \mathbb{E}[\|h_{n,t'}^{(k)} - h_{n,t'+1}^{(k)}\|^2]$$
$$+ 16p\alpha^2\eta^4 C^{2K} K \sum_{t'=s_t p}^{t-1} (1+\frac{1}{p})^{t-1-t'}$$
$$\leq 24p \sum_{t'=s_t p}^{t-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2(k+1)} \mathbb{E}[\|v_{n,t'}^{(k+2)} - v_{n,t'+1}^{(k+2)}\|^2]$$
$$+ 24pL^2 C^{2(K-1)} \sum_{t'=s_t p}^{t-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} \mathbb{E}[\|h_{n,t'}^{(k)} - h_{n,t'+1}^{(k)}\|^2] + 48p^2\alpha^2\eta^4 C^{2K} K \ , \tag{72}$$

where the second step holds due to $v_{n,s_t p} = \bar{v}_{s_t p}$ and $s_t = \lfloor t/p \rfloor$, the last step holds due to $(1+\frac{1}{p})^p < 3$.

By summing over t from 0 to $T-1$, one can get

$$\sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2k} \mathbb{E}[\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \leq 24p \sum_{t=0}^{T-1} \sum_{t'=s_t p}^{t-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2(k+1)} \mathbb{E}[\|v_{n,t'}^{(k+2)} - v_{n,t'+1}^{(k+2)}\|^2]$$
$$+ 24pL^2 C^{2(K-1)} \sum_{t=0}^{T-1} \sum_{t'=s_t p}^{t-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} \mathbb{E}[\|h_{n,t'}^{(k)} - h_{n,t'+1}^{(k)}\|^2] + 48p^2\alpha^2\eta^4 C^{2K} KT$$
$$\leq 24p^2 \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2(k+1)} \mathbb{E}[\|v_{n,t}^{(k+2)} - v_{n,t+1}^{(k+2)}\|^2] + 24p^2 L^2 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} \mathbb{E}[\|h_{n,t}^{(k)} - h_{n,t+1}^{(k)}\|^2]$$
$$+ 48p^2\alpha^2\eta^4 C^{2K} KT$$
$$\leq 24p^2 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t+1}^{(K)}\|^2] + 24p^2 \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2k} \mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t+1}^{(k+1)}\|^2]$$
$$+ 24p^2 L^2 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=2}^{K-1} \mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t+1}^{(k-1)}\|^2] + 48p^2\alpha^2\eta^4 C^{2K} KT$$
$$\leq 24p^2 \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-2} C^{2k} \mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t+1}^{(k+1)}\|^2] + 24p^2 L^2 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=2}^{K-1} \mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t+1}^{(k-1)}\|^2]$$
$$+ 2\alpha^2\eta^4 24p^2 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2]$$
$$+ 96\alpha^2 p^2\eta^4 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-1} (2C^2)^{K-1-k} \mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2]$$
$$+ 48\gamma^2 p^2\eta^2 (2C^2)^{K-1} C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2]$$
$$+ 48\gamma^2 p^2\eta^2 (2C^2)^{K-1} C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[\|\bar{v}_t^{(1)}\|^2]$$
$$+ 96\alpha^2 p^2\eta^4\sigma^2 C^{2(K-1)} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-1} (2C^2)^{K-1-k} + 48\alpha^2 p^2\eta^4\sigma^2 C^{2(K-1)} T + 48p^2\alpha^2\eta^4 C^{2K} KT \ , \tag{73}$$

where the last two steps hold due to Lemma C.17.

□

Lemma C.10. *Suppose Assumptions 3.1-3.3 hold and* $\alpha \leq \frac{C}{N}$, $\eta \leq \frac{1}{20pC}$, *one can get*

$$\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2] \leq 12p^2C^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t+1}^{(k-1)}\|^2]$$
$$+ (24p^2\gamma^2\eta^2C^2 + 120\gamma^2\alpha^2p^4\eta^6C^2)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] + 24p^2\gamma^2\eta^2C^2\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2]$$
$$+ 120p^2\alpha^2\eta^4\sigma^2KT + 60p^2\alpha^2\eta^4\delta^2KT \ . \tag{74}$$

Proof. Similar to Eq. (69), one can get

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2] \leq (1+\frac{1}{p})(1-\alpha\eta^2)^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}h_{m,t-1}^{(k)} - h_{n,t-1}^{(k)}\|^2]$$
$$+ 2(1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) - F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})\|^2]$$
$$+ 2\alpha^2\eta^4(1+p)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\| - F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) + \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)};\xi_{m,t}^{(k)})\|^2] \ . \tag{75}$$

Then, the last term can be bounded as follows:

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) - \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)};\xi_{m,t}^{(k)})\|^2]$$
$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) - F_n^{(k)}(h_{n,t-1}^{(k-1)}) + F_n^{(k)}(h_{n,t-1}^{(k-1)}) - F_n^{(k)}(\bar{h}_{t-1}^{(k-1)}) + F_n^{(k)}(\bar{h}_{t-1}^{(k-1)}) - F^{(k)}(\bar{h}_{t-1}^{(k-1)})$$
$$+ F^{(k)}(\bar{h}_{t-1}^{(k-1)}) - \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)}) + \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)}) - \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)};\xi_{m,t}^{(k)})\|^2]$$
$$\leq 5\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 5\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(h_{n,t-1}^{(k-1)}) - F_n^{(k)}(\bar{h}_{t-1}^{(k-1)})\|^2]$$
$$+ 5\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(\bar{h}_{t-1}^{(k-1)}) - F^{(k)}(\bar{h}_{t-1}^{(k-1)})\|^2] + 5\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(\bar{h}_{t-1}^{(k-1)}) - \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)})\|^2]$$
$$+ 5\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)}) - \frac{1}{N}\sum_{m=1}^{N}F_m^{(k)}(h_{m,t-1}^{(k-1)};\xi_{m,t}^{(k)})\|^2]$$
$$\leq 10\sigma^2 + 5\delta^2 + 10C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k-1)} - \bar{h}_{t-1}^{(k-1)}\|^2] \ , \tag{76}$$

where the last step holds due to Assumption 3.3. Then, by combining above two inequalities, one can get

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2]$$
$$\leq (1+\frac{1}{p})(1-\alpha\eta^2)^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\frac{1}{N}\sum_{m=1}^{N}h_{m,t-1}^{(k)} - h_{n,t-1}^{(k)}\|^2] + 2(1+p)C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2]$$
$$+ 20(1+p)\alpha^2\eta^4\sigma^2 + 10(1+p)\alpha^2\eta^4\delta^2 + 20(1+p)\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k-1)} - \bar{h}_{t-1}^{(k-1)}\|^2]$$

$$
\begin{aligned}
&\leq (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{h}_{t-1}^{(k)}-h_{n,t-1}^{(k)}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-h_{n,t}^{(k-1)}\|^2] \\
&\quad +40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-\bar{h}_{t-1}^{(k-1)}\|^2]+40p\alpha^2\eta^4\sigma^2+20p\alpha^2\eta^4\delta^2 \ . \qquad (77)
\end{aligned}
$$

By summing over k from 1 to $K-1$, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t}^{(k)}-h_{n,t}^{(k)}\|^2] \\
&\leq (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t-1}^{(k)}-h_{n,t-1}^{(k)}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-h_{n,t}^{(k-1)}\|^2] \\
&\quad +40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-\bar{h}_{t-1}^{(k-1)}\|^2]+40p\alpha^2\eta^4\sigma^2K+20p\alpha^2\eta^4\delta^2K \\
&\leq (1+\frac{1}{p})\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t-1}^{(k)}-h_{n,t-1}^{(k)}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(0)}-h_{n,t}^{(0)}\|^2] \\
&\quad +4pC^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-h_{n,t}^{(k-1)}\|^2]+40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(0)}-\bar{h}_{t-1}^{(0)}\|^2] \\
&\quad +40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k)}-\bar{h}_{t-1}^{(k)}\|^2]+40p\alpha^2\eta^4\sigma^2K+20p\alpha^2\eta^4\delta^2K \\
&\leq (1+\frac{1}{p}+40p\alpha^2\eta^4C^2)\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t-1}^{(k)}-h_{n,t-1}^{(k)}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-h_{n,t}^{(k-1)}\|^2] \\
&\quad +40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t-1}-\bar{x}_{t-1}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t-1}-x_{n,t}\|^2] \\
&\quad +40p\alpha^2\eta^4\sigma^2K+20p\alpha^2\eta^4\delta^2K \ . \qquad (78)
\end{aligned}
$$

By setting $\alpha \leq \frac{C}{N}$ and $\eta \leq \frac{1}{20pC}$, we have

$$
1+\frac{1}{p}+40p\alpha^2\eta^4C^2 \leq 1+\frac{1}{p}+40pC^2\frac{C^2}{N^2}\frac{1}{1600p^4C^4}=1+\frac{1}{p}+\frac{1}{40N^2p^3}\leq 1+\frac{41}{40p} \ . \qquad (79)
$$

Then, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t}^{(k)}-h_{n,t}^{(k)}\|^2] \\
&\leq (1+\frac{41}{40p})\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t-1}^{(k)}-h_{n,t-1}^{(k)}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-h_{n,t}^{(k-1)}\|^2] \\
&\quad +40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t-1}-\bar{x}_{t-1}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t-1}-x_{n,t}\|^2] \\
&\quad +40p\alpha^2\eta^4\sigma^2K+20p\alpha^2\eta^4\delta^2K \\
&\leq (1+\frac{41}{40p})\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_{t-1}^{(k)}-h_{n,t-1}^{(k)}\|^2]+4pC^2\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t-1}^{(k-1)}-h_{n,t}^{(k-1)}\|^2]
\end{aligned}
$$

$$+ 8p\gamma^2\eta^2C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 8p\gamma^2\eta^2C^2\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2]$$
$$+ 40p\alpha^2\eta^4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t-1} - \bar{x}_{t-1}\|^2] + 40p\alpha^2\eta^4\sigma^2K + 20p\alpha^2\eta^4\delta^2K \;, \tag{80}$$

where the last step holds due to Lemma C.15.

Then, by denoting $s_t = \lfloor t/p \rfloor$, one can get

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2] \leq 4pC^2\sum_{t'=s_tp}^{t-1}(1+\frac{41}{40p})^{t-1-t'}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t'}^{(k-1)} - h_{n,t'+1}^{(k-1)}\|^2]$$
$$+ 8p\gamma^2\eta^2C^2\sum_{t'=s_tp}^{t-1}(1+\frac{41}{40p})^{t-1-t'}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t'}^{(1)} - \bar{v}_{t'}^{(1)}\|^2] + 8p\gamma^2\eta^2C^2\sum_{t'=s_tp}^{t-1}(1+\frac{41}{40p})^{t-1-t'}\mathbb{E}[\|\bar{v}_{t'}^{(1)}\|^2]$$
$$+ 40p\alpha^2\eta^4C^2\sum_{t'=s_tp}^{t-1}(1+\frac{41}{40p})^{t-1-t'}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t'} - \bar{x}_{t'}\|^2]$$
$$+ (40p\alpha^2\eta^4\sigma^2K + 20p\alpha^2\eta^4\delta^2K)\sum_{t'=s_tp}^{t-1}(1+\frac{41}{40p})^{t-1-t'}$$
$$\leq 12pC^2\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t'}^{(k-1)} - h_{n,t'+1}^{(k-1)}\|^2] + 24p\gamma^2\eta^2C^2\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t'}^{(1)} - \bar{v}_{t'}^{(1)}\|^2]$$
$$+ 24p\gamma^2\eta^2C^2\sum_{t'=s_tp}^{t-1}\mathbb{E}[\|\bar{v}_{t'}^{(1)}\|^2] + 120p\alpha^2\eta^4C^2\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t'} - \bar{x}_{t'}\|^2]$$
$$+ 3p(40p\alpha^2\eta^4\sigma^2K + 20p\alpha^2\eta^4\delta^2K) \;, \tag{81}$$

where the second step holds due to $(1+\frac{41}{40p})^p \leq e^{41/40} \leq 3$.

By summing over t from 0 to $T-1$, one can get

$$\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2] \leq 12pC^2\sum_{t=0}^{T-1}\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t'}^{(k-1)} - h_{n,t'+1}^{(k-1)}\|^2]$$
$$+ 24p\gamma^2\eta^2C^2\sum_{t=0}^{T-1}\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t'}^{(1)} - \bar{v}_{t'}^{(1)}\|^2] + 24p\gamma^2\eta^2C^2\sum_{t=0}^{T-1}\sum_{t'=s_tp}^{t-1}\mathbb{E}[\|\bar{v}_{t'}^{(1)}\|^2]$$
$$+ 120p\alpha^2\eta^4C^2\sum_{t=0}^{T-1}\sum_{t'=s_tp}^{t-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t'} - \bar{x}_{t'}\|^2] + 3pT(40p\alpha^2\eta^4\sigma^2K + 20p\alpha^2\eta^4\delta^2K)$$
$$\leq 12p^2C^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t+1}^{(k-1)}\|^2] + 24p^2\gamma^2\eta^2C^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2]$$
$$+ 24p^2\gamma^2\eta^2C^2\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] + 120p^2\alpha^2\eta^4C^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|x_{n,t} - \bar{x}_t\|^2] + 3pT(40p\alpha^2\eta^4\sigma^2K + 20p\alpha^2\eta^4\delta^2K)$$
$$\leq 12p^2C^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t+1}^{(k-1)}\|^2] + 120p^2\alpha^2\eta^4\sigma^2KT + 60p^2\alpha^2\eta^4\delta^2KT$$
$$+ (24p^2\gamma^2\eta^2C^2 + 120\gamma^2\alpha^2p^4\eta^6C^2)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] + 24p^2\gamma^2\eta^2C^2\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \;, \tag{82}$$

where the last step holds due to Lemma C.7.

□

Lemma C.11. *Suppose Assumptions 3.1-3.3 hold, then for $k \in \{1, \cdots, K-1\}$, one can get*

$$\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2]$$
$$+ 4C^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] + 4C^{2(K-k)}L^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] + 2\alpha^2\eta^4 C^{2(K-k)}\frac{\sigma^2}{N} \,. \quad (83)$$

For $k = K$, one can get

$$\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(K)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2]$$
$$+ 2L^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(K-1)} - h_{n,t-1}^{(K-1)}\|^2] + 2\alpha^2\eta^4\frac{\sigma^2}{N} \,. \quad (84)$$

Proof. When $k \in \{1, \cdots, K-1\}$, one can get

$$\begin{aligned}
&\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&= \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} \Pi_{C^{K-k+1}}((1-\alpha\eta^2)(v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)}) + \nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)}) \\
&\quad - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\leq \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)}) + \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} \\
&\quad - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\leq \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}) \\
&\quad + \frac{1}{N}\sum_{n=1}^{N}\Big((1-\alpha\eta^2)(\nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)} + \nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} \\
&\quad - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}) + \alpha\eta^2(\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)})\Big)\|^2] \\
&\leq (1-\alpha\eta^2)^2\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&\quad + 2(1-\alpha\eta^2)^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)} \\
&\qquad + \nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\quad + 2\alpha^2\eta^4\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2]
\end{aligned}$$

$$
\begin{aligned}
&+ 2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)}\|^2] + 2\alpha^2\eta^4 C^{2(K-k)}\frac{\sigma^2}{N} \\
&\leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] + 2\alpha^2\eta^4 C^{2(K-k)}\frac{\sigma^2}{N} \\
&+ 4\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&+ 4\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&\leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&+ 4C^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] + 4C^{2(K-k)}L^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] + 2\alpha^2\eta^4 C^{2(K-k)}\frac{\sigma^2}{N} \ , \qquad (85)
\end{aligned}
$$

where the second step holds because $\Pi_{C^{K-k+1}}(\cdot)$ is a linear operator.

When $k = K$, one can get

$$
\begin{aligned}
&\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&= \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}\Pi_C((1-\alpha\eta^2)(v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\leq \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)}) - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&= \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})) \\
&+ \frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(\nabla F_n^{(K)}(h_{n,t-1}^{(K-1)}) - \nabla F_n^{(K)}(h_{n,t}^{(K-1)}) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)}) - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})) \\
&+ \alpha\eta^2\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)}) - \alpha\eta^2\frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\leq (1-\alpha\eta^2)^2\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(K)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] \\
&+ 2(1-\alpha\eta^2)^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(K)}(h_{n,t-1}^{(K-1)}) - \nabla F_n^{(K)}(h_{n,t}^{(K-1)}) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)}) - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&+ 2\alpha^2\eta^4\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|\nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)}) - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t-1}^{(K)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] + 2L^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(K-1)} - h_{n,t-1}^{(K-1)}\|^2] + 2\alpha^2\eta^4\frac{\sigma^2}{N} \ . \qquad (86)
\end{aligned}
$$

□

Lemma C.12. *Suppose Assumptions 3.1-3.3 hold, then for $k \in \{1, \cdots, K-1\}$, one can get*

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\leq (1-\alpha\eta^2)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&\quad + 4C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] + 4C^{2(K-k)}L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] + 2\alpha^2\eta^4C^{2(K-k)}\sigma^2 \,. \qquad (87)
\end{aligned}
$$

For $k = K$, one can get

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&\leq (1-\alpha\eta^2)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] + 2L^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(K-1)} - h_{n,t-1}^{(K-1)}\|^2] + 2\alpha^2\eta^4\sigma^2 \,. \qquad (88)
\end{aligned}
$$

It can be proved by following Lemma C.11.

