A Single-Timescale Analysis for Stochastic Approximation with Multiple Coupled Sequences

Han Shen
Rensselaer Polytechnic Institute
Troy, NY 12180, United States
shenh5@rpi.edu

Tianyi Chen
Rensselaer Polytechnic Institute
Troy, NY 12180, United States
chentianyi19@gmail.com

Abstract

Stochastic approximation (SA) with multiple coupled sequences has found broad applications in machine learning such as bilevel learning and reinforcement learning (RL). In this paper, we study the finite-time convergence of nonlinear SA with multiple coupled sequences. Different from existing multi-timescale analysis, we seek for scenarios where a fine-grained analysis can provide the tight performance guarantee for single-timescale multi-sequence SA (STSA). At the heart of our analysis is the smoothness property of the fixed points in multi-sequence SA that holds in many applications. When all sequences have strongly monotone increments, we establish the iteration complexity of $\mathcal{O}(\epsilon^{-1})$ to achieve ϵ-accuracy, which improves the existing $\mathcal{O}(\epsilon^{-1.5})$ complexity for two coupled sequences. When the main sequence does not have strongly monotone increment, we establish the iteration complexity of $\mathcal{O}(\epsilon^{-2})$. We showcase the power of our result by applying it to stochastic bilevel and compositional optimization problems, as well as RL problems, all of which lead to improvements over their existing guarantees.

1 Introduction

Stochastic approximation (SA) is an iterative procedure used to find the zero of a function when only the noisy estimate of the function is observed. Specifically, with the mapping $v : \mathbb{R}^d \mapsto \mathbb{R}^d$, the single-sequence SA seeks to solve for $v(x) = 0$ with the following iterative update:

$$x_{k+1} = x_k + \alpha_k(v(x_k) + \xi_k), \tag{1}$$

where α_k is the step size and ξ_k is a random variable. Since its introduction in [46], single-sequence SA has received great interests because of its broad range of applications to areas including stochastic optimization and reinforcement learning (RL) [6, 53]. The asymptotic convergence of single-sequence SA can be established by the ordinary differential equation method; see e.g., [4]. To gain more insights into the performance difference of various stochastic optimization algorithms, the finite-time convergence of SA has been widely studied in recent years; see e.g., [43, 42, 30, 50, 54, 52, 41, 13].

While most of the SA studies focus on the single-sequence case, the double-sequence SA was introduced in [3], which has been extensively applied to the RL methods involving a double-sequence stochastic update structure [53, 32, 10]. With mappings $v : \mathbb{R}^{d_0} \times \mathbb{R}^{d_1} \mapsto \mathbb{R}^{d_0}$ and $h : \mathbb{R}^{d_0} \times \mathbb{R}^{d_1} \mapsto \mathbb{R}^{d_1}$, the double-sequence SA seeks to solve $v(x, y) = h(x, y) = 0$ with the following update:

$$x_{k+1} = x_k + \alpha_k(v(x_k, y_k) + \xi_k), \tag{2a}$$
$$y_{k+1} = y_k + \beta_k(h(x_k, y_k) + \psi_k), \tag{2b}$$

where α_k, β_k are the step sizes, and ξ_k, ψ_k are random variables. In (2), the update of x_k and that of y_k depend on each other and thus the sequences are *coupled*. To deal with the coupling, a naive thought is to stack (x_k, y_k) as one variable. However, it can be seen later that the convergence of

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

	General result		Application to SBO				Application to multi-level SCO		
	Ours	TTS SA	Ours	TTSA	ALSET	ALSET-AC	Ours	α-TSCGD	SG-MRL
SM	$\mathcal{O}(\epsilon^{-1})$	$\mathcal{O}(\epsilon^{-1.5})$	$\tilde{\mathcal{O}}(\epsilon^{-1})$	$\tilde{\mathcal{O}}(\epsilon^{-1.5})$	~	~	$\mathcal{O}(\epsilon^{-1})$	$\mathcal{O}(\epsilon^{-\frac{N+5}{4}})$	~
N-SM	$\mathcal{O}(\epsilon^{-2})$	~	$\tilde{\mathcal{O}}(\epsilon^{-2})$	$\tilde{\mathcal{O}}(\epsilon^{-2.5})$	$\tilde{\mathcal{O}}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-2})$	$\mathcal{O}(\epsilon^{-\frac{N+8}{4}})$	$\mathcal{O}(\epsilon^{-4})$
Merit	~	Rate ↑	~	Rate ↑	Relax	Relax	~	Rate ↑	Rate ↑

Table 1: Comparisons with TTS SA [12], TTSA [26], ALSET and ALSET-AC [8], α-TSCGD [61] and SG-MRL [14]. Strongly-monotone (SM) and non-strongly-monotone (N-SM) respectively represents the case where the main sequence has strongly-monotone and non-strongly-monotone increments. Rows of SM/N-SM are for the complexity and the row of Merit is for the improvements of this work over the existing work ("Rate ↑" stands for faster rate; "Relax" for relaxed assumptions).

the resulting update requires assumptions violated in the applications. Thus due to the coupling, the double-sequence SA is more challenging to analyze than its single-sequence counterpart.

Prior art on double-sequence SA. Many recent analyses of the double-sequence SA focus on the linear case where $v(x, y)$ and $h(x, y)$ are linear mappings; see e.g., [34, 11, 25, 29]. The key idea here is to use the so-called two-time-scale (TTS) step sizes: *One sequence is updated in the faster time scale while the other is updated in the slower time scale; that is* $\lim_{k\to\infty} \alpha_k/\beta_k = 0$. By doing so, the two sequences are shown to decouple asymptotically, which allows us to leverage the analysis of the single-sequence SA. In particular, [29] proves an iteration complexity of $\mathcal{O}(\epsilon^{-1})$ to achieve ϵ-accuracy for the TTS linear SA, which is shown to be tight. With similar choice of the step sizes, the TTS nonlinear SA was analyzed in [39, 12]. In [39], the finite-time convergence rate of TTS nonlinear SA was established under an assumption that the two sequences converge asymptotically. Later, [12] alleviates this assumption and shows that TTS nonlinear SA achieves an iteration complexity of $\mathcal{O}(\epsilon^{-1.5})$. However, this iteration complexity is larger than $\mathcal{O}(\epsilon^{-1})$ of the TTS linear SA.

The gap between the complexities of nonlinear and linear SA motivates an interesting question:

Q1: Is it possible to prove a faster rate for the nonlinear SA with two coupled sequences?

We first conduct an experiment to examine the possibility.

Experiment. Figure 1 shows the performance of using the double-sequence SA (2) to solve the following problem

$$\max_{x\in\mathbb{R}} \quad -\frac{1}{2}\Big(x^2 + \frac{1}{1+e^{-y^*(x)}}\Big) \qquad \text{s.t.} \;\; y^*(x) = \arg\min_{y\in\mathbb{R}} \frac{1}{2}(y-x)^2. \tag{3}$$

We use the double-sequence SA (2) to solve (3), where

$$v(x, y) = -x - \frac{e^{-y}}{(1+e^{-y})^2}, \quad h(x, y) = x - y \tag{4}$$

and ζ_k, ξ_k are independent Gaussian random variables with zero mean and standard deviations of 0.15. It is easy to check that (4) satisfies the assumptions in the existing TTS-SA analysis [12]. Therefore, we can use the two time-scale step sizes and achieve the iteration complexity of $\mathcal{O}(\epsilon^{-1.5})$. However, as suggested by Figure 1, the iterates still converge with step sizes in a single time-scale ($\alpha_k = \Theta(\frac{1}{k})$, $\beta_k = \Theta(\frac{1}{k})$). In this case, the iteration complexity is $\mathcal{O}(\epsilon^{-1})$, which is the same as that of double-sequence linear SA [29]. This suggests that existing analysis of double-sequence SA might not be tight, at least for the class of updates similar to (4). Indeed, as we will show later, the iterates generated by (4) will converge with the iteration complexity of $\mathcal{O}(\epsilon^{-1})$.

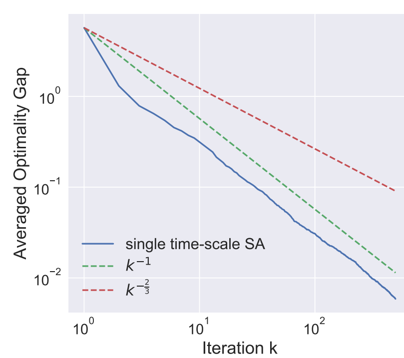

Figure 1: Solving (3) with double-sequence nonlinear SA (2). The single time-scale nonlinear SA converges with a rate of $\mathcal{O}(k^{-1})$, which is faster than the theoretical $\mathcal{O}(k^{-\frac{2}{3}})$ rate in [12].

Furthermore, existing works on TTS SA mainly focus on the double-sequence case. While in cases such as the multi-level stochastic optimization; see e.g., [61], more than two sequences are involved. This necessitates the use of the multi-sequence SA. Specifically, with mappings $v : \mathbb{R}^{d_0} \times \mathbb{R}^{d_1} \cdots \times \mathbb{R}^{d_N} \mapsto \mathbb{R}^{d_0}$, $h^n : \mathbb{R}^{d_{n-1}} \times \mathbb{R}^{d_n} \mapsto \mathbb{R}^{d_n}$, we consider

$$
(\textbf{STSA}) \qquad y_{k+1}^n = y_k^n + \beta_{k,n}\big(h^n(y_k^{n-1}, y_k^n) + \psi_k^n\big), \;\; n = 1, 2, ..., N \tag{5a}
$$
$$
x_{k+1} = x_k + \alpha_k\big(v(x_k, y_k^1, y_k^2, \dots, y_k^N) + \xi_k\big) \tag{5b}
$$

where $\alpha_k, \beta_{k,1}, ..., \beta_{k,N}$ are the step sizes, and $\xi_k, \psi_k^1, \dots, \psi_k^N$ are random variables. For conciseness, we have used $y_k^0 := x_k$ here. Our goal is to find the unique fixed-points $x^*, y^{1,*}, \dots, y^{N,*}$ such that

$$
v(x^*, y^{1,*}, \dots, y^{N,*}) = 0, \; h^1(x^*, y^{1,*}) = 0, \dots, \; h^N(y^{N-1,*}, y^{N,*}) = 0. \tag{6}
$$

Observing that in (5), for every n, the sequence of y_k^n is coupled with that of y_k^{n-1} and is ultimately coupled with the main sequence x_k. Meanwhile the update of x_k also depends on $\{y_k^n\}_{n=1}^N$. Since all sequences in (5) are coupled, (5) is more challenging to analyze than the double-sequence SA.

Prior art related to multi-sequence SA. The multilevel stochastic optimization problem [49] and the multilevel SCO problem [1, 58, 63, 47, 65] are closely related to the multi-sequence SA. To tackle the multi-level structure, these recent methods have modified the vanilla multi-sequence SA update to achieve the state-of-the-art complexity and thus their updates are no longer in the form of (5). In contrast, we focus on the multi-sequence SA update in (5). To the best of our knowledge, the only analysis for (5) is [61] where the TTS technique is generalized to multi-time-scale. In [61], the iteration complexity will get worse as the number of sequences N increases.

This gives rise to another interesting question:

Q2: Is it possible to establish convergence rate independent of the number of sequences?

In this work, we give affirmative answers to both Questions **Q1** and **Q2**.

Our contributions. Specially, by exploiting the *smooth assumption* that can be satisfied in many applications, we show that the vanilla nonlinear SA can run in a *single time scale*! We further prove that the order of the convergence rate is *independent of* the number of sequences N! Intuitively, this is possible because when the fixed point $y^{n,*}$ is smooth in x, the y_k^n-update *converges fast enough* such that its fixed-point residual after one-step update is at the same order as the drift of $y^{n,*}$.

In the context of prior art, our contributions can be summarized as follows (see Table 1).

C1) Single-timescale analysis for multi-sequence SA. Different from existing two-timescale analysis [39, 5], we establish a unifying Single-Timescale analysis for SA with multiple coupled sequences that we term **STSA**. When all the sequences have strongly-monotone increments, we improve the $\mathcal{O}(\epsilon^{-1.5})$ iteration complexity for multi-sequence TTS-SA in [12] to $\mathcal{O}(\epsilon^{-1})$. When the main sequence does not have the strongly-monotone increment, we provide the $\mathcal{O}(\epsilon^{-2})$ iteration complexity. It is worth noting that though the single time-scale step sizes were also explored in [37, 44], the key enabler in those works is the decrease of variance which is a result of variance-reduction update or increasing batch size. While this work and those in Table 1 focus on the case where the variance is non-decreasing.

C2) STSA for stochastic bilevel optimization (SBO). When applying our generic results to the SBO problem with double-sequence SA, for strongly-concave objective functions, we improve the best-known sample complexity $\tilde{\mathcal{O}}(\epsilon^{-1.5})$ of TTSA in [26] to $\tilde{\mathcal{O}}(\epsilon^{-1})$. For the non-concave objective function, we achieve the same sample complexity $\mathcal{O}(\epsilon^{-2})$ of ALSET while relaxing the bounded upper-level gradient assumption made in [8].

C3) STSA for stochastic compositional optimization (SCO). When applying our results to the multi-level SCO problems, we improve the level-dependent sample complexities $\mathcal{O}(\epsilon^{-\frac{N+5}{4}})$ and $\mathcal{O}(\epsilon^{-\frac{N+8}{4}})$ of multi-sequence SA based α-TSCGD method in [61] to the level-independent complexities $\tilde{\mathcal{O}}(\epsilon^{-1})$ and $\mathcal{O}(\epsilon^{-2})$, under the strongly-concave and non-concave objective functions, respectively.

C4) STSA for policy optimization in RL problems. Moreover, applying our results to the actor-critic method achieves the same $\mathcal{O}(\epsilon^{-2})$ sample complexity of ALSET-AC in [8] while relaxing the unverifiable assumption on the stationary distribution of Markov chains; applying our results to the meta policy gradient improves the $\mathcal{O}(\epsilon^{-4})$ sample complexity of SG-MRL in [14] to $\mathcal{O}(\epsilon^{-2})$.

2 Main Results: Convergence of Single-timescale Multi-sequence SA

Before introducing the main results, we will first make some standard assumptions. Throughout the discussion, we define $[N] \coloneqq \{1, 2, ..., N\}$, $[K] \coloneqq \{1, 2, ..., K\}$ and $y^0 \coloneqq x$ for conciseness.

Assumption 1 (Smoothness of the fixed points) *For any $n \in [N]$ and $y^{n-1} \in \mathbb{R}^{d_{n-1}}$, there exists a unique $y^{n,*}(y^{n-1}) \in \mathbb{R}^{d_n}$ such that $h^n(y^{n-1}, y^{n,*}(y^{n-1})) = 0$. Moreover, there exist constants $L_{y,n}$ and $L_{y',n}$ such that for any $y^{n-1}, \bar{y}^{n-1} \in \mathbb{R}^{d_{n-1}}$, the following inequalities hold*

$$\|y^{n,*}(y^{n-1}) - y^{n,*}(\bar{y}^{n-1})\| \leq L_{y,n}\|y^{n-1} - \bar{y}^{n-1}\|, \tag{7a}$$

$$\|\nabla y^{n,*}(y^{n-1}) - \nabla y^{n,*}(\bar{y}^{n-1})\| \leq L_{y',n}\|y^{n-1} - \bar{y}^{n-1}\|. \tag{7b}$$

Due to the change of y_k^{n-1} at each iteration, the solution of $h^n(y_k^{n-1}, y^n) = 0$ with respect to (w.r.t.) y^n, that is $y^{n,*}(y_k^{n-1})$, is drifting over consecutive iterations. Given y_k^{n-1}, since only one-step of y_k^n update is performed at each iteration, one can only hope to establish convergence of y_k^n if the drift of its optimal solution is controlled in some sense. Assumption 1 ensures both the zeroth-order and first-order drifts are controlled in the same scale of the change of y_k^{n-1}. This assumption is satisfied in linear SA [29] and other applications which will be shown later.

Define $v(x) \coloneqq v\big(x, y^{1,*}(x), y^{2,*}(y^{1,*}(x)), \ldots, y^{N,*}(\ldots y^{2,*}(y^{1,*}(x))\ldots)\big)$. With $y^{1:N}$ as a concise notation for $(y^1, ..., y^N)$, we make the following assumption.

Assumption 2 (Lipschitz continuity of increments) *For any $n \in [N]$, $x, \bar{x} \in \mathbb{R}^{d_0}$ and $y^n, \bar{y}^n \in \mathbb{R}^{d_n}$, there exist constants L_v, $L_{v,y}$ and $L_{h,n}$ such that the following inequalities hold*

$$\|v(x) - v(\bar{x})\| \leq L_v\|x - \bar{x}\|, \quad \|v(x, y^{1:N}) - v(x, \bar{y}^{1:N})\| \leq L_{v,y}\sum_{n=1}^{N}\|y^n - \bar{y}^n\|, \tag{8a}$$

$$\|h^n(y^{n-1}, y^n) - h^n(y^{n-1}, \bar{y}^n)\| \leq L_{h,n}\|y^n - \bar{y}^n\|. \tag{8b}$$

Define $\mathcal{F}_k$ as the σ-algebra generated by the random variables in $\{x_i, y_i^{1:N}\}_{i=1}^k$ and $\mathcal{F}_k^n$ as the σ-algebra generated by $\{x_i, y_i^{1:N}\}_{i=1}^k \cup \{y_{k+1}^n\}$. We make the following assumption on the noises.