Lemma C.13. *Suppose Assumptions 3.1-3.3 hold, then for $k \in \{1, \cdots, K-1\}$, one can get*

$$
\begin{aligned}
&\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 2C^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] + 2\alpha^2\eta^4\frac{\sigma^2}{N} \,. \qquad (89)
\end{aligned}
$$

Proof. For $k \in \{1, \cdots, K-1\}$, one can get

$$
\begin{aligned}
&\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&= \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)})) + \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)}) - \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&= \mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})) \\
&\quad + \frac{1}{N}\sum_{n=1}^{N}(1-\alpha\eta^2)(F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)}) - F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) + F_n^{(k)}(h_{n,t-1}^{(k-1)}) - F_n^{(k)}(h_{n,t}^{(k-1)})) \\
&\quad + \alpha\eta^2\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)}) - \alpha\eta^2\frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\leq (1-\alpha\eta^2)^2\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] \\
&\quad + 2(1-\alpha\eta^2)^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|(F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)}) - F_n^{(k)}(h_{n,t-1}^{(k-1)};\xi_{n,t}^{(k)}) + F_n^{(k)}(h_{n,t-1}^{(k-1)}) - F_n^{(k)}(h_{n,t}^{(k-1)}))\|^2] \\
&\quad + 2\alpha^2\eta^4\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|F_n^{(k)}(h_{n,t}^{(k-1)};\xi_{n,t}^{(k)}) - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2]
\end{aligned}
$$

$$\leq (1-\alpha\eta^2)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N}h_{n,t-1}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 2C^2\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] + 2\alpha^2\eta^4\frac{\sigma^2}{N} \ . \quad (90)$$

□

Lemma C.14. *Suppose Assumptions 3.1-3.3 hold, then for $k \in \{1, \cdots, K-1\}$, one can get*

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\leq (1-\alpha\eta^2)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 2C^2\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] + 2\alpha^2\eta^4\sigma^2 \ . \quad (91)
\end{aligned}$$

It can be proved by following Lemma C.13.

Lemma C.15. *Suppose Assumptions 3.1-3.3 hold, one can get*

$$\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] \leq \gamma^2\eta^2p^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \ . \quad (92)$$

Proof.

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_{s_tp} - \gamma\eta\sum_{t'=s_tp}^{t-1}\bar{v}_{t'}^{(1)} - x_{n,s_tp} + \gamma\eta\sum_{t'=s_tp}^{t-1}v_{n,t'}^{(1)}\|^2] \\
&= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\| - \gamma\eta\sum_{t'=s_tp}^{t-1}\bar{v}_{t'}^{(1)} + \gamma\eta\sum_{t'=s_tp}^{t-1}v_{n,t'}^{(1)}\|^2] \leq \gamma^2\eta^2p\frac{1}{N}\sum_{n=1}^{N}\sum_{t'=s_tp}^{t-1}\mathbb{E}[\|v_{n,t'}^{(1)} - \bar{v}_{t'}^{(1)}\|^2] \ , \quad (93)
\end{aligned}$$

where the last step holds due to $x_{n,s_tp} = \bar{x}_{s_tp}$ when $s_t = \lfloor t/p \rfloor$. By summing over t from 0 to $T-1$, one can get

$$\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] \leq \gamma^2\eta^2p\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{t'=s_tp}^{t-1}\mathbb{E}[\|v_{n,t'}^{(1)} - \bar{v}_{t'}^{(1)}\|^2] \leq \gamma^2\eta^2p^2\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(1)} - \bar{v}_t^{(1)}\|^2] \ . \quad (94)$$

□

Lemma C.16. *Suppose Assumptions 3.1-3.3 hold, then for any $\omega_k > 0$ when $k \in \{2, \cdots, K\}$, by setting $\eta \leq \frac{\omega_k^{1/2}}{96\gamma p(3C)^{K-1}L\sqrt{\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}}}$, one can get*

$$\begin{aligned}
&\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\omega_k\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \\
&\leq 4\gamma^2\eta^2\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 192\alpha^2\gamma^2p^2\eta^6\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}(3C^2)^{k-1}\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^Tv_{n,t-1}^{(k+1)}\|^2] \\
&\quad + 128\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] \\
&\quad + \sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Bigg(256\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)(2C^2)^{K-1-k}
\end{aligned}$$

$$
\begin{aligned}
&+ 4\alpha^2\eta^4 \sum_{j=k}^{K-1} \omega_{j+1}(2C^2)^{j-k}\Big)\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] \\
&+ 256\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} \\
&+ 128\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big) + 96\alpha^2\gamma^2p^2\eta^6C^{2K}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big) \\
&+ 192\alpha^2\gamma^2p^2\eta^6\sigma^2C^{2(K-1)}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{k=1}^{K-2}3^k + 4\alpha^2\eta^4\sigma^2T\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j} ,
\end{aligned} \tag{95}
$$

and

$$
\begin{aligned}
\mathbb{E}[\|h_{n,t}^{(K-1)} - h_{n,t-1}^{(K-1)}\|^2] &\leq 2\gamma^2\eta^2(2C^2)^{K-1}\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \\
&+ 2\alpha^2\eta^4\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 2\alpha^2\eta^4\sigma^2\sum_{j=1}^{K-1}(2C^2)^{K-1-j} .
\end{aligned} \tag{96}
$$

Proof. When $k \in \{2, \cdots, K\}$, one can get

$$
\begin{aligned}
&\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \\
&= \mathbb{E}[\|(1-\alpha\eta^2)(h_{n,t-1}^{(k-1)} - F_n^{(k-1)}(h_{n,t-1}^{(k-2)};\xi_{n,t}^{(k-1)})) + F_n^{(k-1)}(h_{n,t}^{(k-2)};\xi_{n,t}^{(k-1)}) - h_{n,t-1}^{(k-1)}\|^2] \\
&= \mathbb{E}[\|F_n^{(k-1)}(h_{n,t}^{(k-2)};\xi_{n,t}^{(k-1)}) - F_n^{(k-1)}(h_{n,t-1}^{(k-2)};\xi_{n,t}^{(k-1)}) - \alpha\eta^2(h_{n,t-1}^{(k-1)} - F_n^{(k-1)}(h_{n,t-1}^{(k-2)};\xi_{n,t}^{(k-1)}))\|^2] \\
&\leq 2\mathbb{E}[\|F_n^{(k-1)}(h_{n,t}^{(k-2)};\xi_{n,t}^{(k-1)}) - F_n^{(k-1)}(h_{n,t-1}^{(k-2)};\xi_{n,t}^{(k-1)})\|^2] + 2\alpha^2\eta^4\mathbb{E}[\|h_{n,t-1}^{(k-1)} - F_n^{(k-1)}(h_{n,t-1}^{(k-2)};\xi_{n,t}^{(k-1)})\|^2] \\
&\leq 2C^2\mathbb{E}[\|h_{n,t}^{(k-2)} - h_{n,t-1}^{(k-2)}\|^2] + 2\alpha^2\eta^4\mathbb{E}[\|h_{n,t-1}^{(k-1)} - F_n^{(k-1)}(h_{n,t-1}^{(k-2)})\|^2] + 2\alpha^2\eta^4\sigma^2 ,
\end{aligned} \tag{97}
$$

where the last step holds due to Assumption 3.3. Then, it is easy to know

$$
\begin{aligned}
&\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \\
&\leq (2C^2)^{k-1}\mathbb{E}[\|x_{n,t} - x_{n,t-1}\|^2] + 2\alpha^2\eta^4\sum_{j=1}^{k-1}(2C^2)^{k-1-j}\mathbb{E}[\|h_{n,t-1}^{(j)} - F_n^{(j)}(h_{n,t-1}^{(j-1)})\|^2] + 2\alpha^2\eta^4\sigma^2\sum_{j=1}^{k-1}(2C^2)^{k-1-j} \\
&= \gamma^2\eta^2(2C^2)^{k-1}\mathbb{E}[\|v_{n,t-1}^{(1)}\|^2] + 2\alpha^2\eta^4\sum_{j=1}^{k-1}(2C^2)^{k-1-j}\mathbb{E}[\|h_{n,t-1}^{(j)} - F_n^{(j)}(h_{n,t-1}^{(j-1)})\|^2] + 2\alpha^2\eta^4\sigma^2\sum_{j=1}^{k-1}(2C^2)^{k-1-j} .
\end{aligned} \tag{98}
$$

For any $\omega_k > 0$ when $k \in \{2, \cdots, K\}$, by summing over k from 2 to $K-1$, one can get

$$
\begin{aligned}
&\sum_{k=2}^{K-1}\omega_k\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \leq \gamma^2\eta^2\Big(\sum_{k=2}^{K-1}\omega_k(2C^2)^{k-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(1)}\|^2] \\
&+ 2\alpha^2\eta^4\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j}\mathbb{E}[\|h_{n,t-1}^{(j)} - F_n^{(j)}(h_{n,t-1}^{(j-1)})\|^2] + 2\alpha^2\eta^4\sigma^2\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j} \\
&\leq \gamma^2\eta^2\Big(\sum_{k=2}^{K-1}\omega_k(2C^2)^{k-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(1)}\|^2] + 2\alpha^2\eta^4\sum_{k=1}^{K-2}\Big(\sum_{j=k}^{K-2}\omega_{j+1}(2C^2)^{j-k}\Big)\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] \\
&\quad + 2\alpha^2\eta^4\sigma^2\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j}
\end{aligned}
$$

$$\leq \gamma^2\eta^2\Big(\sum_{k=2}^{K-1}\omega_k(2C^2)^{k-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(1)}\|^2] + 2\alpha^2\eta^4\sum_{k=1}^{K-1}\Big(\sum_{j=k}^{K-2}\omega_{j+1}(2C^2)^{j-k}\Big)\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2]$$
$$+ 2\alpha^2\eta^4\sigma^2\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j}$$
$$\leq 2\gamma^2\eta^2\Big(\sum_{k=2}^{K-1}\omega_k(2C^2)^{k-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 2\gamma^2\eta^2\Big(\sum_{k=2}^{K-1}\omega_k(2C^2)^{k-1}\Big)\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2]$$
$$+ 2\alpha^2\eta^4\sum_{k=1}^{K-1}\Big(\sum_{j=k}^{K-2}\omega_{j+1}(2C^2)^{j-k}\Big)\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 2\alpha^2\eta^4\sigma^2\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j} . \tag{99}$$

Then, based on Lemma C.7, one can get

$$\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\omega_k\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2]$$
$$\leq 2\gamma^2\eta^2\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2]$$
$$+ 2\alpha^2\eta^4\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big(\sum_{j=k}^{K-2}\omega_{j+1}(2C^2)^{j-k}\Big)\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 2\alpha^2\eta^4\sigma^2T\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j}$$
$$+ 96\gamma^2\eta^2p^2(3C^2)^{K-1}L^2\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2]$$
$$+ 96\alpha^2\gamma^2p^2\eta^6\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}(3C^2)^{k-1}\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2]$$
$$+ 64\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2]$$
$$+ 128\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2]$$
$$+ 128\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k}$$
$$+ 64\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big) + 48\alpha^2\gamma^2p^2\eta^6C^{2K}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)$$
$$+ 96\alpha^2\gamma^2p^2\eta^6\sigma^2C^{2(K-1)}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{k=1}^{K-2}3^k . \tag{100}$$

By setting $\eta \leq \frac{\omega_k^{1/2}}{96\gamma p(3C)^{K-1}L\sqrt{\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}}}$ such that $\omega_k - 96\gamma^2\eta^2p^2(3C^2)^{K-1}L^2\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big) \geq \frac{1}{2}\omega_k$, one can get

$$\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\omega_k\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2]$$
$$\leq 4\gamma^2\eta^2\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\sum_{t=0}^{T-1}\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2]$$

$$
\begin{aligned}
&+ 192\alpha^2\gamma^2p^2\eta^6\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}(3C^2)^{k-1}\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&+ 128\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] \\
&+ \sum_{t=0}^{T-1}\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Bigg(256\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)(2C^2)^{K-1-k} \\
&\quad + 4\alpha^2\eta^4\sum_{j=k}^{K-2}\omega_{j+1}(2C^2)^{j-k}\Bigg)\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] \\
&+ 256\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} \\
&+ 128\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big) + 96\alpha^2\gamma^2p^2\eta^6C^{2K}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big) \\
&+ 192\alpha^2\gamma^2p^2\eta^6\sigma^2C^{2(K-1)}T\Big(\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}\Big)\sum_{k=1}^{K-2}3^k + 4\alpha^2\eta^4\sigma^2T\sum_{k=2}^{K-1}\omega_k\sum_{j=1}^{k-1}(2C^2)^{k-1-j}\,. \qquad (101)
\end{aligned}
$$

□

Lemma C.17. *Suppose Assumptions 3.1-3.3 hold, then for any $\omega_k > 0$ when $k \in \{1, \cdots, K-2\}$, one can get*

$$
\begin{aligned}
&\sum_{k=1}^{K-2}\omega_k\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] \le L^2\sum_{k=2}^{K-1}\Big(\sum_{j=1}^{k-1}\omega_jC^{2(K-j-1)}3^{k-j}\Big)\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \\
&+ 3\alpha^2\eta^4\sum_{k=2}^{K-1}\Big(\sum_{j=1}^{k-1}\omega_j(3C^2)^{k-1-j}\Big)\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&+ 4\alpha^2\eta^4\sum_{k=1}^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] \\
&+ 2\alpha^2\eta^4\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] \\
&+ 2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \\
&+ 4\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 2\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big) \\
&+ \alpha^2\eta^4\sigma^2\sum_{k=1}^{K-2}\omega_kC^{2(K-k-1)}\sum_{i=k}^{K-2}3^{i-k+1}\,, \qquad (102)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t-1}^{(K)}\|^2] \\
&\le 2\alpha^2\eta^4\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 4\alpha^2\eta^4\sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 4\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 2\alpha^2\eta^4\sigma^2\,. \qquad (103)
\end{aligned}
$$

Proof. When $k \in \{1, \cdots, K-2\}$, one can get

$$
\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2]
$$

$$
\begin{aligned}
&= \mathbb{E}[\|\Pi_{C^{K-k}}((1-\alpha\eta^2)(v_{n,t-1}^{(k+1)} - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)}) + \nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}) - v_{n,t-1}^{(k+1)}\|^2] \\
&\leq \mathbb{E}[\|(1-\alpha\eta^2)(v_{n,t-1}^{(k+1)} - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)}) + \nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)} - v_{n,t-1}^{(k+1)}\|^2] \\
&= \mathbb{E}[\| - \alpha\eta^2 v_{n,t-1}^{(k+1)} + \alpha\eta^2 \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} \\
&\quad - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} + \nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&= \mathbb{E}[\| - \alpha\eta^2 v_{n,t-1}^{(k+1)} + \alpha\eta^2 \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} \\
&\quad - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} + \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)} \\
&\quad - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)} + \nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&\leq 3\alpha^2\eta^4 \mathbb{E}[\| - v_{n,t-1}^{(k+1)} + \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)})^T v_{n,t-1}^{(k+2)} + \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)})^T v_{n,t-1}^{(k+2)}\|^2] \\
&\quad + 3\mathbb{E}[\| - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t-1}^{(k+2)} + \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&\quad + 3\mathbb{E}[\| - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)} + \nabla F_n^{(k+1)}(h_{n,t}^{(k)};\xi_{n,t}^{(k+1)})^T v_{n,t}^{(k+2)}\|^2] \\
&\leq 3\alpha^2\eta^4 \mathbb{E}[\|v_{n,t-1}^{(k+1)} - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)})^T v_{n,t-1}^{(k+2)}\|^2] + 3\alpha^2\eta^4 C^{2(K-k-1)}\sigma^2 \\
&\quad + 3C^2 \mathbb{E}[\|v_{n,t}^{(k+2)} - v_{n,t-1}^{(k+2)}\|^2] + 3C^{2(K-k-1)} L^2 \mathbb{E}[\|h_{n,t}^{(k)} - h_{n,t-1}^{(k)}\|^2] \ , \qquad (104)
\end{aligned}
$$

where the last step holds due to Assumption 3.1-3.3.