Assumption 3 (Bias and variance) *There exist constants $\{c_n, \sigma_n\}_{n=0}^N$ such that $\forall k, n$, $\|\mathbb{E}[\xi_k|\mathcal{F}_k^1]\|^2 \leq c_0^2\alpha_k$ and $\|\mathbb{E}[\psi_k^n|\mathcal{F}_k^{n+1}]\|^2 \leq c_n^2\beta_{k,n}$; $\mathbb{E}[\|\xi_k\|^2|\mathcal{F}_k^1] \leq \sigma_0^2$ and $\mathbb{E}[\|\psi_k^n\||\mathcal{F}_k^{n+1}] \leq \sigma_n^2$.*

Here we define $\mathcal{F}_k^{N+1} \coloneqq \mathcal{F}_k$. Assumption 3 is a generalized version of the bias and variance assumption in stochastic programming [19] or the noise assumption in single-sequence SA [30] to multi-sequence case. Similar assumption has also been made in the double-sequence SA [26]. As will be shown later, when applying STSA to the stochastic optimization problems, the conditional independence between samples of different levels given $\mathcal{F}_k$ along with the small bias and bounded variance condition will guarantee this assumption.

Assumption 4 (Monotonicity of h) *For $n \in [N]$, $h^n(y^{n-1}, y^n)$ is one-point strongly monotone on $y^{n,*}(y^{n-1})$ given any y^{n-1}; that is, there exists constant $\lambda_n > 0$ such that (cf. $h^n(y^{n-1}, y^{n,*}) = 0$)*

$$\big\langle y^n - y^{n,*}(y^{n-1}), h^n(y^{n-1}, y^n)\big\rangle \leq -\lambda_n\|y^n - y^{n,*}(y^{n-1})\|^2, \ \forall y^n \in \mathbb{R}^{d_n}. \tag{9}$$

Assumption 4 is implied by the standard regularity assumptions in the previous works on TTS linear SA [34, 29], and has also been exploited in the TTS nonlinear SA works; see e.g. [39, 12].

2.1 The strongly-monotone case

We first consider the case when the main sequence x_k has strongly-monotone increment.

Assumption 5 (Monotonicity of v) *Suppose $v(x)$ is one-point strongly monotone on x^*; that is, there exists a positive constant λ_0 such that (cf. $v(x^*) = 0$)*

$$\big\langle x - x^*, v(x)\big\rangle \leq -\lambda_0\|x - x^*\|^2, \ \forall x \in \mathbb{R}^{d_0}. \tag{10}$$

Same as Assumption 4, Assumption 5 is standard in the previous works on TTS SA [39, 12]. This assumption is a regularity assumption in the case of TTS linear SA; see e.g., [34, Assumption 2.3]. Or in the case of bilevel optimization which will be discussed later, this assumption is satisfied when the objective function is strongly-concave.

Due to space limitation, we directly present the result below and defer the proof to Appendix B.

Theorem 1 *Consider the sequences generated by* (5)*. Suppose Assumptions 1–5 hold. Select step sizes $\alpha_k = \Theta(\frac{1}{k})$ and $\beta_{k,n} = \Theta(\frac{1}{k})$. It holds for any k that*

$$\mathbb{E}\|x_k - x^*\|^2 + \sum_{n=1}^{N} \mathbb{E}\|y_k^n - y^{n,*}(y_k^{n-1})\|^2 = \mathcal{O}\Big(\frac{1}{k}\Big) \tag{11}$$

where $\mathcal{O}(\cdot)$ hides constants in the polynomial of N, and we have used $y_k^0 = x_k$ for convenience. Moreover, for any $n \in [N]$ we have

$$\lim_{k\to\infty} \|x_k - x^*\|^2 = 0 \ \ \textit{almost surely (a.s.)}, \quad \lim_{k\to\infty} \|y_k^n - y^{n,*}(y_k^{n-1})\|^2 = 0 \ \ a.s. \tag{12}$$

It is worth noting that with (7a), Theorem 1 also implies the same convergence result for the error metric $\|x_k - x^*\|^2 + \sum_{n=1}^{N} \|y_k^n - y^{n,*}\|^2$, the formal justification of which is deferred to the proof of Theorem 1. It is worth noting that the order of convergence in Theorem 1 is independent of N, which is in contrast to the convergence rate that gets worse as N increases [12, 61].

Remark 1 (Comparison with prior art in multi-sequence SA) Theorem 1 bridges the gap between the convergence rates of double-sequence linear and nonlinear SA by improving over the $\mathcal{O}(k^{-\frac{2}{3}})$ rate shown in [12] with the additional assumption (7b). As will be shown later, this assumption is satisfied in various applications. Theorem 1 also generalizes the $\mathcal{O}(\frac{1}{k})$ convergence rate in the double-sequence linear SA analysis (e.g., [29]) to the multi-sequence nonlinear SA case.

2.2 The non-strongly-monotone case

Some applications of multi-sequence nonlinear SA such as the actor-critic method [32], Assumption 5 does not hold. This motivates us to consider a more general setting in this subsection where $v(x)$ is non-strongly-monotone. Throughout this subsection, we make the following assumption.

Assumption 6 *Suppose there exists a mapping $F : \mathbb{R}^{d_0} \mapsto \mathbb{R}$ such that $\nabla F(x) = v(x)$. The sequence of $\{x_k\}$ is contained in an open set over which $F(x)$ is upper bounded; e.g. $F(x) \leq C_F$.*

As will be shown later, $F(x)$ can be chosen as the objective function when applying SA to maximization problems. Then assumption 6 is standard to ensure the convergence of x_k; see e.g. [6].

The following theorem gives the general finite-time convergence result of the nonlinear SA when the main sequence has the non-strongly-monotone increment. The proof is deferred to Appendix C.

Theorem 2 *Consider the sequences generated by* (5) *for k=[K]. Suppose Assumptions 1–4 & 6 hold. Select $\alpha_k = \Theta(\frac{1}{\sqrt{K}})$, $\beta_{k,n} = \Theta(\frac{1}{\sqrt{K}})$ with properly chosen initial step sizes, then it holds that*

$$\frac{1}{K}\sum_{k=1}^{K}\Big(\mathbb{E}\|\nabla F(x_k)\|^2 + \sum_{n=1}^{N}\mathbb{E}\|y_k^n - y^{n,*}(y_k^{n-1})\|^2\Big) = \mathcal{O}\Big(\frac{1}{\sqrt{K}}\Big), \tag{13}$$

where $\mathcal{O}(\cdot)$ hides problem dependent constants of a polynomial of N, and we have used $y_k^0 = x_k$.

Theorem 2 implies a finite-time convergence rate of $\mathcal{O}(\epsilon^{-2})$, which is independent of the number of sequences N. The error metric $\|\nabla F(x_k)\|$ used in Theorem 2 is of interest since it is a general measure of the convergence of x_k widely adopted in many applications of SA, especially when the increment of x_k is not strongly-monotone. Moreover, although we have assumed the existence and uniqueness of x^* in (6), the proof of Theorem 2 does not utilize this fact and thus the theorem applies to the more general case where x^* is not unique or even does not exist.

Remark 2 (Comments on stacking all the variables) *One naive way to establish the convergence of a multi-sequence update is to stack all the variables and view it as one sequence. However, the stacked sequence requires stronger assumptions that are violated in the applications to converge. For one, we would need $v, h^1, ..., h^N$ to be jointly lipschitz continuous w.r.t. the stacked variable $(x, y^1, ..., y^N)$. This condition is violated in, e.g., the application of actor-critic (Section 3.2). The upper-bounded function F can also be difficult to find. As it can be seen later in, e.g., Section 3 that such a F only exists for x and might not exist for the stacked variable (x, y).*

Next we will showcase how the results can be applied to optimization and RL problems.

3 Applications to Stochastic Bilevel Optimization

With mappings $f : \mathbb{R}^{d_0} \times \mathbb{R}^{d_1} \mapsto \mathbb{R}$ and $g : \mathbb{R}^{d_0} \times \mathbb{R}^{d_1} \mapsto \mathbb{R}$, consider the following formulation of the bilevel optimization problem:

$$\begin{aligned} &\max_{x \in \mathbb{R}^{d_0}} F(x) \coloneqq f(x, y^*(x)) \coloneqq \mathbb{E}_\zeta\big[f(x, y^*(x); \zeta)\big] \\ &\text{s.t.}\quad y^*(x) \coloneqq \arg\min_{y \in \mathbb{R}^{d_1}} g(x, y) \coloneqq \mathbb{E}_\varphi\big[g(x, y; \varphi)\big] \end{aligned} \tag{14}$$

where ζ and φ are two random variables.

3.1 Reduction from the generic STSA results

A popular approach to solving (14) is the gradient-based method [21, 26, 27, 8]. Under some conditions that will be specified later, the gradient of $F(x)$ takes the following form [21]:

$$\nabla F(x) = \nabla_x f(x, y^*(x)) - \nabla^2_{xy} g(x, y^*(x))[\nabla^2_{yy} g(x, y^*(x))]^{-1} \nabla_y f(x, y^*(x)). \tag{15}$$

Computing (15) requires $y^*(x)$, which is often unknown in practice. Instead, one can iteratively update y_k to approach $y^*(x_k)$ while using y_k in place of $y^*(x_k)$ during the computation of (15) [26, 8]. This leads to an update same as that in (5) with $N = 1$, where the mappings are defined as

$$h(x, y) = -\nabla_y g(x, y), \;\; \psi_k = -h(x_k, y_k) - \nabla_y g(x_k, y_k; \varphi_k), \tag{16a}$$

$$v(x, y) = \nabla_x f(x, y) - \nabla^2_{xy} g(x, y)[\nabla_{yy} g(x, y)]^{-1} \nabla_y f(x, y), \tag{16b}$$

$$\xi_k = -v(x_k, y_k) + \nabla_x f(x_k, y_k; \zeta_k) - \nabla^2_{xy} g(x_k, y_k; \varphi'_k) H^{yy}_k \nabla_y f(x_k, y_k; \zeta_k). \tag{16c}$$

Since we only have two sequences, that is $N = 1$, we omit the index n to simplify notations. In (16), ζ_k is a random variable with the same distribution as that of ζ, and φ_k, φ'_k have the same distribution as that of φ. Here H^{yy}_k is a stochastic approximation of the Hessian inverse $[\nabla_{yy} g(x_k, y_k)]^{-1}$. Given x_k, when y_k reaches the optimal solution $y^*(x_k)$, it follows from (15) that $v(x_k, y^*(x_k)) = \nabla F(x_k)$.

As being discussed below Assumption 1, the lower-level optimal solution $y^*(x_k)$ is drifting at each iteration. Under the Lipschitz continuity assumption of $y^*(x)$, the drifting $\|y^*(x_{k+1}) - y^*(x_k)\|$ scales with $\|x_{k+1} - x_k\|$ which ultimately scales with $\|\nabla F(x_k)\|$. To control the drift scale, former analysis heavily relies on the condition that $\|\nabla F(x_k)\|$ can be bounded for any k. In SBO, this means to either make a strong assumption on the Lipschitz continuity of $f(x, y)$ w.r.t. (x, y), which leads to the Lipschitz continuity of $F(x)$ and the boundedness of $\|\nabla F(x_k)\|$ [8]; or to introduce projection in (16) to forcibly confine x_k in a compact set [26], all of which greatly narrow the range of application. We will show that neither of these the conditions is needed by applying our generic results to SBO.

Lemma 1 (Verifying assumptions of STSA) *Consider the following conditions*

(a) For any $x \in \mathbb{R}^{d_1}$, $g(x, y)$ is strongly convex w.r.t. y with modulus $\lambda_1 > 0$.

(b) There exist constants L_{xy}, l_{xy}, l_{yy} such that $\nabla_y g(x, y)$ is L_{xy}-Lipschitz continuous w.r.t. x; $\nabla_y g(x, y)$ is L_h-Lipschitz continuous w.r.t. y. $\nabla_{xy} g(x, y)$, $\nabla_{yy} g(x, y)$ are respectively l_{xy}-Lipschitz and l_{yy}-Lipschitz continuous w.r.t. (x, y).

(c) There exist constants $l_{fx}, l_{fy}, l'_{fy}, l_y$ such that $\nabla_x f(x, y)$ and $\nabla_y f(x, y)$ are respectively l_{fx} and l_{fy} Lipschitz continuous w.r.t. y; $\nabla_y f(x, y)$ is l'_{fy}-Lipschitz continuous w.r.t. x; $f(x, y)$ is l_y-Lipschitz continuous w.r.t. y.

(d) $F(x)$ satisfies the restricted secant inequality: There exists a constant $\lambda_0 > 0$ such that $\langle \nabla F(x), x - x^* \rangle \leq -\lambda_0 \|x - x^*\|^2$, where $x^* := \arg\max_{x \in \mathbb{R}^{d_1}} F(x)$.

(e) For any k, there exist constants c_0, c_1 such that $\|\mathbb{E}[\xi_k | \mathcal{F}_k^1]\|^2 \leq c_0^2 \alpha_k$ and $\|\mathbb{E}[\psi_k | \mathcal{F}_k]\|^2 \leq c_1^2 \beta_k$; there exist constants σ_0, σ_1 such that $\mathbb{E}[\|\xi_k\|^2 | \mathcal{F}_k^1] \leq \sigma_0^2$ and $\mathbb{E}[\|\psi_k\|^2 | \mathcal{F}_k] \leq \sigma_1^2$.

(f) There exists a constant C_F such that $F(x) \leq C_F$.

We use $a \Rightarrow b$ to indicate that a is a sufficient condition of b. Then we have

$$(a)\&(b) \Rightarrow \textit{ Assumption 1}; \;\; (a)-(c) \Rightarrow \textit{ Assumption 2}; \;\; (e) \Rightarrow \textit{Assumption 3};$$
$$(a) \Rightarrow \textit{ Assumption 4}; \;\; (d) \Rightarrow \textit{ Assumption 5}; \;\; (f) \Rightarrow \textit{ Assumption 6}.$$

The conditions listed above are commonly adopted in the literature [21, 26, 8]. It is worth noting that Lemma 1 does not need the L_{xy}-Lipschitz continuity condition of $f(x, y)$ w.r.t. (x, y). This Lipschitz condition along with the L_y-Lipschitz continuity of $y^*(x)$, which is implied by the standard conditions in Lemma 1, further leads to the Lipschitz continuity of $F(x)$:

$$|F(x) - F(x')| \leq L_{xy}(\|x - x'\| + \|y^*(x) - y^*(x')\|) \leq L_{xy}(L_y + 1)\|x - x'\|. \tag{17}$$

Although it is rather restrictive, this condition has been used in the previous work when $F(x)$ is not strongly-concave. While our analysis does not need this condition. Lastly, condition (e) is guaranteed by using independent samples in the upper and lower level along with [21, Algorithm 3] to obtain a good H_k^{yy}, which takes $\Omega(-\log \alpha_k)$ samples per iteration. With Lemma 1, we have the following corollary regarding the convergence of (16).

Corollary 1 (STSA for SBO) *Consider the STSA sequences with the update in* (16)*. Under Conditions (a)–(e), Theorem 1 holds; that is, with $\alpha_k = \Theta(\frac{1}{k})$ and $\beta_k = \Theta(\frac{1}{k})$ we have*

$$\mathbb{E}\|x_k - x^*\|^2 + \mathbb{E}\|y_k - y^*(x_k)\|^2 = \mathcal{O}\Big(\frac{1}{k}\Big), \tag{18a}$$

$$\lim_{k \to \infty} \|x_k - x^*\|^2 = 0 \quad \text{and} \quad \lim_{k \to \infty} \|y_k - y^*(x_k)\|^2 = 0 \;\; a.s. \tag{18b}$$

Under Conditions (a)–(c), (e) and (f), Theorem 2 holds; i.e., with $\alpha_k = \Theta(\frac{1}{\sqrt{K}})$, $\beta_k = \Theta(\frac{1}{\sqrt{K}})$, we have

$$\frac{1}{K}\sum_{k=1}^{K}\Big(\mathbb{E}\|\nabla F(x_k)\|^2 + \mathbb{E}\|y_k - y^*(x_k)\|^2\Big) = \mathcal{O}\Big(\frac{1}{\sqrt{K}}\Big). \tag{19}$$

Remark 3 (Comparison with prior art in SBO) When $F(x)$ is strongly concave, Corollary 1 implies the sample complexity of $\mathcal{O}(\epsilon^{-1} \log \epsilon^{-1})$, which improves over the best-known sample complexity $\mathcal{O}(\epsilon^{-1.5} \log \epsilon^{-1})$ in [26]. Different from [26], we do not need the projection of x_k to a compact set. When $F(x)$ is non-concave, corollary 1 suggests a sample complexity of $\mathcal{O}(\epsilon^{-2} \log \epsilon^{-1})$, which is the same as the state-of-art complexity established in [8]. Corollary 1 improves the result in [8] in two major aspects: 1) it relaxes the Lipschitz continuity assumption on $f(x, y)$; and, 2) an alternating update is adopted in [8] to ensure stability, while some applications of SBO only allow simultaneous updates. Corollary 1 applies to those cases and thus has a broader range of application.