Then, it is easy to know

$$
\begin{aligned}
\mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] &\leq (3C^2)^{K-k-1}\mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t-1}^{(K)}\|^2] + C^{2(K-k-1)}L^2 \sum_{i=k}^{K-2} 3^{i-k+1}\mathbb{E}[\|h_{n,t}^{(i)} - h_{n,t-1}^{(i)}\|^2] \\
&\quad + 3\alpha^2\eta^4 \sum_{i=k}^{K-2} (3C^2)^{i-k}\mathbb{E}[\|v_{n,t-1}^{(i+1)} - \nabla F_n^{(i+1)}(h_{n,t-1}^{(i)})^T v_{n,t-1}^{(i+2)}\|^2] + \alpha^2\eta^4 C^{2(K-k-1)}\sigma^2 \sum_{i=k}^{K-2} 3^{i-k+1} \ . \qquad (105)
\end{aligned}
$$

Given any $\omega_k > 0$, by summing over k from 1 to $K-2$, one can get

$$
\begin{aligned}
&\sum_{k=1}^{K-2} \omega_k \mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] \leq \Big(\sum_{k=1}^{K-2} \omega_k (3C^2)^{K-k-1}\Big)\mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t-1}^{(K)}\|^2] \\
&\quad + L^2 \sum_{k=1}^{K-2} \omega_k C^{2(K-k-1)} \sum_{i=k}^{K-2} 3^{i-k+1}\mathbb{E}[\|h_{n,t}^{(i)} - h_{n,t-1}^{(i)}\|^2] \\
&\quad + 3\alpha^2\eta^4 \sum_{k=1}^{K-2} \omega_k \sum_{i=k}^{K-2} (3C^2)^{i-k}\mathbb{E}[\|v_{n,t-1}^{(i+1)} - \nabla F_n^{(i+1)}(h_{n,t-1}^{(i)})^T v_{n,t-1}^{(i+2)}\|^2] + \alpha^2\eta^4\sigma^2 \sum_{k=1}^{K-2} \omega_k C^{2(K-k-1)} \sum_{i=k}^{K-2} 3^{i-k+1} \\
&\leq \Big(\sum_{k=1}^{K-2} \omega_k (3C^2)^{K-k-1}\Big)\mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t-1}^{(K)}\|^2] + L^2 \sum_{k=1}^{K-2} \Big(\sum_{j=1}^{k} \omega_j C^{2(K-j-1)} 3^{k-j+1}\Big)\mathbb{E}[\|h_{n,t}^{(k)} - h_{n,t-1}^{(k)}\|^2] \\
&\quad + 3\alpha^2\eta^4 \sum_{k=1}^{K-2} \Big(\sum_{j=1}^{k} \omega_j (3C^2)^{k-j}\Big)\mathbb{E}[\|v_{n,t-1}^{(k+1)} - \nabla F_n^{(k+1)}(h_{n,t-1}^{(k)})^T v_{n,t-1}^{(k+2)}\|^2] + \alpha^2\eta^4\sigma^2 \sum_{k=1}^{K-2} \omega_k C^{2(K-k-1)} \sum_{i=k}^{K-2} 3^{i-k+1} \\
&\leq \Big(\sum_{k=1}^{K-2} \omega_k (3C^2)^{K-k-1}\Big)\mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t-1}^{(K)}\|^2] + L^2 \sum_{k=2}^{K-1} \Big(\sum_{j=1}^{k-1} \omega_j C^{2(K-j-1)} 3^{k-j}\Big)\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \\
&\quad + 3\alpha^2\eta^4 \sum_{k=2}^{K-1} \Big(\sum_{j=1}^{k-1} \omega_j (3C^2)^{k-1-j}\Big)\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] + \alpha^2\eta^4\sigma^2 \sum_{k=1}^{K-2} \omega_k C^{2(K-k-1)} \sum_{i=k}^{K-2} 3^{i-k+1} \ . \qquad (106)
\end{aligned}
$$

Additionally, one can get

$$
\mathbb{E}[\|v_{n,t}^{(K)} - v_{n,t-1}^{(K)}\|^2]
$$

$$
\begin{aligned}
&= \mathbb{E}[\|\Pi_C((1-\alpha\eta^2)(v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})) - v_{n,t-1}^{(K)}\|^2] \\
&\le \mathbb{E}[\| - \alpha\eta^2 v_{n,t-1}^{(K)} + \alpha\eta^2 \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) \\
&\quad - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&\le 2\alpha^2\eta^4 \mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})\|^2] + 2\mathbb{E}[\| - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&\le 2\alpha^2\eta^4 \mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)}) + \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)}) - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&\quad + 2\mathbb{E}[\| - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)};\xi_{n,t}^{(K)}) + \nabla F_n^{(K)}(h_{n,t}^{(K-1)};\xi_{n,t}^{(K)})\|^2] \\
&\le 2\alpha^2\eta^4 \mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] + 2\mathbb{E}[\|h_{n,t}^{(K-1)} - h_{n,t-1}^{(K-1)}\|^2] + 2\alpha^2\eta^4\sigma^2 \\
&\le 2\alpha^2\eta^4 \mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 4\alpha^2\eta^4 \sum_{k=1}^{K-1}(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] + 4\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 2\alpha^2\eta^4\sigma^2 \, . \qquad (107)
\end{aligned}
$$

By combing above two inequalities together, one can get

$$
\begin{aligned}
&\sum_{k=1}^{K-2}\omega_k \mathbb{E}[\|v_{n,t}^{(k+1)} - v_{n,t-1}^{(k+1)}\|^2] \le L^2 \sum_{k=2}^{K-1}\Big(\sum_{j=1}^{k-1}\omega_j C^{2(K-j-1)}3^{k-j}\Big)\mathbb{E}[\|h_{n,t}^{(k-1)} - h_{n,t-1}^{(k-1)}\|^2] \\
&\quad + 3\alpha^2\eta^4 \sum_{k=2}^{K-1}\Big(\sum_{j=1}^{k-1}\omega_j(3C^2)^{k-1-j}\Big)\mathbb{E}[\|v_{n,t-1}^{(k)} - \nabla F_n^{(k)}(h_{n,t-1}^{(k-1)})^T v_{n,t-1}^{(k+1)}\|^2] \\
&\quad + 4\alpha^2\eta^4 \sum_{k=1}^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)(2C^2)^{K-1-k}\mathbb{E}[\|h_{n,t-1}^{(k)} - F_n^{(k)}(h_{n,t-1}^{(k-1)})\|^2] \\
&\quad + 2\alpha^2\eta^4\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(K)} - \nabla F_n^{(K)}(h_{n,t-1}^{(K-1)})\|^2] \\
&\quad + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\mathbb{E}[\|v_{n,t-1}^{(1)} - \bar{v}_{t-1}^{(1)}\|^2] + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\mathbb{E}[\|\bar{v}_{t-1}^{(1)}\|^2] \\
&\quad + 4\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 2\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2}\omega_j(3C^2)^{K-j-1}\Big) \\
&\quad + \alpha^2\eta^4\sigma^2\sum_{k=1}^{K-2}\omega_k C^{2(K-k-1)}\sum_{i=k}^{K-2}3^{i-k+1} \, . \qquad (108)
\end{aligned}
$$

□

With these lemmas, we start to prove Theorem 4.1.

Proof. We define the following potential function:

$$
\begin{aligned}
P_t &= \mathbb{E}[\Phi(\bar{x}_t)] + \sum_{k=1}^{K-1} a_{1,k}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&\quad + a_{1,K}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] + \sum_{k=1}^{K-1} b_{1,k}\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&\quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1} a_{2,k}\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2]
\end{aligned}
$$

$$+ a_{2,K}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1} b_{2,k}\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \,. \tag{109}$$

Then, with Lemmas C.6, C.11, C.13, we can obtain

$$\begin{aligned}
& P_{t+1} - P_t \\
& \leq -\frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] - \frac{\gamma\eta}{4}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
& \quad + \Big(2\gamma\eta K C^{2(K-1)} - \alpha\eta^2 a_{1,K}\Big)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(K)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
& \quad + \sum_{k=1}^{K-1}\Big(3\gamma\eta K C^{2(k-1)} - \alpha\eta^2 a_{1,k}\Big)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} v_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N}\nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
& \quad + \sum_{k=1}^{K-1}\Big(2\gamma\eta A K - \alpha\eta^2 b_{1,k}\Big)\mathbb{E}[\|\frac{1}{N}\sum_{n=1}^{N} h_{n,t}^{(k)} - \frac{1}{N}\sum_{n=1}^{N} F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big(-\alpha\eta^2 a_{2,k}\Big)\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big(-\alpha\eta^2 b_{2,k}\Big)\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\Big(-\alpha\eta^2 a_{2,K}\Big)\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
& \quad + 3\gamma\eta K C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\|^2] + 3\gamma\eta K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2} C^{2k}\mathbb{E}[\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \\
& \quad + 3\gamma\eta K(L^2C^{2(K-1)} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] \\
& \quad + 3\gamma\eta K(L^2C^{2(K-1)} + C^{2K} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1}\frac{1}{N}\Big)\mathbb{E}[\|v_{n,t+1}^{K} - v_{n,t}^{K}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2}\Big(4C^2 a_{2,k} + 4C^2 a_{1,k}\frac{1}{N}\Big)\mathbb{E}[\|v_{n,t+1}^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2 b_{1,1}\frac{1}{N} + 2C^2 b_{2,1}\Big)\mathbb{E}[\|h_{n,t+1}^{(0)} - h_{n,t}^{(0)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\Big(a_{2,k}4C^{2(K-k)}L^2 + a_{1,k}4C^{2(K-k)}L^2\frac{1}{N} + 2C^2 b_{1,k}\frac{1}{N} + 2C^2 b_{2,k}\Big)\mathbb{E}[\|h_{n,t+1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2] \\
& \quad + \Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t+1}^{(K-1)} - h_{n,t}^{(K-1)}\|^2] \\
& \quad + 2\alpha^2\eta^4\frac{\sigma^2}{N}\sum_{k=1}^{K-1} a_{1,k}C^{2(K-k)} + 2\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1} a_{2,k}C^{2(K-k)} + 2\alpha^2\eta^4\frac{\sigma^2}{N}\sum_{k=1}^{K-1} b_{1,k} + 2\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1} b_{2,k} + 2a_{1,K}\alpha^2\eta^4\frac{\sigma^2}{N} \\
& \quad + 2a_{2,K}\alpha^2\eta^4\sigma^2 \,.
\end{aligned} \tag{110}$$

By setting

$$
\begin{aligned}
a_{1,k} &= \frac{3\gamma K C^{2(k-1)}}{\alpha\eta},\ k \in \{1,\cdots,K\}\ , \\
b_{1,k} &= \frac{3\gamma A K}{\alpha\eta}\ ,\ k \in \{1,\cdots,K-1\}\ ,
\end{aligned} \tag{111}
$$

we can obtain

$$
\begin{aligned}
& P_{t+1} - P_t \\
& \leq -\frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] - \frac{\gamma\eta}{4}\mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big(-\alpha\eta^2 a_{2,k}\Big)\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\Big(-\alpha\eta^2 b_{2,k}\Big)\mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\Big(-\alpha\eta^2 a_{2,K}\Big)\mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
& \quad + 2\alpha^2\eta^4\frac{\sigma^2}{N}\sum_{k=1}^{K-1}a_{1,k}C^{2(K-k)} + 2\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1}a_{2,k}C^{2(K-k)} + 2\alpha^2\eta^4\frac{\sigma^2}{N}\sum_{k=1}^{K-1}b_{1,k} \\
& \quad + 2\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1}b_{2,k} + 2a_{1,K}\alpha^2\eta^4\frac{\sigma^2}{N} + 2a_{2,K}\alpha^2\eta^4\sigma^2 \\
& \quad + 3\gamma\eta K C^{2(K-1)}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{v}_t^{(K)} - v_{n,t}^{(K)}\|^2] + 3\gamma\eta K\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2}C^{2k}\mathbb{E}[\|\bar{v}_t^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \\
& \quad + 3\gamma\eta K(L^2C^{2(K-1)} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|\bar{x}_t - x_{n,t}\|^2] \\
& \quad + 3\gamma\eta K(L^2C^{2(K-1)} + C^{2K} + AC^2)\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-1}\mathbb{E}[\|\bar{h}_t^{(k)} - h_{n,t}^{(k)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1}\frac{1}{N}\Big)\mathbb{E}[\|v_{n,t+1}^{K} - v_{n,t}^{K}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K-2}\Big(4C^2 a_{2,k} + 4C^2 a_{1,k}\frac{1}{N}\Big)\mathbb{E}[\|v_{n,t+1}^{(k+1)} - v_{n,t}^{(k+1)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2 b_{1,1}\frac{1}{N} + 2C^2 b_{2,1}\Big)\mathbb{E}[\|h_{n,t+1}^{(0)} - h_{n,t}^{(0)}\|^2] \\
& \quad + \frac{1}{N}\sum_{n=1}^{N}\sum_{k=2}^{K-1}\Big(a_{2,k}4C^{2(K-k)}L^2 + a_{1,k}4C^{2(K-k)}L^2\frac{1}{N} + 2C^2 b_{1,k}\frac{1}{N} + 2C^2 b_{2,k}\Big)\mathbb{E}[\|h_{n,t+1}^{(k-1)} - h_{n,t}^{(k-1)}\|^2] \\
& \quad + \Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\|h_{n,t+1}^{(K-1)} - h_{n,t}^{(K-1)}\|^2]\ .
\end{aligned} \tag{112}
$$

Then, we replace the terms in the last eight lines with their upper bound and sum over t from 0 to $T-1$ to obtain

$$
\frac{1}{T}\sum_{t=0}^{T-1}\frac{\gamma\eta}{2}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2]
$$

$$
\begin{aligned}
&\leq \frac{P_0 - P_T}{T} + Z_{\bar{v}} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\|\bar{v}_t^{(1)}\|^2] \\
&+ \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-1} Z_{v-err} \mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2] \\
&+ \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K-1} Z_{h-err} \mathbb{E}[\|h_{n,t}^{(k)} - F_n^{(k)}(h_{n,t}^{(k-1)})\|^2] \\
&+ \frac{1}{T} \sum_{t=0}^{T-1} \frac{1}{N} \sum_{n=1}^{N} Z_{vK-err} \mathbb{E}[\|v_{n,t}^{(K)} - \nabla F_n^{(K)}(h_{n,t}^{(K-1)})\|^2] \\
&+ 256\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1}\Big) \sum_{k=1}^{K-1} (2C^2)^{K-1-k} \\
&+ 128\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1}\Big) + 96\alpha^2\gamma^2p^2\eta^6C^{2K}\Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1}\Big) \\
&+ 192\alpha^2\gamma^2p^2\eta^6\sigma^2C^{2(K-1)}\Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1}\Big) \sum_{k=1}^{K-2} 3^k + 4\alpha^2\eta^4\sigma^2 \sum_{k=2}^{K-1} \omega_j^h \sum_{j=1}^{k-1} (2C^2)^{k-1-j} \\
&+ 2\alpha^2\eta^4 \frac{\sigma^2}{N} \sum_{k=1}^{K-1} a_{1,k} C^{2(K-k)} + 2\alpha^2\eta^4\sigma^2 \sum_{k=1}^{K-1} a_{2,k} C^{2(K-k)} + 2\alpha^2\eta^4 \frac{\sigma^2}{N} \sum_{k=1}^{K-1} b_{1,k} + 2\alpha^2\eta^4\sigma^2 \sum_{k=1}^{K-1} b_{2,k} \\
&+ 2a_{1,K}\alpha^2\eta^4 \frac{\sigma^2}{N} + 2a_{2,K}\alpha^2\eta^4\sigma^2 + 4\alpha^2\eta^4\sigma^2\Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1}\frac{1}{N}\Big) \sum_{k=1}^{K-1} (2C^2)^{K-1-k} \\
&+ 2\alpha^2\eta^4\sigma^2\Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1}\frac{1}{N}\Big) \\
&+ 2\alpha^2\eta^4\sigma^2\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big) \sum_{j=1}^{K-1} (2C^2)^{K-1-j} + 144\gamma\eta p^2\alpha^2\eta^4C^2KC^{2(K-1)} \\
&+ 72\gamma\eta\alpha^2p^2\eta^4\sigma^2L^2KC^{2(K-1)} \sum_{k=1}^{K-1} (2C^2)^{K-1-k} \\
&+ 288\gamma\eta\alpha^2p^2\eta^4\sigma^2C^{2(K-1)}K \sum_{k=1}^{K-1} (2C^2)^{K-1-k} + 144\gamma\eta\alpha^2p^2\eta^4\sigma^2C^{2(K-1)}K + 144\gamma\eta p^2\alpha^2\eta^4C^{2K}K^2 \\
&+ 360\gamma\eta p^2\alpha^2\eta^4\sigma^2K^2(L^2C^{2(K-1)} + C^{2K} + AC^2) + 180\gamma\eta p^2\alpha^2\eta^4\delta^2K^2(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
&+ 4\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2} \omega_j^v (3C^2)^{K-j-1}\Big) \sum_{k=1}^{K-1} (2C^2)^{K-1-k} + 2\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2} \omega_j^v (3C^2)^{K-j-1}\Big) \\
&+ \alpha^2\eta^4\sigma^2 \sum_{k=1}^{K-2} \omega_k^v C^{2(K-k-1)} \sum_{i=k}^{K-2} 3^{i-k+1} \\
&+ 64p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}P_{v_1^c} \sum_{k=1}^{K-1} (2C^2)^{K-1-k} + 32p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}P_{v_1^c} + 24\alpha^2p^2\eta^4C^{2K}P_{v_1^c} \\
&+ 48\alpha^2p^2\eta^4C^{2(K-1)}\sigma^2P_{v_1^c} \sum_{k=1}^{K-2} 3^k \,, \qquad (113)
\end{aligned}
$$

where

$$
\begin{aligned}
Z_{\bar{v}} = \Bigg[& 2\gamma^2\eta^2\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2 b_{1,1}\frac{1}{N} + 2C^2 b_{2,1}\Big) \\
& + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1}\frac{1}{N}\Big) + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big) \\
& + 72\gamma\eta\gamma^2p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} + +144\gamma\eta\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K \\
& + 72\gamma\eta p^2\gamma^2\eta^2C^2K(L^2C^{2(K-1)} + C^{2K} + AC^2) + +2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big) \\
& + \Big(96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)P_{v_1^c} \\
& + 4\gamma^2\eta^2\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big) - \frac{\gamma\eta}{4}\Bigg], \\
Z_{v-var} = \Bigg[& 3\alpha^2\eta^4\Big(\sum_{j=1}^{k-1}\omega_j^v(3C^2)^{k-1-j}\Big)\mathbb{I}_{k>1} + 48\alpha^2p^2\eta^4P_{v_1^c}(3C^2)^{k-1}\mathbb{I}_{k>1} \\
& + 192\alpha^2\gamma^2p^2\eta^6\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(3C^2)^{k-1}\mathbb{I}_{k>1} - \alpha\eta^2a_{2,k}\Bigg], \\
Z_{h-err} = \Bigg[& 4\alpha^2\eta^4\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big)(2C^2)^{K-1-k} + 2\alpha^2\eta^4\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)(2C^2)^{K-1-k} \\
& + 72\gamma\eta\alpha^2p^2\eta^4L^2KC^{2(K-1)}(2C^2)^{K-1-k} + 288\gamma\eta\alpha^2p^2\eta^4C^{2(K-1)}K(2C^2)^{K-1-k} \\
& + 4\alpha^2\eta^4\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)(2C^2)^{K-1-k} + 64p^2\alpha^2\eta^4(3C^2)^{K-1}P_{v_1^c}(2C^2)^{K-1-k} \\
& + 256\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(2C^2)^{K-1-k} + 4\alpha^2\eta^4\sum_{j=k}^{K-2}\omega_{j+1}^h(2C^2)^{j-k} - \alpha\eta^2b_{2,k}\Bigg], \\
Z_{vK-err} = \Bigg[& 2\alpha^2\eta^4\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 6\gamma\eta\alpha^2\eta^4 24p^2C^{2(K-1)}K \\
& + 2\alpha^2\eta^4\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big) + 32p^2\alpha^2\eta^4(3C^2)^{K-1}P_{v_1^c} \\
& + 128\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big) - \alpha\eta^2a_{2,K}\Bigg],
\end{aligned}
\tag{114}
$$

and

$$
\begin{aligned}
P_{v_1^c} = \Bigg[& 3\gamma^3\eta^3p^2K(L^2C^{2(K-1)} + AC^2) + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) \\
& + 2\gamma^2\eta^2\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2b_{1,1}\frac{1}{N} + 2C^2b_{2,1}\Big) \\
& + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big) + 72\gamma\eta\gamma^2p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} \\
& + 144\gamma\eta\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K + 3\gamma\eta K(24p^2\gamma^2\eta^2C^2 + 120\gamma^2\alpha^2p^4\eta^6C^2)(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
& + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\Bigg],
\end{aligned}
\tag{115}
$$

$$\omega_k^v = \left[4C^2 a_{2,k} + 4C^2 a_{1,k}\frac{1}{N} + 72p^2\gamma\eta K C^{2k}\right] , \tag{116}$$

$$\begin{aligned}\omega_k^h = \Bigg[& a_{2,k}4C^{2(K-k)}L^2 + a_{1,k}4C^{2(K-k)}L^2\frac{1}{N} + 2C^2 b_{1,k}\frac{1}{N} + 2C^2 b_{2,k} + 72\gamma\eta p^2 L^2 C^{2(K-1)}K \\ &+ 36\gamma\eta p^2 C^2 K(L^2C^{2(K-1)} + C^{2K} + AC^2) + L^2\Big(\sum_{j=1}^{k-1}\omega_j^v C^{2(K-j-1)}3^{k-j}\Big) + 48p^2(3C^2)^{K-1}L^2 P_{v_1^c}\Bigg] . \end{aligned} \tag{117}$$

Additionally, $\mathbb{I}$ denotes the indicator function.