3.2 Application to advantage actor-critic

RL problems are often modeled as a MDP described by $\mathcal{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma\}$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space; $\mathcal{P}(s'|s, a)$ is the probability of transitioning to $s' \in \mathcal{S}$ given $(s, a) \in \mathcal{S} \times \mathcal{A}$; $r(s, a) \in [0, 1]$ is the reward associated with (s, a); and $\gamma \in (0, 1)$ is a discount factor. A policy π maps $\mathcal{S}$ to a distribution over $\mathcal{A}$, and we use $\pi(a|s)$ to denote the probability of choosing a under s. Given a policy π, we define the value functions as $V_\pi(s) := \mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s\Big]$, where $\mathbb{E}_\pi$ is taken over the trajectory $(s_0, a_0, s_1, a_1, \ldots)$ generated under policy π and transition kernel $\mathcal{P}$. With ρ denoting the initial state distribution, the discounted visitation distribution induced by policy π is defined via $d_\pi(s, a) = (1 - \gamma)\sum_{t=0}^{\infty} \gamma^t \mathbf{Pr}_\pi(s_t = s \mid s_0 \sim \rho)\pi(a|s)$. To overcome the difficulty of learning a function, we parameterize the policy with $x \in \mathbb{R}^{d_0}$, and solve

$$\max_{x \in \mathbb{R}^{d_0}} F(x) := (1 - \gamma)\mathbb{E}_{s \sim \rho}[V_{\pi_x}(s)]. \tag{20}$$

To solve for (20), a popular method is the actor-critic (AC) method [32]. The actor-critic algorithm with linear critic function is a special case of (5). Specifically, the critic variable y is updated with

$$\begin{aligned} h(x,y) &= \mathbb{E}_{s\sim\mu_{\pi_x},a\sim\pi_x,s'\sim\mathcal{P}}[\phi(s)(\gamma\phi(s')-\phi(s))^\top]y + \mathbb{E}_{s\sim\mu_{\pi_x},a\sim\pi_x}[r(s,a)\phi(s)],\\ \psi_k &= -h(x_k,y_k) + \phi(s_k)(\gamma\phi(s'_k)-\phi(s_k))^\top y + r(s_k,a_k)\phi(s_k), \end{aligned} \tag{21}$$

where μ_{π_x} is the stationary distribution of the Markov chain induced by π_x, $\phi(s)\in\mathbb{R}^{d_1}$ is the feature vector encoding state s and the sample (s_k,a_k,s'_k) is returned by some sampling protocol. Under some regularity conditions, it is known that there exists a unique $y^*(x)$ such that $h(x,y^*(x))=0$ [2]. The actor variable x is then updated with

$$\begin{aligned} v(x,y) &= \mathbb{E}_{s,a\sim d_{\pi_x},s'\sim\mathcal{P}}[(r(s,a)+(\gamma\phi(s')-\phi(s))^\top y)\nabla\log\pi_x(a|s)],\\ \xi_k &= -v(x_k,y_k) + r(\bar{s}_k,\bar{a}_k) + \gamma(\phi(\bar{s}'_k)-\phi(\bar{s}_k))^\top y_k\nabla\log\pi_{x_k}(\bar{a}_k|\bar{s}_k). \end{aligned} \tag{22}$$

For the AC update in (21) and (22), Assumption 2–4 and 6 or their sufficient conditions have been explored in the RL context by previous works [57]. However, the smoothness of $y^*(x)$ in Assumption 1, which is the key condition leading to a faster convergence rate, has yet been verified. With the same conditions as those adopted in [57], we prove that $y^*(x)$ is indeed smooth.

Lemma 2 *Consider the AC update in* (21)-(22). *Under the standard conditions specified in Appendix E, $y^*(x)$ is differentiable and there exists $L_{y'}>0$ such that $\|\nabla y^*(x)-\nabla y^*(x')\|\leq L_{y'}\|x-x'\|$.*

As a comparison, the above condition was directly assumed in [8], while we provide a formal justification for Lemma 2 in this work. With detailed verification of all assumptions deferred to Appendix E, we then directly present the theorem regarding the convergence of AC.

Theorem 3 (Complexity of AC) *Consider the AC update* (21)-(22). *Under the standard conditions specified in Appendix E, Theorem 2 holds; that is, with $\alpha_k=\Theta(\frac{1}{\sqrt{K}})$ and $\beta_k=\Theta(\frac{1}{\sqrt{K}})$, we have*

$$\frac{1}{K}\sum_{k=1}^{K}\Big(\mathbb{E}\|\nabla F(x_k)\|^2+\mathbb{E}\|y_k-y^*(x_k)\|^2\Big)=\mathcal{O}\Big(\frac{1}{\sqrt{K}}\Big). \tag{23}$$

In [8, 57], the projection step is adopted in the y_k update to ensure that $\|y_k\|<\infty,\forall k$. Since the projection radius is unknown in practice, adopting the projection is essentially assuming that $\|y_k\|$ can be bounded for any k, which is quite strong. Theorem 3 holds without this projection.

4 Applications to Stochastic Compositional Optimization

Define mappings $f^n:\mathbb{R}^{d_n}\mapsto\mathbb{R}^{d_{n+1}}$ for $n=0,1,...,N$ with $d_{N+1}=1$. The multi-level stochastic compositional problem can be formulated as

$$\max_{x\in\mathbb{R}^{d_0}} F(x) \coloneqq f^N(f^{N-1}(\dots f^0(x)\dots)) \quad \text{with} \quad f^n(x) \coloneqq \mathbb{E}_{\zeta^n}[f^n(x;\zeta^n)]\,, n=0,1,...,N \tag{24}$$

where $\zeta^0,\zeta^1,\dots,\zeta^N$ are random variables. Here we slightly overload the notation and use $f^n(x;\zeta^n)$ to represent the stochastic version of the mapping.

4.1 Reduction from the generic STSA results

To solve the problem in (24), a natural scheme is to use the stochastic gradient descent method with the gradient given by

$$\nabla F(x)=\nabla f^0(x)\nabla f^1(f^0(x))...\nabla f^N(f^{N-1}(...f^0(x)...)) \tag{25}$$

where we use $\nabla f^n(f^{n-1}(\dots f^0(x)\dots)) = \nabla f^n(x)|_{x=f^{n-1}(\dots f^0(x)\dots)}$. To obtain a stochastic estimator of $\nabla F(x)$, we will need to obtain the stochastic estimators for $\nabla f^n(f^{n-1}(...f^0(x)...))$ for each n. For example, when $n=1$, one will need the estimator of $\nabla f^1(\mathbb{E}_{\zeta^0}[f^0(x;\zeta^0)])$. However, due to the possible non-linearity of $\nabla f^1(\cdot)$, the natural candidate $\nabla f^1(f^0(x;\zeta^0))$ is not an unbiased estimator of $\nabla f^1(\mathbb{E}_{\zeta^0}[f^0(x;\zeta^0)])$. To tackle this issue, a popular method is to directly approximate the mean $\mathbb{E}_{\zeta^n}[f^n(\cdot;\zeta^n)]$ with a tracking variable $y^n\in\mathbb{R}^{d_n}$ for $n=0,1,...,N$, see e.g., [61].

The update of y^n is then a special case of the SA update in (5) with the generic mapping defined as

$$h^n(y^{n-1}, y^n) = f^{n-1}(y^{n-1}) - y^n,\ \ \psi_k^n = -h^n(y_k^{n-1}, y_k^n) + f^{n-1}(y_k^{n-1}; \zeta_k^{n-1}) - y_k^n \tag{26}$$

where $\zeta_k^0, \dots, \zeta_k^N$ have the same distributions as that of $\zeta^0, \dots, \zeta^N$ respectively. It is then clear that each y_k^n has a unique fixed-point $y_k^{n,*} = f^{n-1}(y_k^{n-1})$, and thus y_k^n can be viewed as an approximation of $f^n(y_k^{n-1})$. With these approximations, variable x is updated in the form of (5) by defining

$$v(x, y^1, \dots, y^N) = \nabla f^0(x) \nabla f^1(y^1) ... \nabla f^N(y^N),$$
$$\xi_k = -v(x_k, y_k^1, \dots, y_k^N) + \nabla f^0(x_k; \hat{\zeta}_k^0) \cdots \nabla f^N(y_k^N; \zeta_k^N) \tag{27}$$

where $\hat{\zeta}_k^0$ has the same distribution as that of ζ^0. It is clear that when every y_k^n reaches its fixed-point $y_k^{n,*}$, it follows from (25) that $v(x_k, y_k^{1,*}, ..., y_k^{N,*}) = \nabla F(x_k)$, which indicates that the expected update direction of x_k in (27) is $\nabla F(x_k)$.

Next we provide a lemma that summarizes the sufficient conditions of Assumption 1–6. The listed conditions are standard in the literature [61, 7].

Lemma 3 (Verifying assumptions of STSA) *Consider the following conditions*

- *(g) Given any $n \in \{0, 1, \dots, N\}$, there exist positive constants $L_{y,n}$ and $L_{y',n}$ such that the mapping $f^n(\cdot)$ is $L_{y,n}$-Lipschitz continuous and $L_{y',n}$-smooth.*
- *(h) Given $\mathcal{F}_k$, for any $n \in [N]$: $f^n(y_k^{n-1}; \zeta_k^n)$ and $\nabla f^n(y_k^{n-1}; \zeta_k^n)$ are respectively the unbiased estimators of $f^n(y_k^{n-1})$ and $\nabla f^n(y_k^{n-1})$ with bounded variance; $f^0(x_k; \hat{\zeta}_k^0)$ and $\nabla f^0(x_k; \hat{\zeta}_k^0)$ are respectively the unbiased estimators of $f^0(x_k)$ and $\nabla f^0(x_k)$ with bounded variance.*
- *(i) At each iteration k, $\hat{\zeta}_k^0, \zeta_k^0, \zeta_k^1, \dots, \zeta_k^N$ are conditionally independent of each other given $\mathcal{F}_k$.*
- *(j) Function $F(x)$ satisfies the restricted secant inequality: There exists a constant $\lambda_0 > 0$ such that $\langle \nabla F(x), x - x^* \rangle \leq -\lambda_0 \|x - x^*\|^2$, where $x^* := \arg\max_{x \in \mathbb{R}^{d_1}} F(x)$.*
- *(k) There exists a constant C_F such that $F(x) \leq C_F$.*

We use $a \Rightarrow b$ to indicate that a is a sufficient condition of b. Then we have

$$(g) \Rightarrow \textit{Assumption 1 and 2};\quad (h) \textit{ and } (i) \Rightarrow \textit{Assumption 3};\quad (j) \Rightarrow \textit{Assumption 5};$$
$$(k) \Rightarrow \textit{Assumption 6};\quad \textit{Assumption 4 holds for } (26).$$

With Lemma 3, we can directly arrive at the following corollary on the convergence of the stochastic compositional optimization method.

Corollary 2 (STSA for multi-level SCO) *Consider the STSA sequences generated by* (26)-(27). *Under Conditions (g)–(j), Theorem 1 holds. Under Conditions (g)–(i) and (k), Theorem 2 holds.*

Remark 4 (Comparison with prior art in SCO) *Corollary 2 establishes the sample complexity of $\mathcal{O}(\epsilon^{-1})$ for the strongly monotone case and the complexity of $\mathcal{O}(\epsilon^{-2})$ for the non-monotone case, which are both independent of N. This improves over the $\mathcal{O}(\epsilon^{-\frac{N+5}{4}})$ complexity for the strongly concave case and the $\mathcal{O}(\epsilon^{-\frac{N+8}{4}})$ complexity for the non-concave case shown in [61]. There are other works that establish the same complexity as that in Corollary 2, but they require modification to the basic SA update* (26) *and* (27) *to achieve acceleration; see e.g., [7, 1, 47].*

4.2 Application to model-agnostic meta policy gradient

Consider a set of MDPs $\{\mathcal{M}_i\}_{i=1}^M$ with $\mathcal{M}_i = \{\mathcal{S}, \mathcal{A}, \mathcal{P}_i, r_i, \gamma\}$. The MDPs model a set of RL tasks that share the same state-action space while having different transition kernels $\mathcal{P}_i$ and reward functions r_i. To better compare with the previous work [14], we consider the finite-horizon objective function with the policy π parametrized by $x \in \mathbb{R}^{d_0}$: $F_i(x) := \mathbb{E}_{\zeta \sim \pi_x}\Big[\sum_{t=0}^H \gamma^t r_i(s_t, a_t) | \rho_i, \mathcal{P}_i\Big]$, where $H \in \mathbb{N}^+$ is the horizon, and $\mathbb{E}_{\zeta \sim \pi_x}$ is taken over the trajectory $\zeta := (s_0, a_0, s_1, a_1, ..., s_H, a_H)$ generated under policy π_x, initial distribution ρ_i and transition kernels $\mathcal{P}_i$.

The goal of MAMPG is to find an initial policy π_x that can achieve good performance in new tasks by performing a few policy gradient steps [15, 14]. In the case where N steps of gradient update are performed, the problem of finding an initial policy parameter x can be formulated as

$$\max_{x\in\mathbb{R}^{d_0}} F(x) \coloneqq \frac{1}{M}\sum_{i=1}^{M} F_i(\tilde{x}_i^N(x)) \quad \text{with } \tilde{x}_i^{n+1} = \tilde{x}_i^n + \eta\nabla F_i(\tilde{x}_i^n),\ n = 0, 1, ..., N-1, \tag{28}$$

where x is the shared initial policy parameter, i.e. $\tilde{x}_i^0 = x$ for any task i and $\tilde{x}_i^N(x)$ is the parameter after running N steps of gradient ascent with respect to F_i starting from x.

Solving (28) **with SCO method.** The MAMPG problem in (28) can be solved by the stochastic compositional optimization method introduced before. In order to get $\nabla F(x)$, one will need $\nabla F_i(\tilde{x}_i^N(x))$ for each task i. Observe that $F_i(\tilde{x}_i^N(x))$ can be written as a compositional function:

$$F_i(\tilde{x}_i^N(x)) = f_i^N(f_i^{N-1}(\dots f_i^0(x)\dots)) \ \text{ with } \ f_i^n(x) \coloneqq x + \eta\nabla F_i(x),\ n = 0, ..., N-1, \tag{29}$$

where $f_i^N(x) = F_i(x)$. In order to approximate $\nabla F_i(\tilde{x}_i^N(x))$, we can follow the discussion in Section 4 and introduce tracking variables $y_i^n \in \mathbb{R}^{d_0}$ for $n \in [N]$ which are updated as follows

$$y_{k+1,i}^n = y_{k,i}^n - \beta_{k,n}(y_{k,i}^n - f_i^{n-1}(y_k^{n-1}; \zeta_{k,i}^{n-1})),\ n = 0, 1, ..., N-1 \tag{30}$$

where we define $f_i^n(\cdot;\zeta)$ as a stochastic approximation of $f_i^n(\cdot)$ with random trajectory ζ. Then we estimate $\nabla F_i(\tilde{x}_i^N(x))$ by $\hat{\nabla} F_{i,k}$ defined as

$$\hat{\nabla} F_{i,k} \coloneqq \nabla f_i^0(x; \hat{\zeta}_{k,i}^0)\nabla f_i^1(y_{k,i}^1; \zeta_{k,i}^1)\cdots\nabla f_i^N(y_{k,i}^N; \zeta_{k,i}^N). \tag{31}$$

To obtain an estimation of $\nabla F(x)$, we need $\hat{\nabla} F_{i,k}$ for each $i \in \{1, 2, ..., M\}$. Thus we do (30) for each i. With $\{\hat{\nabla} F_{i,k}\}_{i=1}^M$, the initial policy is updated as $x_{k+1} = x_k + \alpha_k \frac{1}{M}\sum_{i=1}^M \hat{\nabla} F_{i,k}$.

Reduction from the generic results. Let $y^n \in \mathbb{R}^{d_n M}$ be a concatenation of y_i^n for $i \in \{1, 2, ..., M\}$. With $\zeta_k^n \coloneqq \{\zeta_{k,i}^n\}_{i=1}^M$, let $f^n(y_k^n; \zeta_k^n)$ be a concatenation of $f_i^n(y_{k,i}^n; \zeta_{k,i}^n)$ for $i \in \{1, 2, ..., M\}$. Then we can write the tracking variable update of all tasks jointly in the form of (5a), that is

$$h^n(y^{n-1}, y^n) = f^{n-1}(y^{n-1}) - y^n,\ \ \psi_k^n = -h^n(y_k^{n-1}, y_k^n) + f^{n-1}(y_k^{n-1}; \zeta_k^{n-1}) - y_k^n. \tag{32}$$

The initial policy update is a special case of (5b), that is

$$v(x, y^1, ..., y^N) = \frac{1}{M}\sum_{i=1}^M \nabla f_i^0(x)...\nabla f_i^N(y_i^N),\ \ \xi_k = -v(x_k, y_k^1, \dots, y_k^N) + \frac{1}{M}\sum_{i=1}^M \hat{\nabla} F_{i,k}. \tag{33}$$

Due to space limitation, we directly give the result below and defer the proof to Appendix G.

Theorem 4 (Complexity of MAMPG) *Consider the STSA sequences generated by the MAMPG update in* (32) *and* (33)*. Under some standard conditions specified in Appendix G, Theorem 2 holds.*

Theorem 4 implies a sample complexity of $\mathcal{O}(\epsilon^{-2})$ to achieve the ϵ-stationary initial policy, which improves over the $\mathcal{O}(\epsilon^{-4})$ sample complexity in [14]. Moreover, Theorem 4 holds for any $N \geq 1$ in (28), while the method in [14] only applies to the case $N = 1$.

5 Conclusions

In this work, we consider the general nonlinear SA with multiple coupled sequences, and study its non-asymptotic performance. Different from the dominating two-timescale SA analysis, we are particularly interested in under which conditions, single-timescale analysis can be applied to nonlinear SA with multiple coupled sequences. When all the sequences have strongly monotone increments, we establish the iteration complexity of $\mathcal{O}(\epsilon^{-1})$. When the main sequence is not strongly-monotone, we establish the iteration complexity of $\mathcal{O}(\epsilon^{-2})$. We then apply our generic SA analysis to stochastic bilevel and compositional optimization and improve their existing results. Specifically, we improve the state-of-the-art convergence rate of: 1) the SBO method and its application to the AC method; and, 2) the multi-level SCO method and its application to the MAMPG method.

Acknowledgments

This work was partially supported by National Science Foundation MoDL-SCALE Grant 2134168 and the Rensselaer-IBM AI Research Collaboration (`http://airc.rpi.edu`), part of the IBM AI Horizons Network (`http://ibm.biz/AIHorizons`).