Based on Lemmas C.18, C.19, C.20, we can know the value of $\{a_{2,k}\}_{k=1}^K$, $\{b_{2,k}\}_{k=1}^{K-1}$, and $P_{v_1^c}$. Specifically, we have

$$\begin{aligned}
& a_{1,k} = \frac{3\gamma K C^{2(k-1)}}{\alpha\eta}, k\in\{1,\cdots,K\} , \\
& a_{2,k} = \omega_{a_k}(C)\frac{\gamma K}{\eta}, k\in\{1,\cdots,K-1\} , \qquad a_{2,K} = \omega_{a_K}(C,\frac{1}{\alpha N})\frac{\gamma K}{\eta} , \\
& b_{1,k} = \frac{3\gamma A K}{\alpha\eta}, k\in\{1,\cdots,K-1\} , \\
& b_{2,k} = \nu_{b_k}(C,L,\frac{1}{\alpha N})\frac{\gamma K}{\eta} , k\in\{1,\cdots,K-2\} , \qquad b_{2,K-1} = \frac{\gamma K}{\eta} , \\
& P_{v_1^c} \le \eta\gamma^3 K\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N}) ,
\end{aligned} \tag{118}$$

where $\omega_{a_1}(C)=1$, $\omega_{a_k}(C)$ is a positive constant with respect to C for $k\in\{2,\cdots,K-1\}$, $\omega_{a_K}(C,\frac{1}{\alpha N})$ is a positive constant with respect to C and $\frac{1}{\alpha N}$, $\nu_{b_k}(C,L,\frac{1}{\alpha N})$ is a positive constant with respect to C, L, and $\frac{1}{\alpha N}$ for $k\in\{1,\cdots,K-2\}$, and $\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})$ is a constant with respect to C, L, and $\frac{1}{\alpha N}$.

Based on these hyperparameters, we enforce $Z_{\bar{v}}\le 0$, $Z_{v-var}\le 0$, $Z_{h-var}\le 0$, $Z_{vK-err}\le 0$.

In the following, we aim to set

$$\begin{aligned}Z_{v-var} =& 3\alpha^2\eta^4\Big(\sum_{j=1}^{k-1}\omega_j^v(3C^2)^{k-1-j}\Big)\mathbb{I}_{k>1} + 48\alpha^2p^2\eta^4 P_{v_1^c}(3C^2)^{k-1}\mathbb{I}_{k>1} \\ &+ 192\alpha^2\gamma^2p^2\eta^6\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(3C^2)^{k-1}\mathbb{I}_{k>1} - \alpha\eta^2 a_{2,k} \le 0 .\end{aligned} \tag{119}$$

Obviously, when $k=1$, it is true. When $k\in\{2,\cdots,K-1\}$, based on the value of $a_{2,k}$, we can know that

$$3\alpha\eta^2\Big(\sum_{j=1}^{k-1}\omega_j^v(3C^2)^{k-1-j}\Big) \le \frac{1}{3}a_{2,k} . \tag{120}$$

Based on the value of $P_{v_1^c}$, we have

$$\begin{aligned}
& 48\alpha p^2\eta^2 P_{v_1^c}(3C^2)^{k-1} \\
\le\ & 48\alpha p^2\eta^2\eta\gamma^3K\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{k-1} \\
\le\ & 48\frac{C}{N}p^2\frac{1}{400p^2C^2}\eta\gamma^3K\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{k-1} \\
\le\ & \frac{1}{C}\eta\gamma^3K\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{k-1} .
\end{aligned} \tag{121}$$

By setting $\frac{1}{C}\eta\gamma^3 K\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})(3C^2)^{k-1} \le \frac{1}{3}a_{2,k} = \frac{1}{3}\omega_{a_k}(C)\frac{\gamma K}{\eta}$, we can get

$$\eta \le \frac{\sqrt{C\omega_{a_k}(C)}}{\gamma\sqrt{3\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})(3C^2)^{k-1}}} \ . \tag{122}$$

Then, based on the value of $b_{2,k}$, we consider

$$192\alpha\gamma^2 p^2\eta^4\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(3C^2)^{k-1} \ . \tag{123}$$

For $k \in \{2, \cdots, K-1\}$, we have

$$\begin{aligned}
\omega_k^h &= a_{2,k}4C^{2(K-k)}L^2 + a_{1,k}4C^{2(K-k)}L^2\frac{1}{N} + 2C^2 b_{1,k}\frac{1}{N} + 2C^2 b_{2,k} \\
&\quad + 72\gamma\eta p^2 L^2 C^{2(K-1)}K + 36\gamma\eta p^2 C^2 K(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
&\quad + L^2\Big(\sum_{j=1}^{k-1}\omega_j^v C^{2(K-j-1)}3^{k-j}\Big) + 48p^2(3C^2)^{K-1}L^2 P_{v_1^c} \\
&\le \frac{\gamma K}{\eta}(\omega_{a_k}(C)4C^{2(K-k)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_k}(C, L, \frac{1}{\alpha N})) \\
&\quad + 36\gamma\eta p^2 K(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) + 48p^2\gamma^3\eta K(3C^2)^{K-1}\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})L^2 \\
&\quad + \frac{\gamma K}{\eta}L^2\sum_{i=1}^{k-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{k-i} + 72p^2\gamma\eta K L^2 C^{2(K-1)}\sum_{i=1}^{k-1}3^{k-i} \ .
\end{aligned} \tag{124}$$

Then, we have

$$\begin{aligned}
&192\alpha\gamma^2 p^2\eta^4(3C^2)^{k-1}\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1} \\
&\le 192\alpha\gamma^2 p^2\eta^4(3C^2)^{k-1}\sum_{j=2}^{K-1}\Big(\frac{\gamma K}{\eta}(\omega_{a_k}(C)4C^{2(K-k)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} \\
&\qquad\qquad + 2C^2\nu_{b_k}(C, L, \frac{1}{\alpha N}))\Big)(2C^2)^{j-1} \\
&\quad + 192\alpha\gamma^2 p^2\eta^4(3C^2)^{k-1}\sum_{j=2}^{K-1}\Big(36\gamma\eta p^2 K(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) \\
&\qquad\qquad + 48p^2\gamma^3\eta K(3C^2)^{K-1}\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})L^2\Big)(2C^2)^{j-1} \\
&\quad + 192\alpha\gamma^2 p^2\eta^4(3C^2)^{k-1}\sum_{j=2}^{K-1}\Big(\frac{\gamma K}{\eta}L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i} \\
&\qquad\qquad + 72p^2\gamma\eta K L^2 C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1} \\
&\le 192\alpha\gamma^2 p^2\eta^4(3C^2)^{k-1}\sum_{j=2}^{K-1}\Big(\frac{\gamma K}{\eta}(\omega_{a_k}(C)4C^{2(K-k)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_k}(C, L, \frac{1}{\alpha N})) \\
&\quad + \frac{\gamma K}{\eta}L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}
\end{aligned}$$

$$
\begin{aligned}
&+ 192\alpha\gamma^2 p^2 \eta^4 (3C^2)^{k-1} \sum_{j=2}^{K-1} \Big(36\gamma\eta p^2 K (2L^2 C^{2(K-1)} + C^2(L^2 C^{2(K-1)} + C^{2K} + AC^2)) \\
&\quad + 72 p^2 \gamma\eta K L^2 C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) (2C^2)^{j-1} \\
&+ 192\alpha\gamma^2 p^2 \eta^4 (3C^2)^{k-1} \Big(48 p^2 \gamma^3 \eta K (3C^2)^{K-1} \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) L^2 \Big) \sum_{j=2}^{K-1} (2C^2)^{j-1} \\
\leq\; & \gamma^2 \eta^2 (3C^2)^{k-1} \frac{\gamma K}{\eta} \frac{1}{C} \sum_{j=2}^{K-1} \Big((\omega_{a_k}(C) 4C^{2(K-k)} L^2 + 12 C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_k}(C, L, \frac{1}{\alpha N})) \\
&\quad + L^2 \sum_{i=1}^{j-1} (4C^2 \omega_{a_i}(C) + \frac{12 C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-1} \\
&+ \gamma^3 \eta K (3C^2)^{k-1} \frac{1}{C^3} \sum_{j=2}^{K-1} \Big((2L^2 C^{2(K-1)} + C^2(L^2 C^{2(K-1)} + C^{2K} + AC^2)) + 2L^2 C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) (2C^2)^{j-1} \\
&+ \gamma^5 \eta K (3C^2)^{k-1} (3C^2)^{K-1} \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) L^2 \frac{1}{C^3} \sum_{j=2}^{K-1} (2C^2)^{j-1} \,.
\end{aligned} \tag{125}
$$

By setting

$$
\begin{aligned}
&\gamma^2 \eta^2 (3C^2)^{k-1} \frac{\gamma K}{\eta} \frac{1}{C} \sum_{j=2}^{K-1} \Big((\omega_{a_k}(C) 4C^{2(K-k)} L^2 + 12 C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_k}(C, L, \frac{1}{\alpha N})) \\
&\qquad + L^2 \sum_{i=1}^{j-1} (4C^2 \omega_{a_i}(C) + \frac{12 C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-1} \leq \frac{1}{9} a_{2,k} = \frac{1}{9} \omega_{a_k}(C) \frac{\gamma K}{\eta} \,, \\
&\gamma^3 \eta K (3C^2)^{k-1} \frac{1}{C^3} \sum_{j=2}^{K-1} \Big((2L^2 C^{2(K-1)} + C^2(L^2 C^{2(K-1)} + C^{2K} + AC^2)) + 2L^2 C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) (2C^2)^{j-1} \\
&\quad \leq \frac{1}{9} a_{2,k} = \frac{1}{9} \omega_{a_k}(C) \frac{\gamma K}{\eta} \,, \\
&\gamma^5 \eta K (3C^2)^{k-1} (3C^2)^{K-1} \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) L^2 \frac{1}{C^3} \sum_{j=2}^{K-1} (2C^2)^{j-1} \leq \frac{1}{9} a_{2,k} = \frac{1}{9} \omega_{a_k}(C) \frac{\gamma K}{\eta} \,,
\end{aligned} \tag{126}
$$

we can get

$$
\begin{aligned}
\eta &\leq \frac{\sqrt{C \omega_{a_k}(C)}}{3\gamma \sqrt{B_1(C, L, \frac{1}{\alpha N})}} \,, \\
\eta &\leq \frac{\sqrt{C^3 \omega_{a_k}(C)}}{3\gamma \sqrt{B_2(C, L)}} \,, \\
\eta &\leq \frac{\sqrt{C^3 \omega_{a_k}(C)}}{3\gamma^2 \sqrt{B_3(C, L, \frac{1}{\alpha N})}} \,,
\end{aligned} \tag{127}
$$

where

$$
B_1(C, L, \frac{1}{\alpha N}) = (3C^2)^{k-1} \sum_{j=2}^{K-1} \Big((\omega_{a_k}(C) 4C^{2(K-k)} L^2 + 12 C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_k}(C, L, \frac{1}{\alpha N}))
$$

$$+ L^2 \sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1} ,$$

$$B_2(C,L) = (3C^2)^{k-1}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) + 2L^2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1} ,$$

$$B_3(C,L,\frac{1}{\alpha N}) = (3C^2)^{k-1}(3C^2)^{K-1}\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})L^2\sum_{j=2}^{K-1}(2C^2)^{j-1} . \tag{128}$$

As a result, we can know that the coefficient of $\mathbb{E}[\|v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)})^T v_{n,t}^{(k+1)}\|^2]$ is negative.

When $k = K$, we aim to enforce

$$2\alpha^2\eta^4\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 6\gamma\eta\alpha^2\eta^4 24p^2C^{2(K-1)}K + 2\alpha^2\eta^4\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)$$
$$+ 32p^2\alpha^2\eta^4(3C^2)^{K-1}P_{v_1^c} + 128\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big) - \alpha\eta^2a_{2,K} \le 0 . \tag{129}$$

In particular, we have

$$2\alpha\eta^2\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 144\gamma\eta\alpha\eta^2p^2C^{2(K-1)}K$$
$$+ 2\alpha\eta^2\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big) + 32p^2\alpha\eta^2(3C^2)^{K-1}P_{v_1^c} + 128\alpha\gamma^2p^2\eta^4(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)$$
$$\le 2\alpha\eta^2\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + 4C^2\frac{3C^{2(K-2)}}{\alpha N}\Big) + 144\alpha\eta^2p^2\gamma\eta KC^{2(K-1)} + 32\alpha\eta^2p^2\gamma^3\eta K(3C^2)^{K-1}\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})$$
$$+ 2\alpha\eta^2\frac{\gamma K}{\eta}\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1} + 144\alpha\eta^2p^2\gamma\eta K\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}$$
$$+ 128\alpha p^2\eta^4(3C^2)^{K-1}\frac{\gamma^3K}{\eta}\sum_{j=2}^{K-1}\Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N})$$
$$+ L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}$$
$$+ 128\times 36\alpha p^4\eta^4(3C^2)^{K-1}\gamma^3\eta K\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))$$
$$+ 2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1}$$
$$+ 128\times 48\alpha\gamma^2p^4\eta^4(3C^2)^{K-1}\gamma^3\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\sum_{j=2}^{K-1}(2C^2)^{j-1}$$
$$\le 2\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + 4C^2\frac{3C^{2(K-2)}}{\alpha N} + \sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\Big)$$
$$+ \gamma\eta K\frac{1}{C}\Big(C^{2(K-1)} + \sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}\Big)$$

$$
\begin{aligned}
&+\gamma^3\eta K\frac{1}{C}\Big((3C^2)^{K-1}\sum_{j=2}^{K-1}(\omega_{a_j}(C)4C^{2(K-j)}L^2+12C^{2(K-1)}L^2\frac{1}{\alpha N}+6AC^2\frac{1}{\alpha N}+2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N})\\
&\quad+L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C)+\frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i})(2C^2)^{j-1}+(3C^2)^{K-1}\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\Big)\\
&+\gamma^3\eta K\frac{(3C^2)^{K-1}}{C^3}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)}+C^2(L^2C^{2(K-1)}+C^{2K}+AC^2))+2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1}\\
&+\gamma^5\eta K\frac{(3C^2)^{2(K-1)}}{C^3}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\sum_{j=2}^{K-1}(2C^2)^{j-1}\,. \qquad (130)
\end{aligned}
$$