References

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

[2] J. Bhandari, D. Russo, and R. Singal. A finite time analysis of temporal difference learning with linear function approximation. In *Proc. of Conference on Learning Theory*, 2018.

[3] V. Borkar. Stochastic approximation with two time scales. *System control letter*, 29, 1997.

[4] V. Borkar. *Stochastic approximation: a dynamical systems viewpoint*. Springer, 2009.

[5] V. Borkar and V. Konda. The actor-critic algorithm as multi-time-scale stochastic approximation. *Sadhana*, 22(4):525–543, 1997.

[6] L. Bottou, F. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2), 2018.

[7] T. Chen, Y. Sun, and W. Yin. Solving stochastic compositional optimization is nearly as easy as solving stochastic optimization. *IEEE Transactions on Signal Processing*, 69:4937–4948, 2021.

[8] T. Chen, Y. Sun, and W. Yin. Tighter analysis of alternating stochastic gradient method for stochastic nested problems. In *Proc. of Advances in Neural Information Processing Systems*, 2021.

[9] M. Dagreou, P. Ablin, S. Vaiter, and T. Moreau. A framework for bilevel optimization that enables stochastic and global variance reduction algorithms. *arXiv preprint arXiv:2201.13409*, 2022.

[10] G. Dalal, B. Szorenyi, and G. Thoppe. A tale of two-timescale reinforcement learning with the tightest finite-time bound. In *Proc. of AAAI Conference on Artificial Intelligence*, 2020.

[11] G. Dalal, B. Szorenyi, G. Thoppe, and S. Mannor. Finite sample analysis of two-timescale stochastic approximation with applications to reinforcement learning. In *Proc. of Conference on Learning Theory*, 2018.

[12] T. Doan. Nonlinear two-time-scale stochastic approximation: Convergence and finite-time performance. *arXiv preprint:2011.01868*, 2021.

[13] A. Durmus, E. Moulines, A. Naumov, S. Samsonov, and H. Wai. On the stability of random matrix product with markovian noise: Application to linear stochastic approximation and td learning, 2021.

[14] A. Fallah, K. Georgiev, A. Mokhtari, and A. Ozdaglar. On the convergence theory of debiased model-agnostic meta-reinforcement learning. In *Proc. of Advances in Neural Information Processing Systems*, 2021.

[15] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. of International Conference on Machine Learning*, 2017.

[16] L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi, and M. Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *Proc. of International Conference on Machine Learning*, 2018.

[17] Z. Fu, Z. Yang, and Z. Wang. Single-timescale actor-critic provably finds globally optimal policy. *arXiv preprint:2008.00483*, 2020.

[18] S. Gadet. Stochastic optimization algorithms. `https://perso.math.univ-toulouse.fr/gadat/files/2012/12/cours_Algo_Stos_M2R.pdf`, 2017.

[19] S. Ghadimi and G. Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

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

[21] S. Ghadimi and M. Wang. Approximation methods for bilevel programming. *arXiv preprint:1802.02246*, 2018.

[22] R. Grazzi, L. Franceschi, M. Pontil, and S. Salzo. On the iteration complexity of hypergradient computation. In *Proc. of International Conference on Machine Learning*, pages 3748–3758, 2020.

[23] R. Grazzi, M. Pontil, and S. Salzo. Bilevel optimization with a lower-level contraction: Optimal sample complexity without warm-start. *arXiv preprint arXiv:2202.03397*, 2022.

[24] Z. Guo and T. Yang. Randomized stochastic variance-reduced methods for stochastic bilevel optimization. *arXiv preprint arXiv:2105.02266*, 2021.

[25] H. Gupta, R. Srikant, and L. Ying. Finite-time performance bounds and adaptive learning rate selection for two time-scale reinforcement learning. In *Proc. of Advances in Neural Information Processing Systems*, 2019.

[26] M. Hong, H.-T. Wai, Z. Wang, and Z. Yang. A two-timescale framework for bilevel optimization: Complexity analysis and application to actor-critic. *arXiv preprint:2007.05170*, 2020.

[27] K. Ji, J. Yang, and Y. Liang. Provably faster algorithms for bilevel optimization and applications to meta-learning. In *Proc. of International Conference on Machine Learning*, 2021.

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

[29] M. Kaledin, E. Moulines, A. Naumov, V. Tadic, and H. Wai. Finite time analysis of linear two-timescale stochastic approximation with markovian noise. *Proc. of Conference on Learning Theory*, 2020.

[30] B. Karimi, B. Miasojedow, E. Moulines, and H. Wai. Non-asymptotic analysis of biased stochastic approximation scheme. In *Proc. of Conference on Learning Theory*, 2019.

[31] P. Khanduri, S. Zeng, M. Hong, H. Wai, Z. Wang, and Z. Yang. A momentum-assisted single-timescale stochastic approximation algorithm for bilevel optimization. *arXiv preprint arXiv:2102.07367*, 2021.

[32] V. Konda. *Actor-critic algorithms.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2002.

[33] V. Konda and V. Borkar. Actor-critic–type learning algorithms for markov decision processes. *SIAM Journal on Control and Optimization*, 38(1):94–123, 1999.

[34] V. Konda and J. Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *The Annals of Applied Probability*, 14(2), 2004.

[35] H. Kumar, A. Koppel, and A. Ribeiro. On the sample complexity of actor-critic method for reinforcement learning with function approximation. *arXiv preprint:1910.08412*, 2019.

[36] J. Li, B. Gu, and H. Huang. A fully single loop algorithm for bilevel optimization without hessian inverse. In *Proc. of AAAI Conference on Artificial Intelligence*, 2022.

[37] T. Lin, C. Jin, and M. I. Jordan. On gradient descent ascent for nonconvex-concave minimax problems. *arXiv preprint arXiv:1906.00331*, 2019.

[38] R. Liu, P. Mu, X. Yuan, S. Zeng, and J. Zhang. A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton. In *Proc. of International Conference on Machine Learning*, 2020.

[39] A. Mokkadem and M. Pelletier. Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *The Annals of Applied Probability*, 16(3), 2006.

[40] H. Moshe and V. Ludo. Perturbation bounds for the stationary probabilities of a finite markov chain. *Advances in Applied Probability*, 16(4):804–818, 1984.

[41] W. Mou, J. Li, M. Wainwright, P. Bartlett, and M. Jordan. On linear stochastic approximation: Fine-grained polyak-ruppert and non-asymptotic concentration, 2020.

[42] E. Moulines and F. Bach. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Proc. of Advances in Neural Information Processing Systems*, 2011.

[43] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[44] S. Qiu, Z. Yang, X. Wei, J. Ye, and Z. Wang. Single-timescale stochastic nonconvex-concave optimization for smooth nonlinear td learning. *arXiv preprint arXiv:2008.10103*, 2020.

[45] S. Qiu, Z. Yang, J. Ye, and Z. Wang. On the finite-time convergence of actor-critic algorithm. In *Optimization Foundations for Reinforcement Learning Workshop at Advances in Neural Information Processing Systems*, 2019.

[46] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statis- tics*, 22(3):400–407, 1951.

[47] A. Ruszczynski. A stochastic subgradient method for nonsmooth nonconvex multilevel composition optimization. *SIAM Journal on Control and Optimization*, 59(3):2301–2320, 2021.

[48] S. Sabach and S. Shtern. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27(2):640–660, 2017.

[49] R. Sato, M. Tanaka, and A. Takeda. A gradient method for multilevel optimization. *arXiv preprint:2105.13954*, 2021.

[50] R. Srikant and L. Ying. Finite-time error bounds for linear stochastic approximation and td learning. In *Proc. of Conference on Learning Theory*, 2019.

[51] H. Stackelberg. *The Theory of Market Economy*. Oxford University Press, 1952.

[52] J. Sun, G. Wang, G. B. Giannakis, Q. Yang, and Z. Yang. Finite-time analysis of decentralized temporal-difference learning with linear function approximation. In *International Conference on Artificial Intelligence and Statistics*, 2020.

[53] R. Sutton, H. Maei, D. Precup, S. Bhatnagar, D. Silver, and E. Szepesvári, C.and Wiewiora. Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *Proc. of International Conference on Machine Learning*, 2009.

[54] G. Wang, B. Li, and G. B. Giannakis. A multistep lyapunov approach for finite-time analysis of biased stochastic approximation. *arXiv preprint arXiv:1909.04299*, 2019.

[55] M. Wang, E. Fang, and H. Liu. Stochastic compositional gradient descent: Algorithms for minimizing compositions of expected-value functions. *Mathematical Programming*, 161:419–449, 2017.

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

[57] Y. Wu, W. Zhang, P. Xu, and Q. Gu. A finite time analysis of two time-scale actor critic methods. In *Proc. of Advances in Neural Information Processing Systems*, 2020.

[58] T. Xiao, K. Balasubramanian, and S. Ghadimi. A projection-free algorithm for constrained stochastic multi-level composition optimization. *arXiv preprint:2202.04296*, 2022.