Based on the value of $a_{2,K}$, we have already known $2\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C)+4C^2\frac{3C^{2(K-2)}}{\alpha N}+\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C)+\frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\Big)\leq\frac{1}{4}a_{2,K}$ in Lemma C.18. Then, we set

$$
\begin{aligned}
&\gamma\eta K\frac{1}{C}\Big(C^{2(K-1)}+\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}\Big)\leq\frac{1}{16}\omega_{a_K}(C,\frac{1}{\alpha N})\frac{\gamma K}{\eta}\,,\\
&\gamma^3\eta K\frac{1}{C}\Big((3C^2)^{K-1}\sum_{j=2}^{K-1}(\omega_{a_j}(C)4C^{2(K-j)}L^2+12C^{2(K-1)}L^2\frac{1}{\alpha N}+6AC^2\frac{1}{\alpha N}+2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N})\\
&\quad+L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C)+\frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i})(2C^2)^{j-1}+(3C^2)^{K-1}\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\Big)\leq\frac{1}{16}\omega_{a_K}(C,\frac{1}{\alpha N})\frac{\gamma K}{\eta}\,,\\
&\gamma^3\eta K\frac{(3C^2)^{K-1}}{C^3}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)}+C^2(L^2C^{2(K-1)}+C^{2K}+AC^2))+2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1}\\
&\quad\leq\frac{1}{16}\omega_{a_K}(C,\frac{1}{\alpha N})\frac{\gamma K}{\eta}\,,\\
&\gamma^5\eta K\frac{(3C^2)^{2(K-1)}}{C^3}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\sum_{j=2}^{K-1}(2C^2)^{j-1}\leq\frac{1}{16}\omega_{a_K}(C,\frac{1}{\alpha N})\frac{\gamma K}{\eta}\,, \qquad (131)
\end{aligned}
$$

We can get

$$
\eta\leq\frac{\sqrt{C\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\sqrt{E_1(C)}}\,,\quad \eta\leq\frac{\sqrt{C\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma\sqrt{E_2(C,L,\frac{1}{\alpha N})}}\,,\eta\leq\frac{\sqrt{C^3\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma\sqrt{E_3(C,L)}}\,,\quad \eta\leq\frac{\sqrt{C^3\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma^2\sqrt{E_4(C,L,\frac{1}{\alpha N})}}\,, \qquad (132)
$$

where

$$
\begin{aligned}
&E_1(C)=C^{2(K-1)}+\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}\,,\ E_4(C,L,\frac{1}{\alpha N})=(3C^2)^{2(K-1)}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\sum_{j=2}^{K-1}(2C^2)^{j-1}\,,\\
&E_2(C,L,\frac{1}{\alpha N})=\Big((3C^2)^{K-1}\sum_{j=2}^{K-1}(\omega_{a_j}(C)4C^{2(K-j)}L^2+12C^{2(K-1)}L^2\frac{1}{\alpha N}+6AC^2\frac{1}{\alpha N}+2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N})\\
&\quad+L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C)+\frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i})(2C^2)^{j-1}+(3C^2)^{K-1}\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\Big)\,,\\
&E_3(C,L)=(3C^2)^{K-1}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)}+C^2(L^2C^{2(K-1)}+C^{2K}+AC^2))+2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1}\,. \qquad (133)
\end{aligned}
$$

Then, we aim to enforce

$$
\begin{aligned}
Z_{h-err} \leq\ & 4\alpha^2\eta^4\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-)}}{\alpha N} + \frac{3C^{2(K-1)}}{\alpha N}L^2 + \omega_{a_K}(C, \frac{1}{\alpha N})L^2\Big)(2C^2)^{K-1-k} \\
& + 4\alpha^2\eta^4(2C^2)^{K-1-k}\frac{\gamma K}{\eta}\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1} \\
& + 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(\omega_{a_{j+1}}(C)4C^{2(K-j-1)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N}\Big)(2C^2)^{j-k} \\
& + 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(L^2\sum_{i=1}^{j}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-k} \\
& + 72\alpha^2p^2\eta^4\gamma\eta K\Big((L^2+4)C^{2(K-1)}(2C^2)^{K-1-k} + 4(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} \\
& \quad + 2(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\sum_{j=k}^{K-2}(2C^2)^{j-k} + 4L^2C^{2(K-1)}\sum_{j=k}^{K-2}\Big(\sum_{i=1}^{j}3^{j-i}\Big)(2C^2)^{j-k}\Big) \\
& + 64p^2\alpha^2\eta^4\gamma^3\eta K(3C^2)^{K-1}\Big[\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})((2C^2)^{K-1-k} + 3L^2\sum_{j=k}^{K-2}(2C^2)^{j-k}) \\
& \quad + 4(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
& \quad + L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}\Big] \\
& + 256\times 36\alpha^2p^4\eta^6\gamma^3\eta K(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) \\
& \qquad + 2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1} \\
& + 256\times 48\alpha^2p^4\eta^6\gamma^5\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}(2C^2)^{j-1} \\
& + 4\alpha^2\eta^4\sum_{j=k}^{K-2}2C^2b_{2,j+1}(2C^2)^{j-k} - \alpha\eta^2b_{2,k} \leq 0\,,
\end{aligned}
\tag{134}
$$

where the first step follows from the following inequality:

$$
\begin{aligned}
& 256\alpha^2p^2\eta^6\gamma^2(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1} \\
& = 256\alpha^2p^2\eta^6\gamma^2(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Bigg(\frac{\gamma K}{\eta}\Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} \\
& \quad + 2C^2\nu_{b_j}(C, L, \frac{1}{\alpha N}) + L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)\Bigg)(2C^2)^{j-1} \\
& \quad + 256\alpha^2p^2\eta^6\gamma^2(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Bigg(36\gamma\eta p^2K\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))
\end{aligned}
$$

$$+ 2C^{2(K-1)}\sum_{i=1}^{j-1} 3^{j-i}\Big)\Bigg)(2C^2)^{j-1}$$

$$+ 256\alpha^2p^2\eta^6\gamma^2(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Bigg(48p^2\gamma^3\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\Bigg)(2C^2)^{j-1}$$

$$\le 256\alpha^2p^2\eta^4\gamma^3\eta K(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Bigg(\Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N}$$

$$+ 2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N}) + L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)\Bigg)(2C^2)^{j-1}$$

$$+ 256\times 36\alpha^2p^4\eta^6\gamma^3\eta K(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Bigg(\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))$$

$$+ 2C^{2(K-1)}\sum_{i=1}^{j-1} 3^{j-i}\Big)\Bigg)(2C^2)^{j-1}$$

$$+ 256\times 48\alpha^2p^4\eta^6\gamma^5\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}(2C^2)^{j-1}\ . \tag{135}$$

Due to the value of $b_{2,k}$ in Lemma C.19, we have already known that

$$4\alpha^2\eta^4\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-)}}{\alpha N} + \frac{3C^{2(K-1)}}{\alpha N}L^2 + \omega_{a_K}(C,\frac{1}{\alpha N})L^2\Big)(2C^2)^{K-1-k}$$

$$+ 4\alpha^2\eta^4(2C^2)^{K-1-k}\frac{\gamma K}{\eta}\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}$$

$$+ 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(\omega_{a_{j+1}}(C)4C^{2(K-j-1)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N}\Big)(2C^2)^{j-k}$$

$$+ 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(L^2\sum_{i=1}^{j}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-k}$$

$$+ 4\alpha^2\eta^4\sum_{j=k}^{K-2}\omega_{j+1}^h(2C^2)^{j-k} \le \frac{1}{2}\alpha\eta^2 b_{2,k}\ . \tag{136}$$

Then, we enforce

$$72\alpha p^2\eta^2\gamma\eta K\Big((L^2+4)C^{2(K-1)}(2C^2)^{K-1-k} + 4(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}$$

$$+ 2(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\sum_{j=k}^{K-2}(2C^2)^{j-k}$$

$$+ 4L^2C^{2(K-1)}\sum_{j=k}^{K-2}\Big(\sum_{i=1}^{j}3^{j-i}\Big)(2C^2)^{j-k}\Big)$$

$$+ 64p^2\alpha\eta^2\gamma^3\eta K(3C^2)^{K-1}\Big[\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})((2C^2)^{K-1-k} + 3L^2\sum_{j=k}^{K-2}(2C^2)^{j-k})$$

$$
\begin{aligned}
&+ 4(2C^2)^{K-1-k} \sum_{j=2}^{K-1} \Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
&+ L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}\Big] \\
&+ 256\times 36\alpha p^4\eta^4\gamma^3\eta K(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) \\
&+ 2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1} \\
&+ 256\times 48\alpha p^4\eta^4\gamma^5\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}(2C^2)^{j-1} \\
\le\ & \gamma\eta K\frac{1}{C}\Big((L^2+4)C^{2(K-1)}(2C^2)^{K-1-k} + 4(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} \\
&+ 2(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\sum_{j=k}^{K-2}(2C^2)^{j-k} \\
&+ 4L^2C^{2(K-1)}\sum_{j=k}^{K-2}\Big(\sum_{i=1}^{j}3^{j-i}\Big)(2C^2)^{j-k}\Big) \\
&+ \gamma^3\eta K(3C^2)^{K-1}\frac{1}{C}\Big[\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})((2C^2)^{K-1-k} + 3L^2\sum_{j=k}^{K-2}(2C^2)^{j-k}) \\
&+ 4(2C^2)^{K-1-k} \sum_{j=2}^{K-1} \Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
&+ L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}\Big] \\
&+ \gamma^3\eta K(3C^2)^{K-1}(2C^2)^{K-1-k}\frac{1}{C^3}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) \\
&+ 2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1} \\
&+ \gamma^5\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})(3C^2)^{K-1}(2C^2)^{K-1-k}\frac{1}{C^3}\sum_{j=2}^{K-1}(2C^2)^{j-1} \\
\le\ & \frac{1}{2}b_{2,k}\ . \qquad (137)
\end{aligned}
$$

In particular, we set

$$
\begin{aligned}
&\gamma\eta K\frac{1}{C}\Big((L^2+4)C^{2(K-1)}(2C^2)^{K-1-k} + 4(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} \\
&+ 2(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\sum_{j=k}^{K-2}(2C^2)^{j-k}
\end{aligned}
$$

$$
\begin{aligned}
&+ 4L^2C^{2(K-1)}\sum_{j=k}^{K-2}\Big(\sum_{i=1}^{j}3^{j-i}\Big)(2C^2)^{j-k}\Big) \le \frac{1}{8}b_{2,k} = \frac{\nu_{b_k}(C,L,\frac{1}{\alpha N})}{8}\frac{\gamma K}{\eta}\ ,\\
&\gamma^3\eta K(3C^2)^{K-1}\frac{1}{C}\Big[\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})((2C^2)^{K-1-k} + 3L^2\sum_{j=k}^{K-2}(2C^2)^{j-k})\\
&+ 4(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N})\\
&+ L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}\Big] \le \frac{1}{8}b_{2,k} = \frac{\nu_{b_k}(C,L,\frac{1}{\alpha N})}{8}\frac{\gamma K}{\eta}\ ,\\
&\gamma^3\eta K(3C^2)^{K-1}(2C^2)^{K-1-k}\frac{1}{C^3}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\\
&+ 2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1} \le \frac{1}{8}b_{2,k} = \frac{\nu_{b_k}(C,L,\frac{1}{\alpha N})}{8}\frac{\gamma K}{\eta}\ ,\\
&\gamma^5\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{K-1}(2C^2)^{K-1-k}\frac{1}{C^3}\sum_{j=2}^{K-1}(2C^2)^{j-1} \le \frac{1}{8}b_{2,k} = \frac{\nu_{b_k}(C,L,\frac{1}{\alpha N})}{8}\frac{\gamma K}{\eta}\ .
\end{aligned}
\tag{138}
$$

As a result, we can get

$$
\eta \le \frac{\sqrt{C\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\sqrt{D_1(C,L)}}\ , \eta \le \frac{\sqrt{C\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma\sqrt{D_2(C,L,\frac{1}{\alpha N})}}\ , \eta \le \frac{\sqrt{C^3\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma\sqrt{D_3(C,L)}}\ , \eta \le \frac{\sqrt{C^3\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma^2\sqrt{D_4(C,L,\frac{1}{\alpha N})}}\ , \tag{139}
$$

where

$$
\begin{aligned}
D_1(C,L) &= \Big((L^2+4)C^{2(K-1)}(2C^2)^{K-1-k} + 4(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}\\
&+ 2(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\sum_{j=k}^{K-2}(2C^2)^{j-k} + 4L^2C^{2(K-1)}\sum_{j=k}^{K-2}\Big(\sum_{i=1}^{j}3^{j-i}\Big)(2C^2)^{j-k}\Big)\ ,\\
D_2(C,L,\frac{1}{\alpha N}) &= (3C^2)^{K-1}\Big[\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})((2C^2)^{K-1-k} + 3L^2\sum_{j=k}^{K-2}(2C^2)^{j-k})\\
&+ 4(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Big(\omega_{a_j}(C)4C^{2(K-j)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_j}(C,L,\frac{1}{\alpha N})\\
&\qquad + L^2\sum_{i=1}^{j-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-1}\Big]\ ,\\
D_3(C,L) &= (3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\\
&\qquad + 2C^{2(K-1)}\sum_{i=1}^{j-1}3^{j-i}\Big)(2C^2)^{j-1}\ ,\\
D_4(C,L,\frac{1}{\alpha N}) &= (3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{K-1}(2C^2)^{K-1-k}\sum_{j=2}^{K-1}(2C^2)^{j-1}\ .
\end{aligned}
\tag{140}
$$

In the following, we aim to enforce

$$
\begin{aligned}
&2\gamma\eta\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2b_{1,1}\frac{1}{N} + 2C^2b_{2,1}\Big)\\
&+ 2\gamma\eta(2C^2)^{K-1}\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 2\gamma\eta(2C^2)^{K-1}\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)\\
&+ 72\gamma^2p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} + 144\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K\\
&+ 72p^2\gamma^2\eta^2C^2K(L^2C^{2(K-1)} + C^{2K} + AC^2) + 2\gamma\eta(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\\
&+ 32p^2\gamma\eta\Big(3C^{2(K-1)}L^2 + (3C^2)^{2(K-1)}\Big)P_{v_1^c}\\
&+ 4\gamma\eta\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big) - \frac{1}{4} \leq 0\,. \qquad (141)
\end{aligned}
$$

In particular, we have

$$
\begin{aligned}
&2\gamma\eta\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2b_{1,1}\frac{1}{N} + 2C^2b_{2,1}\Big)\\
&\quad + 2\gamma\eta(2C^2)^{K-1}\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 2\gamma\eta(2C^2)^{K-1}\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)\\
&\quad + 72\gamma^2p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} + 144\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K\\
&\quad + 72p^2\gamma^2\eta^2C^2K(L^2C^{2(K-1)} + C^{2K} + AC^2) + 2\gamma\eta(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\\
&\quad + 32p^2\gamma\eta\Big(3C^{2(K-1)}L^2 + (3C^2)^{2(K-1)}\Big)P_{v_1^c}\\
&\quad + 4\gamma\eta\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\\
&\leq 2\gamma\eta\Big(\frac{\gamma K}{\eta}4C^{2(K-1)}L^2 + \frac{3\gamma K}{\alpha\eta}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2\frac{3\gamma AK}{\alpha\eta}\frac{1}{N} + 2C^2\nu_{b_1}(C,L,\frac{1}{\alpha N})\frac{\gamma K}{\eta}\Big)\\
&\quad + 2\gamma\eta(2C^2)^{K-1}\Big(4C^2\omega_{a_{K-1}}(C)\frac{\gamma K}{\eta} + 4C^2\frac{3\gamma KC^{2(K-2)}}{\alpha\eta}\frac{1}{N}\Big)\\
&\quad + 2\gamma\eta(2C^2)^{K-1}\Big(2\frac{3\gamma KC^{2(K-1)}}{\alpha\eta}L^2\frac{1}{N} + 2\omega_{a_K}(C,\frac{1}{\alpha N})\frac{\gamma K}{\eta}L^2\Big)\\
&\quad + 72\gamma^2p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} + 144\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K\\
&\quad + 72p^2\gamma^2\eta^2C^2K(L^2C^{2(K-1)} + C^{2K} + AC^2) + 2\gamma\eta(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\\
&\quad + 32p^2\gamma\eta\Big(3C^{2(K-1)}L^2 + (3C^2)^{2(K-1)}\Big)\gamma^3\eta K\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})\\
&\quad + 4\gamma\eta\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\Big(1 + 96p^2\gamma^2\eta^2C^{2(K-1)}L^2 + 32p^2\gamma^2\eta^2(3C^2)^{2(K-1)}\Big)\\
&\leq 2\gamma^2K\Big(4C^{2(K-1)}L^2 + \frac{1}{\alpha N}12C^{2(K-1)}L^2 + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_1}(C,L,\frac{1}{\alpha N})\Big)\\
&\quad + 2\gamma^2K(2C^2)^{K-1}\Big(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-1)}}{\alpha N}\Big) + 2\gamma^2K(2C^2)^{K-1}\Big(\frac{6C^{2(K-1)}}{\alpha N}L^2 + 2\omega_{a_K}(C,\frac{1}{\alpha N})L^2\Big)\\
&\quad + \gamma^2K(2C^2)^{K-2}L^2C^{2(K-1)} + \gamma^2K(2C^2)^{K-2}C^{2(K-1)} + \gamma^2K(L^2C^{2(K-1)} + C^{2K} + AC^2)\\
&\quad + 2\gamma\eta(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big) + \frac{1}{C^2}\gamma^4K\Big(3C^{2(K-1)}L^2 + (3C^2)^{2(K-1)}\Big)\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})
\end{aligned}
$$

$$+ 4\gamma\eta \sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1} + 4\gamma\eta \sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1} \gamma^2 (C^{2(K-2)} L^2 + (3C^2)^{2(K-1)-1}) \,. \tag{142}$$

Then, we bound the following two terms.

$$\begin{aligned}
&2\gamma\eta(2C^2)^{K-1} \sum_{j=1}^{K-2} \omega_j^v (3C^2)^{K-j-1} \leq 2\gamma\eta(2C^2)^{K-1} \frac{\gamma K}{\eta} \sum_{j=1}^{K-2} \Big(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N}\Big)(3C^2)^{K-j-1} \\
&\quad + 144\gamma^2 p^2 \eta^2 K (2C^2)^{K-1} \sum_{j=1}^{K-2} C^{2j} (3C^2)^{K-j-1} \\
&\leq 2\gamma^2 K (2C^2)^{K-1} \sum_{j=1}^{K-2} \Big(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N}\Big)(3C^2)^{K-j-1} \\
&\quad + \gamma^2 K (2C^2)^{K-2} \sum_{j=1}^{K-2} C^{2j} (3C^2)^{K-j-1} \,.
\end{aligned} \tag{143}$$

Additionally, we have

$$\begin{aligned}
&4\gamma\eta \sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1} \\
&\leq 4\gamma\eta \sum_{j=2}^{K-1} \Bigg(\frac{\gamma K}{\eta} \Big(\omega_{a_j}(C) 4C^{2(K-j)} L^2 + 12C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
&\quad + L^2 \sum_{i=1}^{j-1} (4C^2 \omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) \Bigg) (2C^2)^{j-1} \\
&\quad + 4\gamma\eta \sum_{j=2}^{K-1} \Bigg(36\gamma\eta p^2 K \Big((2L^2 C^{2(K-1)} + C^2 (L^2 C^{2(K-1)} + C^{2K} + AC^2)) + 2C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) \Bigg) (2C^2)^{j-1} \\
&\quad + 4\gamma\eta \sum_{j=2}^{K-1} \Bigg(48 p^2 \gamma^3 \eta K (3C^2)^{K-1} L^2 \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) \Bigg) (2C^2)^{j-1} \\
&\leq 4\gamma^2 K \sum_{j=2}^{K-1} \Bigg(\Big(\omega_{a_j}(C) 4C^{2(K-j)} L^2 + 12C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
&\quad + L^2 \sum_{i=1}^{j-1} (4C^2 \omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) \Bigg) (2C^2)^{j-1} \\
&\quad + 144\gamma^2 \eta^2 p^2 K \sum_{j=2}^{K-1} \Bigg(\Big((2L^2 C^{2(K-1)} + C^2 (L^2 C^{2(K-1)} + C^{2K} + AC^2)) + 2C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) \Bigg) (2C^2)^{j-1} \\
&\quad + 192\eta^2 p^2 \gamma^4 K (3C^2)^{K-1} L^2 \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) \sum_{j=2}^{K-1} (2C^2)^{j-1} \\
&\leq 4\gamma^2 K \sum_{j=2}^{K-1} \Big(\omega_{a_j}(C) 4C^{2(K-j)} L^2 + 12C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
&\quad + L^2 \sum_{i=1}^{j-1} (4C^2 \omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-1} \\
&\quad + \gamma^2 K \sum_{j=2}^{K-1} \Big((2L^2 C^{2(K-1)} + C^2 (L^2 C^{2(K-1)} + C^{2K} + AC^2)) + 2C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) (2C^2)^{j-2}
\end{aligned}$$

$$
\begin{aligned}
&+ \gamma^4 K(3C^2)^{K-1} L^2 \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) \sum_{j=2}^{K-1} (2C^2)^{j-2} \\
&= \gamma^2 K J_1 + \gamma^4 K J_2 \,,
\end{aligned} \tag{144}
$$

where

$$
\begin{aligned}
J_1(C, L, \frac{1}{\alpha N}) &= 4 \sum_{j=2}^{K-1} \Big(\omega_{a_j}(C) 4C^{2(K-j)} L^2 + 12 C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_j}(C, L, \frac{1}{\alpha N}) \\
&\quad + L^2 \sum_{i=1}^{j-1} (4C^2 \omega_{a_i}(C) + \frac{12 C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-1} \\
&\quad + \sum_{j=2}^{K-1} \Big((2L^2 C^{2(K-1)} + C^2 (L^2 C^{2(K-1)} + C^{2K} + AC^2)) + 2C^{2(K-1)} \sum_{i=1}^{j-1} 3^{j-i} \Big) (2C^2)^{j-2} \,, \\
J_2(C, L, \frac{1}{\alpha N}) &= (3C^2)^{K-1} L^2 \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) \sum_{j=2}^{K-1} (2C^2)^{j-2} \,.
\end{aligned} \tag{145}
$$

Therefore, we have

$$
\begin{aligned}
&2\gamma\eta \Big(a_{2,1} 4C^{2(K-1)} L^2 + a_{1,1} 4C^{2(K-1)} L^2 \frac{1}{N} + 2C^2 b_{1,1} \frac{1}{N} + 2C^2 b_{2,1} \Big) \\
&+ 2\gamma\eta (2C^2)^{K-1} \Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1} \frac{1}{N} \Big) + 2\gamma\eta (2C^2)^{K-1} \Big(2a_{1,K} L^2 \frac{1}{N} + 2a_{2,K} L^2 \Big) \\
&+ 72 \gamma^2 p^2 \eta^2 (2C^2)^{K-1} L^2 K C^{2(K-1)} + 144 \gamma^2 p^2 \eta^2 (2C^2)^{K-1} C^{2(K-1)} K \\
&+ 72 p^2 \gamma^2 \eta^2 C^2 K (L^2 C^{2(K-1)} + C^{2K} + AC^2) + 2\gamma\eta (2C^2)^{K-1} \Big(\sum_{j=1}^{K-2} \omega_j^v (3C^2)^{K-j-1} \Big) \\
&+ 32 p^2 \gamma\eta \Big(3C^{2(K-1)} L^2 + (3C^2)^{2(K-1)} \Big) P_{v_1^c} \\
&+ 4\gamma\eta \Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1} \Big) \Big(1 + 96 p^2 \gamma^2 \eta^2 C^{2(K-1)} L^2 + 32 p^2 \gamma^2 \eta^2 (3C^2)^{2(K-1)} \Big) \\
&\leq 2\gamma^2 K \Big(4C^{2(K-1)} L^2 + \frac{1}{\alpha N} 12 C^{2(K-1)} L^2 + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_1}(C, L, \frac{1}{\alpha N}) \Big) \\
&+ 2\gamma^2 K (2C^2)^{K-1} \Big(4C^2 \omega_{a_{K-1}}(C) + \frac{12 C^{2(K-1)}}{\alpha N} \Big) + 2\gamma^2 K (2C^2)^{K-1} \Big(\frac{6C^{2(K-1)}}{\alpha N} L^2 + 2\omega_{a_K}(C, \frac{1}{\alpha N}) L^2 \Big) \\
&+ \gamma^2 K (2C^2)^{K-2} L^2 C^{2(K-1)} + \gamma^2 K (2C^2)^{K-2} C^{2(K-1)} + \gamma^2 K (L^2 C^{2(K-1)} + C^{2K} + AC^2) \\
&+ 2\gamma^2 K (2C^2)^{K-1} \sum_{j=1}^{K-2} \Big(4C^2 \omega_{a_j}(C) + \frac{12 C^{2j}}{\alpha N} \Big) (3C^2)^{K-j-1} + \gamma^2 K (2C^2)^{K-2} \sum_{j=1}^{K-2} C^{2j} (3C^2)^{K-j-1} \\
&+ \frac{1}{C^2} \gamma^4 K \Big(3C^{2(K-1)} L^2 + (3C^2)^{2(K-1)} \Big) \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) + \gamma^2 K J_1 + \gamma^4 K J_2 \\
&+ \gamma^4 K J_1 (C^{2(K-2)} L^2 + (3C^2)^{2(K-1)-1}) + \gamma^6 K J_2 (C^{2(K-2)} L^2 + (3C^2)^{2(K-1)-1}) \leq \frac{1}{4} \,.
\end{aligned} \tag{146}
$$

By denoting

$$
\begin{aligned}
J_3(C, L, \frac{1}{\alpha N}) &= 2K \Big(4C^{2(K-1)} L^2 + \frac{1}{\alpha N} 12 C^{2(K-1)} L^2 + 6AC^2 \frac{1}{\alpha N} + 2C^2 \nu_{b_1}(C, L, \frac{1}{\alpha N}) \Big) \\
&+ 2K (2C^2)^{K-1} \Big(4C^2 \omega_{a_{K-1}}(C) + \frac{12 C^{2(K-1)}}{\alpha N} \Big) + 2K (2C^2)^{K-1} \Big(\frac{6C^{2(K-1)}}{\alpha N} L^2 + 2\omega_{a_K}(C, \frac{1}{\alpha N}) L^2 \Big) \\
&+ K (2C^2)^{K-2} L^2 C^{2(K-1)} + K (2C^2)^{K-2} C^{2(K-1)}
\end{aligned}
$$

$$
\begin{aligned}
&+ K(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
&+ 2K(2C^2)^{K-1}\sum_{j=1}^{K-2}\Big(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N}\Big)(3C^2)^{K-j-1} \\
&+ K(2C^2)^{K-2}\sum_{j=1}^{K-2} C^{2j}(3C^2)^{K-j-1} + KJ_1 \ , \\
J_4(C, L, \frac{1}{\alpha N}) &= \frac{1}{C^2}K\Big(3C^{2(K-1)}L^2 + (3C^2)^{2(K-1)}\Big)\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) + KJ_1(C^{2(K-2)}L^2 + (3C^2)^{2(K-1)-1}) \ , \\
J_5(C, L, \frac{1}{\alpha N}) &= KJ_2(C^{2(K-2)}L^2 + (3C^2)^{2(K-1)-1}) \ , \qquad (147)
\end{aligned}
$$

we can get

$$
\gamma \le \frac{1}{4\sqrt{J_3(C, L, \frac{1}{\alpha N})K}} \ , \gamma \le \frac{1}{2(J_4(C, L, \frac{1}{\alpha N})K)^{1/4}} \ , \gamma \le \frac{1}{(16J_5(C, L, \frac{1}{\alpha N})K)^{1/6}} \ . \qquad (148)
$$

Finally, we can obtain

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \le \frac{2(P_0 - P_T)}{\eta\gamma T} \\
&+ \frac{2}{\gamma\eta}\Bigg[256\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} \\
&+ 128\alpha^2\gamma^2p^2\eta^6\sigma^2(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big) + 96\alpha^2\gamma^2p^2\eta^6C^{2K}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big) \\
&+ 192\alpha^2\gamma^2p^2\eta^6\sigma^2C^{2(K-1)}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\sum_{k=1}^{K-2}3^k + 4\alpha^2\eta^4\sigma^2\sum_{k=2}^{K-1}\omega_j^h\sum_{j=1}^{k-1}(2C^2)^{k-1-j} \\
&+ 2\alpha^2\eta^4\frac{\sigma^2}{N}\sum_{k=1}^{K-1}a_{1,k}C^{2(K-k)} + 2\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1}a_{2,k}C^{2(K-k)} + 2\alpha^2\eta^4\frac{\sigma^2}{N}\sum_{k=1}^{K-1}b_{1,k} + 2\alpha^2\eta^4\sigma^2\sum_{k=1}^{K-1}b_{2,k} + 2a_{1,K}\alpha^2\eta^4\frac{\sigma^2}{N} \\
&+ 2a_{2,K}\alpha^2\eta^4\sigma^2 + 4\alpha^2\eta^4\sigma^2\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} \\
&+ 2\alpha^2\eta^4\sigma^2\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) \\
&+ 2\alpha^2\eta^4\sigma^2\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)\sum_{j=1}^{K-1}(2C^2)^{K-1-j} + 144\gamma\eta p^2\alpha^2\eta^4C^2KC^{2(K-1)} \\
&+ 72\gamma\eta\alpha^2p^2\eta^4\sigma^2L^2KC^{2(K-1)}\sum_{k=1}^{K-1}(2C^2)^{K-1-k} \\
&+ 288\gamma\eta\alpha^2p^2\eta^4\sigma^2C^{2(K-1)}K\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 144\gamma\eta\alpha^2p^2\eta^4\sigma^2C^{2(K-1)}K + 144\gamma\eta p^2\alpha^2\eta^4C^{2K}K^2 \\
&+ 360\gamma\eta p^2\alpha^2\eta^4\sigma^2K^2(L^2C^{2(K-1)} + C^{2K} + AC^2) + 180\gamma\eta p^2\alpha^2\eta^4\delta^2K^2(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
&+ 4\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 2\alpha^2\eta^4\sigma^2\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big) \\
&+ \alpha^2\eta^4\sigma^2\sum_{k=1}^{K-2}\omega_k^vC^{2(K-k-1)}\sum_{i=k}^{K-2}3^{i-k+1}
\end{aligned}
$$

$$+ 64p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}P_{v_1^c}\sum_{k=1}^{K-1}(2C^2)^{K-1-k} + 32p^2\alpha^2\eta^4\sigma^2(3C^2)^{K-1}P_{v_1^c} + 24\alpha^2p^2\eta^4C^{2K}P_{v_1^c}$$
$$+ 48\alpha^2p^2\eta^4C^{2(K-1)}\sigma^2P_{v_1^c}\sum_{k=1}^{K-2}3^k\Bigg] . \tag{149}$$

Then, based on the value of $\{a_{1,k}\}_{k=1}^K$, $\{a_{2,k}\}_{k=1}^K$, $\{b_{1,k}\}_{k=1}^{K-1}$, and $\{b_{2,k}\}_{k=1}^{K-1}$, we have

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2] \le \frac{2(P_0-P_T)}{\eta\gamma T}$$
$$+ O(\gamma^2\alpha^2p^2\eta^4) + O(\gamma^2\alpha^2p^4\eta^6) + O(\gamma^4\alpha^2p^4\eta^6) + O(\alpha^2\eta^2) + O(\alpha^2p^2\eta^4) + O(\frac{\alpha\eta^2}{N}) . \tag{150}$$

When $t=0$, we have

$$P_0 \le \mathbb{E}[\Phi(\bar{x}_0)] + \sum_{k=1}^{K-1}a_{1,k}\frac{\sigma^2C^{K-k}}{SN} + a_{1,K}\frac{\sigma^2}{SN} + \sum_{k=1}^{K-1}b_{1,k}\frac{\sigma^2}{SN} + \sum_{k=1}^{K-1}a_{2,k}\frac{\sigma^2C^{K-k}}{S} + a_{2,K}\frac{\sigma^2}{S} + \sum_{k=1}^{K-1}b_{2,k}\frac{\sigma^2}{S}$$
$$\le \mathbb{E}[\Phi(\bar{x}_0)] + O(\frac{\gamma}{\alpha\eta SN}) + O(\frac{\gamma}{\eta S}) , \tag{151}$$

where S is the mini-batch size in initialization.

Additionally, we need to verify the condition in Lemma C.16, $\eta \le \frac{\omega_k^{1/2}}{96\gamma p(3C)^{K-1}L\sqrt{\sum_{j=2}^{K-1}\omega_j(2C^2)^{j-1}}}$ where $\omega_k = \omega_k^h$. To this end, we denote

$$\omega_k^h = a_{2,k}4C^{2(K-k)}L^2 + a_{1,k}4C^{2(K-k)}L^2\frac{1}{N} + 2C^2b_{1,k}\frac{1}{N} + 2C^2b_{2,k}$$
$$+ 72\gamma\eta p^2L^2C^{2(K-1)}K + 36\gamma\eta p^2C^2K(L^2C^{2(K-1)} + C^{2K} + AC^2)$$
$$+ L^2\Big(\sum_{j=1}^{k-1}\omega_j^vC^{2(K-j-1)}3^{k-j}\Big) + 48p^2(3C^2)^{K-1}L^2P_{v_1^c}$$
$$= \frac{\gamma K}{\eta}\Big(\omega_{a_k}(C)4C^{2(K-k)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N} + 2C^2\nu_{b_k}(C,L,\frac{1}{\alpha N})$$
$$+ L^2\sum_{i=1}^{k-1}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{k-i}\Big)$$
$$+ 36\gamma\eta p^2K\Big((2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) + 2C^{2(K-1)}\sum_{i=1}^{k-1}3^{k-i}\Big)$$
$$+ 48p^2\gamma^3\eta K(3C^2)^{K-1}L^2\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})$$
$$\triangleq \frac{\gamma K}{\eta}Q_{1,k} + \gamma\eta p^2KQ_{2,k} + p^2\gamma^3\eta KQ_3 . \tag{152}$$

Then, it is easy to know

$$\eta \le \frac{C_Q(C,L,\frac{1}{\alpha N})}{96\gamma p(3C)^{K-1}L} , \tag{153}$$

where $C_Q(C,L,\frac{1}{\alpha N})$ is a constant with respect to C, L, and $\frac{1}{\alpha N}$.

Finally, by setting

$$\eta \le \min\Bigg\{\frac{\sqrt{C\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\sqrt{D_1(C,L)}}, \frac{\sqrt{C\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma\sqrt{D_2(C,L,\frac{1}{\alpha N})}}, \frac{\sqrt{C^3\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma\sqrt{D_3(C,L)}}, \frac{\sqrt{C^3\nu_{b_k}(C,L,\frac{1}{\alpha N})}}{3\gamma^2\sqrt{D_4(C,L,\frac{1}{\alpha N})}},$$

$$
\frac{\sqrt{C\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\sqrt{E_1(C)}},\frac{\sqrt{C\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma\sqrt{E_2(C,L,\frac{1}{\alpha N})}},\frac{\sqrt{C^3\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma\sqrt{E_3(C,L)}},\frac{\sqrt{C^3\omega_{a_K}(C,\frac{1}{\alpha N})}}{4\gamma^2\sqrt{E_4(C,L,\frac{1}{\alpha N})}},\frac{C_Q(C,L,\frac{1}{\alpha N})}{96\gamma p(3C)^{K-1}L},
$$

$$
\frac{\sqrt{C\omega_{a_k}(C)}}{3\gamma\sqrt{B_1(C,L,\frac{1}{\alpha N})}},\frac{\sqrt{C^3\omega_{a_k}(C)}}{3\gamma\sqrt{B_2(C,L)}},\frac{\sqrt{C^3\omega_{a_k}(C)}}{3\gamma^2\sqrt{B_3(C,L,\frac{1}{\alpha N})}},\frac{\sqrt{C\omega_{a_k}(C)}}{\gamma\sqrt{3\hat{P}_{v_1^c}(C,L,\frac{1}{\alpha N})(3C^2)^{k-1}}}
$$

$$
\left.\frac{\sqrt{\omega_{a_k}(C)}}{\sqrt{6(36\sum_{j=1}^{k-1}C^{2j}(3C^2)^{k-1-j}+\sum_{j=1}^{k-1}C^{2j-1}(3C^2)^{k-1-j})}},\frac{1}{20pC},\frac{1}{6\gamma p(3C^2)^{K-1}},\frac{1}{20\gamma pC^{K-1}L},\frac{1}{2\gamma L_\Phi}\right\},
$$

$$
\gamma\leq\min\left\{\frac{1}{4\sqrt{J_3(C,L,\frac{1}{\alpha N})K}},\frac{1}{2(J_4(C,L,\frac{1}{\alpha N})K)^{1/4}},\frac{1}{(16J_5(C,L,\frac{1}{\alpha N})K)^{1/6}}\right\}, \tag{154}
$$

we have the convergence upper bound as follows:

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla\Phi(\bar{x}_t)\|^2]\leq O(\frac{1}{\alpha\eta^2SNT})+O(\frac{1}{\eta^2ST})\\
&\quad+O(\gamma^2\alpha^2p^2\eta^4)+O(\gamma^2\alpha^2p^4\eta^6)+O(\gamma^4\alpha^2p^4\eta^6)+O(\alpha^2\eta^2)+O(\alpha^2p^2\eta^4)+O(\frac{\alpha\eta^2}{N})\,.
\end{aligned} \tag{155}
$$

□

Lemma C.18. *Given* $\alpha \leq \frac{C}{N}$, $\eta \leq \min\{\frac{\sqrt{\omega_{a_k}(C)}}{\sqrt{6(36\sum_{j=1}^{k-1} C^{2j}(3C^2)^{k-1-j} + \sum_{j=1}^{k-1} C^{2j-1}(3C^2)^{k-1-j})}}, \frac{1}{20pC}\}$*, we have*

$$
\begin{aligned}
& a_{2,1} = \frac{\gamma K}{\eta} \ , \\
& a_{2,k} = \omega_{a_k}(C)\frac{\gamma K}{\eta} \ , k \in \{2, \cdots, K-1\} \ , \\
& a_{2,K} = \omega_{a_K}(C, \frac{1}{\alpha N})\frac{\gamma K}{\eta} \ ,
\end{aligned} \tag{156}
$$

where $\omega_{a_k}(C)$ *is a constant value regarding* C *and* $\omega_{a_K}(C, \frac{1}{\alpha N})$ *is a constant regarding* C *and* $\frac{1}{\alpha N}$*.*

Proof. For $k \in \{2, \cdots, K-1\}$, we consider

$$
\begin{aligned}
\omega_k^v & = 4C^2 a_{2,k} + 4C^2 a_{1,k}\frac{1}{N} + 72p^2\gamma\eta K C^{2k} \\
& = 4C^2 a_{2,k} + \frac{12\gamma K C^{2k}}{\alpha\eta N} + 72p^2\gamma\eta K C^{2k} \ .
\end{aligned} \tag{157}
$$

Then, we have

$$
\begin{aligned}
& 3\alpha\eta^2\Big(\sum_{j=1}^{k-1}\omega_j^v(3C^2)^{k-1-j}\Big) \\
& \leq 3\alpha\eta^2\Big(\sum_{j=1}^{k-1}(4C^2 a_{2,j} + \frac{12\gamma K C^{2j}}{\alpha\eta N} + 72p^2\gamma\eta K C^{2j})(3C^2)^{k-1-j}\Big) \\
& \leq 3\alpha\eta^2\sum_{j=1}^{k-1}4C^2 a_{2,j}(3C^2)^{k-1-j} + 36\gamma\eta\sum_{j=1}^{k-1}\frac{KC^{2j}}{N}(3C^2)^{k-1-j} + 216\gamma\eta\alpha\eta^2p^2K\sum_{j=1}^{k-1}C^{2j}(3C^2)^{k-1-j} \\
& \leq 3\alpha\eta^2\sum_{j=1}^{k-1}4C^2 a_{2,j}(3C^2)^{k-1-j} + 36\gamma\eta K\sum_{j=1}^{k-1}C^{2j}(3C^2)^{k-1-j} + \gamma\eta K\sum_{j=1}^{k-1}C^{2j-1}(3C^2)^{k-1-j} \ ,
\end{aligned} \tag{158}
$$

where the last step follows from $\alpha \leq \frac{C}{N}$ and $\eta \leq \frac{1}{20pC}$.

By setting

$$
a_{2,k} = 24\sum_{j=1}^{k-1}a_{2,j}(3C^2)^{k-j} \ , \tag{159}
$$

we can obtain $3\alpha\eta^2\sum_{j=1}^{k-1}4C^2a_{2,j}(3C^2)^{k-1-j} \leq 3\sum_{j=1}^{k-1}4C^2a_{2,j}(3C^2)^{k-1-j} \leq \frac{1}{6}a_{2,k}$, where the first step holds due to $\alpha\eta^2 \leq 1$.

Based on Eq. (159), we can define $a_{2,1} = \frac{\gamma K}{\eta}$, then we can know

$$
a_{2,k} = \omega_{a_k}(C)\frac{\gamma K}{\eta} \ , \tag{160}
$$

where $\omega_{a_k}(C)$ is a constant value regarding C.

Then, by setting

$$
36\gamma\eta K\sum_{j=1}^{k-1}C^{2j}(3C^2)^{k-1-j} + \gamma\eta K\sum_{j=1}^{k-1}C^{2j-1}(3C^2)^{k-1-j} \leq \frac{1}{6}a_{2,k} = \omega_{a_k}(C)\frac{\gamma K}{\eta} \ , \tag{161}
$$

we can obtain

$$\eta \leq \frac{\sqrt{\omega_{a_k}(C)}}{\sqrt{6(36\sum_{j=1}^{k-1} C^{2j}(3C^2)^{k-1-j} + \sum_{j=1}^{k-1} C^{2j-1}(3C^2)^{k-1-j})}} . \tag{162}$$

Based on these values, we have

$$\begin{aligned}
&3\alpha\eta^2\Big(\sum_{j=1}^{k-1}\omega_j^v(3C^2)^{k-1-j}\Big)\\
&\leq 3\alpha\eta^2\sum_{j=1}^{k-1}4C^2a_{2,j}(3C^2)^{k-1-j} + 36\gamma\eta K\sum_{j=1}^{k-1}C^{2j}(3C^2)^{k-1-j} + \gamma\eta K\sum_{j=1}^{k-1}C^{2j-1}(3C^2)^{k-1-j}\\
&\leq \frac{1}{6}a_{2,k} + \frac{1}{6}a_{2,k} \leq \frac{1}{3}a_{2,k} .
\end{aligned} \tag{163}$$

For $k = K$, we have

$$\begin{aligned}
&2\alpha\eta^2\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 144\gamma\eta\alpha\eta^2p^2C^{2(K-1)}K\\
&\quad + 2\alpha\eta^2\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big) + 32p^2\alpha\eta^2(3C^2)^{K-1}P_{v_1^c} + 128\alpha\gamma^2p^2\eta^4(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\\
&\leq 2\alpha\eta^2\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-1)}}{\alpha N} + \sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\Big)\\
&\quad + 144\alpha\eta^2p^2\gamma\eta K\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} + \gamma\eta C^{2(K-1)-1}K\\
&\quad + \frac{1}{C}(3C^2)^{K-1}\gamma^3\eta K\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) + 128\alpha\gamma^2p^2\eta^4(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)\\
&\leq \frac{\gamma K}{\eta}\Big(8C^2\omega_{a_{K-1}}(C) + \frac{24C^{2(K-1)}}{\alpha N} + 2\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\Big)\\
&\quad + 144\alpha\eta^2p^2\gamma\eta K\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} + \gamma\eta C^{2(K-1)-1}K\\
&\quad + \frac{1}{C}(3C^2)^{K-1}\gamma^3\eta K\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) + 128\alpha\gamma^2p^2\eta^4(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big) ,
\end{aligned} \tag{164}$$

where the second to last step holds due to $\alpha \leq \frac{C}{N}$, $\eta \leq \frac{1}{20pC}$, and the definition of $a_{1,K-1}$ and $a_{2,K-1}$, and we use $\alpha\eta^2 < 1$ in the last step.

Then, we set

$$\begin{aligned}
a_{2,K} &= 100(\frac{C^{2(K-1)}}{\alpha N} + C^2\omega_{a_{K-1}}(C) + \sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1})\frac{\gamma K}{\eta}\\
&\triangleq \omega_{a_K}(C, \frac{1}{\alpha N})\frac{\gamma K}{\eta} ,
\end{aligned} \tag{165}$$

where $\omega_{a_K}(C, \frac{1}{\alpha N})$ is a constant regarding C and $\frac{1}{\alpha N}$. Then, It is easy to know that $\frac{\gamma K}{\eta}\Big(8C^2\omega_{a_{K-1}}(C) + \frac{24C^{2(K-1)}}{\alpha N} + 2\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\Big) \leq \frac{a_{2,K}}{4}$.

□

Lemma C.19. *Given $\alpha \leq \frac{C}{N}$, $\eta \leq \frac{1}{20pC}$, we have*

$$
\begin{aligned}
& b_{2,K-1} = \frac{\gamma K}{\eta} \ , \\
& b_{2,k} = \nu_{b_k}(C, L, \frac{1}{\alpha N})\frac{\gamma K}{\eta} \ , k \in \{1, \cdots, K-2\} \ ,
\end{aligned} \tag{166}
$$

where $\nu_{b_k}(C, L, \frac{1}{\alpha N})$ is a constant value depending on C, L, and $\frac{1}{\alpha N}$.

Proof. For $k \in \{2, \cdots, K-1\}$,

$$
\begin{aligned}
& 4\alpha^2\eta^4\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big)(2C^2)^{K-1-k} + 2\alpha^2\eta^4\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big)(2C^2)^{K-1-k} \\
& \quad + 72\gamma\eta\alpha^2p^2\eta^4L^2KC^{2(K-1)}(2C^2)^{K-1-k} + 288\gamma\eta\alpha^2p^2\eta^4C^{2(K-1)}K(2C^2)^{K-1-k} \\
& \quad + 4\alpha^2\eta^4\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)(2C^2)^{K-1-k} + 64p^2\alpha^2\eta^4(3C^2)^{K-1}P_{v_1^c}(2C^2)^{K-1-k} \\
& \quad + 256\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(2C^2)^{K-1-k} + 4\alpha^2\eta^4\sum_{j=k}^{K-2}\omega_{j+1}^h(2C^2)^{j-k} - \alpha\eta^2b_{2,k} \\
& \leq 2\alpha^2\eta^4\frac{\gamma K}{\eta}\Big(8C^2\omega_{a_{K-1}}(C) + \frac{24C^{2(K-)}}{\alpha N} + \frac{6C^{2(K-1)}}{\alpha N}L^2 + 2\omega_{a_K}(C, \frac{1}{\alpha N})L^2\Big)(2C^2)^{K-1-k} \\
& \quad + 72\gamma\eta\alpha^2p^2\eta^4L^2KC^{2(K-1)}(2C^2)^{K-1-k} + 288\gamma\eta\alpha^2p^2\eta^4C^{2(K-1)}K(2C^2)^{K-1-k} \\
& \quad + 64p^2\alpha^2\eta^4(3C^2)^{K-1}P_{v_1^c}(2C^2)^{K-1-k} + 256\alpha^2\gamma^2p^2\eta^6(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(2C^2)^{K-1-k} \\
& \quad + 4\alpha^2\eta^4(2C^2)^{K-1-k}\frac{\gamma K}{\eta}\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1} \\
& \quad + 4\alpha^2\eta^4 72p^2\gamma\eta K(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} \\
& \quad + 4\alpha^2\eta^4\sum_{j=k}^{K-2}\omega_{j+1}^h(2C^2)^{j-k} - \alpha\eta^2b_{2,k} \ .
\end{aligned} \tag{167}
$$

For ω_k^h, we have

$$
\begin{aligned}
\omega_k^h &= a_{2,k}4C^{2(K-k)}L^2 + a_{1,k}4C^{2(K-k)}L^2\frac{1}{N} + 2C^2b_{1,k}\frac{1}{N} + 2C^2b_{2,k} \\
& \quad + 72\gamma\eta p^2L^2C^{2(K-1)}K + 36\gamma\eta p^2C^2K(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
& \quad + L^2\Big(\sum_{j=1}^{k-1}\omega_j^vC^{2(K-j-1)}3^{k-j}\Big) + 48p^2(3C^2)^{K-1}L^2P_{v_1^c} \\
&= \omega_{a_k}(C)4C^{2(K-k)}L^2\frac{\gamma K}{\eta} + 12C^{2(K-1)}L^2\frac{1}{\alpha N}\frac{\gamma K}{\eta} + 6AC^2\frac{1}{\alpha N}\frac{\gamma K}{\eta} + 2C^2b_{2,k} \\
& \quad + 36\gamma\eta p^2K(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2)) \\
& \quad + L^2\sum_{j=1}^{k-1}\omega_j^vC^{2(K-j-1)}3^{k-j} + 48p^2(3C^2)^{K-1}L^2P_{v_1^c} \ .
\end{aligned} \tag{168}
$$

Therefore, we can get

$$
\begin{aligned}
&4\alpha\eta^2 \sum_{j=k}^{K-2} \omega_{j+1}^h (2C^2)^{j-k} - b_{2,k} \\
&\leq 4\alpha\eta^2 \frac{\gamma K}{\eta} \sum_{j=k}^{K-2} \Big(\omega_{a_{j+1}}(C) 4C^{2(K-j-1)} L^2 + 12C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} \Big) (2C^2)^{j-k} \\
&\quad + 144\alpha\eta^2 p^2 \gamma\eta K (2L^2 C^{2(K-1)} + C^2(L^2 C^{2(K-1)} + C^{2K} + AC^2)) \sum_{j=k}^{K-2} (2C^2)^{j-k} \\
&\quad + 4\alpha\eta^2 \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} \omega_i^v C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} + 192\alpha\eta^2 p^2 (3C^2)^{K-1} L^2 P_{v_1^c} \sum_{j=k}^{K-2} (2C^2)^{j-k} \\
&\quad + 4\alpha\eta^2 \sum_{j=k}^{K-2} 2C^2 b_{2,j+1} (2C^2)^{j-k} - b_{2,k} \ .
\end{aligned} \tag{169}
$$

For $4\alpha\eta^2 \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} \omega_i^v C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k}$, we have

$$
\begin{aligned}
&4\alpha\eta^2 \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} \omega_i^v C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\leq 4\alpha\eta^2 \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} (4C^2 \omega_{a_i}(C) \frac{\gamma K}{\eta} C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\quad + 4\alpha\eta^2 \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} (\frac{12\gamma K C^{2i}}{\alpha\eta N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\quad + 4\alpha\eta^2 \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} (72p^2 \gamma\eta K C^{2i}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\leq 4\alpha\eta^2 \frac{\gamma K}{\eta} \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} (4C^2 \omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\quad + 288\alpha\eta^2 p^2 \gamma\eta K L^2 C^{2(K-1)} \sum_{j=k}^{K-2} \Big(\sum_{i=1}^{j} 3^{j-i} \Big) (2C^2)^{j-k} \ .
\end{aligned} \tag{170}
$$

By putting it into the last inequality, we have

$$
\begin{aligned}
&4\alpha\eta^2 \sum_{j=k}^{K-2} \omega_{j+1}^h (2C^2)^{j-k} - b_{2,k} \\
&\leq 4\alpha\eta^2 \frac{\gamma K}{\eta} \sum_{j=k}^{K-2} \Big(\omega_{a_{j+1}}(C) 4C^{2(K-j-1)} L^2 + 12C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} \Big) (2C^2)^{j-k} \\
&\quad + 144\alpha\eta^2 p^2 \gamma\eta K (2L^2 C^{2(K-1)} + C^2(L^2 C^{2(K-1)} + C^{2K} + AC^2)) \sum_{j=k}^{K-2} (2C^2)^{j-k} \\
&\quad + 192\alpha\eta^2 p^2 (3C^2)^{K-1} L^2 P_{v_1^c} \sum_{j=k}^{K-2} (2C^2)^{j-k}
\end{aligned}
$$

$$
\begin{aligned}
&+ 4\alpha\eta^2 \frac{\gamma K}{\eta} \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} (4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&+ 288\alpha\eta^2 p^2 \gamma\eta K L^2 C^{2(K-1)} \sum_{j=k}^{K-2} \Big(\sum_{i=1}^{j} 3^{j-i} \Big) (2C^2)^{j-k} \\
&+ 4\alpha\eta^2 \sum_{j=k}^{K-2} 2C^2 b_{2,j+1} (2C^2)^{j-k} - b_{2,k} \ .
\end{aligned} \tag{171}
$$

Then, we have

$$
\begin{aligned}
&4\alpha^2\eta^4 \Big(4C^2 a_{2,K-1} + 4C^2 a_{1,K-1} \frac{1}{N} \Big) (2C^2)^{K-1-k} + 2\alpha^2\eta^4 \Big(2a_{1,K} L^2 \frac{1}{N} + 2a_{2,K} L^2 \Big) (2C^2)^{K-1-k} \\
&\quad + 72\gamma\eta\alpha^2 p^2 \eta^4 L^2 K C^{2(K-1)} (2C^2)^{K-1-k} + 288\gamma\eta\alpha^2 p^2 \eta^4 C^{2(K-1)} K (2C^2)^{K-1-k} \\
&\quad + 4\alpha^2\eta^4 \Big(\sum_{j=1}^{K-2} \omega_j^v (3C^2)^{K-j-1} \Big) (2C^2)^{K-1-k} + 64 p^2 \alpha^2 \eta^4 (3C^2)^{K-1} P_{v_1^c} (2C^2)^{K-1-k} \\
&\quad + 256\alpha^2\gamma^2 p^2 \eta^6 (3C^2)^{K-1} \Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1} \Big) (2C^2)^{K-1-k} + 4\alpha^2\eta^4 \sum_{j=k}^{K-2} \omega_{j+1}^h (2C^2)^{j-k} - \alpha\eta^2 b_{2,k} \\
&\le 4\alpha^2\eta^4 \frac{\gamma K}{\eta} \Big(4C^2 \omega_{a_{K-1}}(C) + \frac{12C^{2(K-)}}{\alpha N} + \frac{3C^{2(K-1)}}{\alpha N} L^2 + \omega_{a_K}(C, \frac{1}{\alpha N}) L^2 \Big) (2C^2)^{K-1-k} \\
&\quad + 4\alpha^2\eta^4 (2C^2)^{K-1-k} \frac{\gamma K}{\eta} \sum_{j=1}^{K-2} (4C^2 \omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N}) (3C^2)^{K-j-1} \\
&\quad + 4\alpha^2\eta^4 \frac{\gamma K}{\eta} \sum_{j=k}^{K-2} \Big(\omega_{a_{j+1}}(C) 4C^{2(K-j-1)} L^2 + 12C^{2(K-1)} L^2 \frac{1}{\alpha N} + 6AC^2 \frac{1}{\alpha N} \Big) (2C^2)^{j-k} \\
&\quad + 4\alpha^2\eta^4 \frac{\gamma K}{\eta} \sum_{j=k}^{K-2} \Big(L^2 \sum_{i=1}^{j} (4C^2 \omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N}) C^{2(K-i-1)} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\quad + 72\gamma\eta\alpha^2 p^2 \eta^4 L^2 K C^{2(K-1)} (2C^2)^{K-1-k} + 288\gamma\eta\alpha^2 p^2 \eta^4 C^{2(K-1)} K (2C^2)^{K-1-k} \\
&\quad + 64 p^2 \alpha^2 \eta^4 (3C^2)^{K-1} P_{v_1^c} (2C^2)^{K-1-k} + 256\alpha^2\gamma^2 p^2 \eta^6 (3C^2)^{K-1} \Big(\sum_{j=2}^{K-1} \omega_j^h (2C^2)^{j-1} \Big) (2C^2)^{K-1-k} \\
&\quad + 4\alpha^2\eta^4 72 p^2 \gamma\eta K (2C^2)^{K-1-k} \sum_{j=1}^{K-2} C^{2j} (3C^2)^{K-j-1} \\
&\quad + 144\alpha^2\eta^4 p^2 \gamma\eta K (2L^2 C^{2(K-1)} + C^2 (L^2 C^{2(K-1)} + C^{2K} + AC^2)) \sum_{j=k}^{K-2} (2C^2)^{j-k} \\
&\quad + 192\alpha^2\eta^4 p^2 (3C^2)^{K-1} L^2 P_{v_1^c} \sum_{j=k}^{K-2} (2C^2)^{j-k} + 288\alpha^2\eta^4 p^2 \gamma\eta K L^2 C^{2(K-1)} \sum_{j=k}^{K-2} \Big(\sum_{i=1}^{j} 3^{j-i} \Big) (2C^2)^{j-k} \\
&\quad + 4\alpha^2\eta^4 \sum_{j=k}^{K-2} 2C^2 b_{2,j+1} (2C^2)^{j-k} - \alpha\eta^2 b_{2,k} \\
&\le 4\alpha^2\eta^4 \frac{\gamma K}{\eta} \Big(4C^2 \omega_{a_{K-1}}(C) + \frac{12C^{2(K-)}}{\alpha N} + \frac{3C^{2(K-1)}}{\alpha N} L^2 + \omega_{a_K}(C, \frac{1}{\alpha N}) L^2 \Big) (2C^2)^{K-1-k} \\
&\quad + 4\alpha^2\eta^4 (2C^2)^{K-1-k} \frac{\gamma K}{\eta} \sum_{j=1}^{K-2} (4C^2 \omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N}) (3C^2)^{K-j-1}
\end{aligned}
$$

$$
\begin{aligned}
&+ 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(\omega_{a_{j+1}}(C)4C^{2(K-j-1)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N}\Big)(2C^2)^{j-k}\\
&+ 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(L^2\sum_{i=1}^{j}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-k}\\
&+ 72\alpha^2p^2\eta^4\gamma\eta K\Big((L^2+4)C^{2(K-1)}(2C^2)^{K-1-k} + 4(2C^2)^{K-1-k}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1}\\
&\quad + 2(2L^2C^{2(K-1)} + C^2(L^2C^{2(K-1)} + C^{2K} + AC^2))\sum_{j=k}^{K-2}(2C^2)^{j-k} + 4L^2C^{2(K-1)}\sum_{j=k}^{K-2}\Big(\sum_{i=1}^{j}3^{j-i}\Big)(2C^2)^{j-k}\Big)\\
&+ 64p^2\alpha^2\eta^4(3C^2)^{K-1}P_{v_1^c}((2C^2)^{K-1-k} + 3L^2\sum_{j=k}^{K-2}(2C^2)^{j-k})\\
&+ 256\alpha^2p^2\eta^6\gamma^2(3C^2)^{K-1}\Big(\sum_{j=2}^{K-1}\omega_j^h(2C^2)^{j-1}\Big)(2C^2)^{K-1-k}\\
&+ 4\alpha^2\eta^4\sum_{j=k}^{K-2}2C^2b_{2,j+1}(2C^2)^{j-k} - \alpha\eta^2 b_{2,k}\ ,
\end{aligned}
\tag{172}
$$

When $k = K-1$, we define $b_{2,K-1} = \frac{\gamma K}{\eta}$. Then, for $k \in \{1, \cdots, K-2\}$, we define

$$
\begin{aligned}
b_{2,k} &= 8\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-)}}{\alpha N} + \frac{3C^{2(K-1)}}{\alpha N}L^2 + \omega_{a_K}(C, \frac{1}{\alpha N})L^2\Big)(2C^2)^{K-1-k}\\
&+ 8\alpha\eta^2(2C^2)^{K-1-k}\frac{\gamma K}{\eta}\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\\
&+ 8\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(\omega_{a_{j+1}}(C)4C^{2(K-j-1)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N}\Big)(2C^2)^{j-k}\\
&+ 8\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(L^2\sum_{i=1}^{j}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-k}\\
&+ 8\sum_{j=k}^{K-2}2C^2b_{2,j+1}(2C^2)^{j-k}\\
&\triangleq \nu_{b_k}(C, L, \frac{1}{\alpha N})\frac{\gamma K}{\eta}\ ,
\end{aligned}
\tag{173}
$$

where $\nu_{b_k}(C, L, \frac{1}{\alpha N})$ is a constant value depending on C, L, and $\frac{1}{\alpha N}$. As a result, we have

$$
\begin{aligned}
&4\alpha^2\eta^4\frac{\gamma K}{\eta}\Big(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-)}}{\alpha N} + \frac{3C^{2(K-1)}}{\alpha N}L^2 + \omega_{a_K}(C, \frac{1}{\alpha N})L^2\Big)(2C^2)^{K-1-k}\\
&+ 4\alpha^2\eta^4(2C^2)^{K-1-k}\frac{\gamma K}{\eta}\sum_{j=1}^{K-2}(4C^2\omega_{a_j}(C) + \frac{12C^{2j}}{\alpha N})(3C^2)^{K-j-1}\\
&+ 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(\omega_{a_{j+1}}(C)4C^{2(K-j-1)}L^2 + 12C^{2(K-1)}L^2\frac{1}{\alpha N} + 6AC^2\frac{1}{\alpha N}\Big)(2C^2)^{j-k}\\
&+ 4\alpha^2\eta^4\frac{\gamma K}{\eta}\sum_{j=k}^{K-2}\Big(L^2\sum_{i=1}^{j}(4C^2\omega_{a_i}(C) + \frac{12C^{2i}}{\alpha N})C^{2(K-i-1)}3^{j-i}\Big)(2C^2)^{j-k}
\end{aligned}
$$

$$+ 4\alpha^2\eta^4 \sum_{j=k}^{K-2} \omega_{j+1}^h (2C^2)^{j-k} \le \frac{1}{2}\alpha\eta^2 b_{2,k} \ . \tag{174}$$

□

Lemma C.20. *Given $\alpha \le \frac{C}{N}$, $\eta \le \frac{1}{20pC}$, we have*

$$P_{v_1^c} \le \eta\gamma^3 K \hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) \ , \tag{175}$$

where $\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})$ is a constant with respect to C, L, and $\frac{1}{\alpha N}$.

Proof.

$$\begin{aligned}
P_{v_1^c} = \Big[& 3\gamma^3\eta^3p^2K(L^2C^{2(K-1)} + AC^2) + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) \\
& + 2\gamma^2\eta^2\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2b_{1,1}\frac{1}{N} + 2C^2b_{2,1}\Big) \\
& + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big) + 72\gamma\eta\gamma^2p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} \\
& + 144\gamma\eta\gamma^2p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K + 3\gamma\eta K(24p^2\gamma^2\eta^2C^2 + 120\gamma^2\alpha^2p^4\eta^6C^2)(L^2C^{2(K-1)} + C^{2K} + AC^2) \\
& + 2\gamma^2\eta^2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\Big] \\
= \gamma\eta\gamma^2 & \Big(3\eta^2p^2K(L^2C^{2(K-1)} + AC^2) + 72p^2\eta^2(2C^2)^{K-1}L^2KC^{2(K-1)} + 144p^2\eta^2(2C^2)^{K-1}C^{2(K-1)}K \\
& + 3K(24p^2\eta^2C^2 + 120\alpha^2p^4\eta^6C^2)(L^2C^{2(K-1)} + C^{2K} + AC^2)\Big) \\
& + \gamma^2\eta^2\Big[2(2C^2)^{K-1}\Big(4C^2a_{2,K-1} + 4C^2a_{1,K-1}\frac{1}{N}\Big) + 2(2C^2)^{K-1}\Big(2a_{1,K}L^2\frac{1}{N} + 2a_{2,K}L^2\Big) \\
& + 2\Big(a_{2,1}4C^{2(K-1)}L^2 + a_{1,1}4C^{2(K-1)}L^2\frac{1}{N} + 2C^2b_{1,1}\frac{1}{N} + 2C^2b_{2,1}\Big) + 2(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}\omega_j^v(3C^2)^{K-j-1}\Big)\Big] \\
\le \gamma\eta\gamma^2K & \Big(3\eta^2p^2(L^2C^{2(K-1)} + AC^2) + 72p^2\eta^2(2C^2)^{K-1}L^2C^{2(K-1)} + 144p^2\eta^2(2C^2)^{K-1}C^{2(K-1)} \\
& + 3(24p^2\eta^2C^2 + 120\alpha^2p^4\eta^6C^2)(L^2C^{2(K-1)} + C^{2K} + AC^2)\Big) \\
& + \gamma\eta\gamma\Big[2(2C^2)^{K-1}\Big(4\eta C^2\omega_{a_{K-1}}(C)\frac{\gamma K}{\eta} + \frac{12\gamma KC^{2(K-2)}}{\alpha N}\Big) \\
& + 2(2C^2)^{K-1}\Big(\frac{6\gamma KC^{2(K-1)}}{\alpha N}L^2 + 2\eta L^2\omega_{a_K}(C, \frac{1}{\alpha N})\frac{\gamma K}{\eta}\Big) \\
& + 2\Big(\frac{\gamma K}{\eta}4\eta C^{2(K-1)}L^2 + \frac{12\gamma KC^{2(K-1)}L^2}{\alpha N} + \frac{6\gamma AKC^2}{\alpha N} + 2\eta C^2\nu_{b_1}(C, L, \frac{1}{\alpha N})\frac{\gamma K}{\eta}\Big)\Big] \\
& + 2\gamma\eta\gamma(2C^2)^{K-1}\Big(\sum_{j=1}^{K-2}4\eta C^2a_{2,j}(3C^2)^{K-j-1} + \sum_{j=1}^{K-2}\frac{12\gamma KC^{2j}}{\alpha N}(3C^2)^{K-j-1} + \sum_{j=1}^{K-2}72\gamma p^2\eta^2KC^{2j}(3C^2)^{K-j-1}\Big) \\
\le \gamma\eta\gamma^2K & \Big((L^2C^{2(K-2)} + A) + (2C^2)^{K-1}L^2C^{2(K-2)} + (2C^2)^{K-1}C^{2(K-2)} \\
& + 3(1 + C^{-2})(L^2C^{2(K-1)} + C^{2K} + AC^2)\Big) \\
& + \gamma\eta\gamma^2K\Big(2(2C^2)^{K-1}(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-2)}}{\alpha N} + \frac{6C^{2(K-1)}}{\alpha N}L^2 + 2L^2\omega_{a_K}(C, \frac{1}{\alpha N}))
\end{aligned}$$

$$
\begin{aligned}
&+ 8C^{2(K-1)}L^2 + \frac{24C^{2(K-1)}L^2}{\alpha N} + \frac{12AC^2}{\alpha N}\Big) + 4\gamma\eta\gamma^2 KC^2\nu_{b_1}(C, L, \frac{1}{\alpha N}) \\
&+ 2\gamma\eta\gamma^2 K(2C^2)^{K-1}\Big(2\sum_{j=1}^{K-2}\omega_{a_j}(C)(3C^2)^{K-j} + \frac{12}{\alpha N}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} + \sum_{j=1}^{K-2}C^{2j-2}(3C^2)^{K-j-1}\Big) \\
&\triangleq \eta\gamma^3 K\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) \,, \qquad (176)
\end{aligned}
$$

where the last step holds due to $\alpha \leq \frac{C}{N}$, $\eta \leq \frac{1}{20pC}$, and $\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N})$ is a constant value as follows:

$$
\begin{aligned}
\hat{P}_{v_1^c}(C, L, \frac{1}{\alpha N}) &= \Big((L^2C^{2(K-2)} + A) + (2C^2)^{K-1}L^2C^{2(K-2)} + (2C^2)^{K-1}C^{2(K-2)} \\
&+ 3(1 + C^{-2})(L^2C^{2(K-1)} + C^{2K} + AC^2)\Big) \\
&+ \Big(2(2C^2)^{K-1}(4C^2\omega_{a_{K-1}}(C) + \frac{12C^{2(K-2)}}{\alpha N} + \frac{6C^{2(K-1)}}{\alpha N}L^2 + 2L^2\omega_{a_K}(C, \frac{1}{\alpha N})) \\
&\quad + 8C^{2(K-1)}L^2 + \frac{24C^{2(K-1)}L^2}{\alpha N} + \frac{12AC^2}{\alpha N} + 4C^2\nu_{b_1}(C, L, \frac{1}{\alpha N})\Big) \\
&+ 2(2C^2)^{K-1}\Big(2\sum_{j=1}^{K-2}\omega_{a_j}(C)(3C^2)^{K-j} + \frac{12}{\alpha N}\sum_{j=1}^{K-2}C^{2j}(3C^2)^{K-j-1} + \sum_{j=1}^{K-2}C^{2j-2}(3C^2)^{K-j-1}\Big) \,. \qquad (177)
\end{aligned}
$$

□

Lemma C.21. *The linear operator:* $\Pi_{C_{\Psi^{(k)}}}(v^{(k)}) = \frac{1}{3}v^{(k)}$*, can be used for the projection step in Algorithm 1.*

Proof. At the first iteration, it is easy to know that

$$
\|v_{n,0}^{(k)}\| \leq C_{\Psi^{(k)}} \,. \qquad (178)
$$

Then, in the t-th iteration, we have already known $\|v_{n,t}^{(k)}\| \leq C_{\Psi^{(k)}}$. As a result, we have

$$
\begin{aligned}
&\|(1-\alpha\eta^2)(v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t}^{(k+1)}) + \nabla F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t+1}^{(k+1)}\| \\
&\leq \|v_{n,t}^{(k)}\| + \|\nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t}^{(k+1)}\| + \|\nabla F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t+1}^{(k+1)}\| \\
&\leq 3C_{\Psi^{(k)}} \,. \qquad (179)
\end{aligned}
$$

Then, we can set the projection operator as a linear operator: $\Pi_{C_{\Psi^{(k)}}}(v) = \frac{1}{3}v$. It is easy to verify that

$$
\begin{aligned}
&\|v_{n,t+1}^{(k)}\| \\
&= \|\Pi_{C_{\Psi^{(k)}}}((1-\alpha\eta^2)(v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t}^{(k+1)}) + \nabla F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t+1}^{(k+1)})\| \\
&= \|\frac{1}{3}((1-\alpha\eta^2)(v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t}^{(k+1)}) + \nabla F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t+1}^{(k+1)})\| \\
&= \frac{1}{3}\|(1-\alpha\eta^2)(v_{n,t}^{(k)} - \nabla F_n^{(k)}(h_{n,t}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t}^{(k+1)}) + \nabla F_n^{(k)}(h_{n,t+1}^{(k-1)}; \xi_{n,t+1}^{(k)})^T v_{n,t+1}^{(k+1)}\| \\
&\leq C_{\Psi^{(k)}} \,. \qquad (180)
\end{aligned}
$$

□