[59] T. Xu, Z. Wang, and Y. Liang. Improving sample complexity bounds for (natural) actor-critic algorithms. In *Proc. of Advances in Neural Information Processing Systems*, 2020.

[60] J. Yang, K. Ji, and Y. Liang. Provably faster algorithms for bilevel optimization. *arXiv preprint arXiv:2106.04692*, 2021.

[61] S. Yang, M. Wang, and E. Fang. Multi-level stochastic gradient methods for nested composition optimization. *SIAM Journal on Optimization*, 29(1), 2018.

[62] Z. Yang, K. Zhang, M. Hong, and T. Başar. A finite sample analysis of the actor-critic algorithm. In *Proc. of IEEE Conference on Decision and Control*, pages 2759–2764, 2018.

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

[64] K. Zhang, A. Koppel, H. Zhu, and T. Başar. Global convergence of policy gradient methods to (almost) locally optimal policies. *SIAM Journal on Control and Optimization*, 58(6):3586–3612, 2019.

[65] Z. Zhang and G. Lan. Optimal algorithms for convex nested stochastic composite optimization. *arXiv preprint arXiv:2011.10076*, 2020.

[66] S. Zou, T. Xu, and Y. Liang. Finite-sample analysis for SARSA with linear function approximation. In *Proc. of Advances in Neural Information Processing Systems*, 2019.

Checklist

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] This work only considers the strongly-monotone increments for the follower sequences, but it will be interesting to also consider the monotone increments with non-unique fixed point y^*. Right now the generic result in this work can only be applied to the unconstrained stochastic optimization, while it will be also interesting to consider whether it is possible to establish similar result that is applicable to the constrained stochastic optimization problems. We will add discussion on the limitations in the revision.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Supplementary Material for “A Single-Timescale Analysis for Stochastic Approximation with Multiple Coupled Sequences"

Table of Contents

A Additional related works

In this section, we review the prior art on the applications of multi-sequence SA.

Gradient-based bilevel optimization. The bilevel optimization was first introduced in [51]. Recently, the gradient-based bilevel optimization methods have gained growing popularity [48, 16, 22, 38]. The finite-time convergence of the double-loop bilevel optimization methods has been studied in some previous works; see e.g., [21, 27]. Later, [26] proved the finite-time convergence rate for the single-loop two time-scale bilevel optimization method, which was then improved by [8] to the optimal rate with additional assumptions and a more refined analysis. There are also other works that incorporate momentum to accelerate the convergence; see e.g., [31, 24, 60]. After our initial conference submission, we have also noticed some concurrent works that are relavant to this work [9, 23, 36]. Specifically, [9] proposed a SBO method with the variance-reduction technique and achieved optimal rate. And [23] proposed a SBO method that achieves the optimal rate without warm-start. The algorithms in [9, 23] are not a case of the SA update discussed in this work and thus its analysis is not applicable to our problem. Lastly, [36] proposed a single-loop SBO method without Hessian inverse, but it required the bounded-gradient assumption which is not needed in this work.

Actor-critic method. After its frist introduction in [33], the finite-sample guarantee for the AC algorithm has been established in [62, 35, 17] with i.i.d. sampling. In [45], the finite-time convergence rate has been established for the nested-loop AC under the Markovian setting, which was later improved improved by [59]. On the other hand, the finite-time convergence of two-timescale AC has been studied in [57] under Markovian sampling and [26, 8] under i.i.d. sampling.

Gradient-based stochastic compositional optimization. The two time-scale stochastic compositional optimization method was proposed in [55, 56]. Due to the two time-scale step sizes choice, the convergence rate of [55, 56] is slower than that of the SGD. In order to achieve acceleration, [20, 7, 47, 1] have modified the basic update in [55, 61] and successfully established the convergence rate same as that of SGD. Concurrent to this work, [28] proposed a variance-reduced SCO method that achieved the optimal rate under variance-reduction. While this work focuses on establishing an optimal rate for the SA update without having diminishing variance. Due to the difference in update scheme, their analysis is not directly applicable to our case.

B Proof of Theorem 1

B.1 Analysis of the lower-level sequences

For brevity, we define the shorthand notations $y_k^{n,*} := y^{n,*}(y_k^{n-1})$ with $y_k^{1,*} := y^{1,*}(x_k)$. Also, we write $\mathbb{E}[\cdot|\mathcal{F}_k]$ as $\mathbb{E}_k[\cdot]$ for brevity.

One-step contraction of lower-level sequences. With $y_k^0 = x_k$, it holds for any $n \in [N]$ that

$$\mathbb{E}_k\|y_{k+1}^n - y_k^{n,*}\|^2 = \|y_k^n - y_k^{n,*}\|^2 + 2\beta_{k,n}\mathbb{E}_k\langle y_k^n - y_k^{n,*}, h^n(y_k^{n-1}, y_k^n) + \psi_k^n\rangle + \mathbb{E}_k\|y_{k+1}^n - y_k^n\|^2. \tag{34}$$

The second term in (34) can be bounded as

$$\begin{aligned}\mathbb{E}_k\langle y_k^n - y_k^{n,*}, h^n(y_k^{n-1}, y_k^n) + \psi_k^n\rangle &= \langle y_k^n - y_k^{n,*}, h^n(y_k^{n-1}, y_k^n))\rangle + \langle y_k^n - y_k^{n,*}, \mathbb{E}_k[\psi_k^n]\rangle \\ &\leq -\lambda_n\|y_k^n - y_k^{n,*}\|^2 + \|y_k^n - y_k^{n,*}\|\|\mathbb{E}_k[\psi_k^n]\| \\ &\leq -\lambda_n\|y_k^n - y_k^{n,*}\|^2 + \frac{\lambda_n}{4}\|y_k^n - y_k^{n,*}\|^2 + \frac{1}{\lambda_n}\|\mathbb{E}_k[\psi_k^n]\|^2 \\ &\leq -\frac{3\lambda_n}{4}\|y_k^n - y_k^{n,*}\|^2 + \frac{c_n^2}{\lambda_n}\beta_{k,n},\end{aligned} \tag{35}$$

where the first inequality follows from the strong monotonicity of $h(y^{n-1}, y^n)$ in Assumption 4, the second inequality follows from the Young's inequality, and the last inequality follows from the bias of the increment ψ_k^n in Assumption 3.

The third term in (34) can be bounded as

$$\mathbb{E}_k\|y_{k+1}^n - y_k^n\|^2 \leq 2\beta_{k,n}^2\big(\|h^n(y_k^{n-1}, y_k^n)\|^2 + \sigma_n^2\big) \leq 2L_{h,n}^2\beta_{k,n}^2\|y_k^n - y_k^{n,*}\| + 2\sigma_n^2\beta_{k,n}^2 \tag{36}$$

where the last inequality follows from Assumption 2 which gives

$$\|h^n(y_k^{n-1}, y_k^n)\| = \|h^n(y_k^{n-1}, y_k^n) - \underbrace{h^n(y_k^{n-1}, y_k^{n,*}(y_k^{n-1}))}_{=0}\| \leq L_{h,n}\|y_k^n - y_k^{n,*}(y_k^{n-1})\|. \tag{37}$$

Collecting the upper bounds in (35) and (36) yields

$$\begin{aligned}\mathbb{E}_k\|y_{k+1}^n - y_k^{n,*}\|^2 &\leq (1 - \frac{3}{2}\lambda_n\beta_{k,n} + 2L_{h,n}^2\beta_{k,n}^2)\|y_k^n - y_k^{n,*}\|^2 + 2(\sigma_n^2 + c_n^2\lambda_n^{-1})\beta_{k,n}^2 \\ &\leq (1 - \lambda_n\beta_{k,n})\|y_k^n - y_k^{n,*}\|^2 + 2(\sigma_n^2 + c_n^2\lambda_n^{-1})\beta_{k,n}^2,\end{aligned} \tag{38}$$

where the last inequality is due to the choice of step size that satisfies $2L_{h,n}^2\beta_{k,n}^2 \leq \frac{\lambda_n}{2}\beta_{k,n}$.

Bounding the drifting optimality gap. For any $n \geq 1$, we have

$$\|y_{k+1}^n - y_{k+1}^{n,*}\|^2 = \|y_{k+1}^n - y_k^{n,*}\|^2 + 2\langle y_k^{n,*} - y_{k+1}^n, y_{k+1}^{n,*} - y_k^{n,*}\rangle + \|y_k^{n,*} - y_{k+1}^{n,*}\|^2. \tag{39}$$

(1) When $n \geq 2$**.** By the mean-value theorem, for some $\hat{y}_{k+1}^{n-1} = ay_k^{n-1} + (1-a)y_{k+1}^{n-1}, a \in [0,1]$, the second term in (39) can be rewritten as

$$\begin{aligned}\langle y_k^{n,*} - y_{k+1}^n, y_{k+1}^{n,*} - y_k^{n,*}\rangle &= \langle y_k^{n,*} - y_{k+1}^n, \nabla y^{n,*}(\hat{y}_{k+1}^{n-1})^\top(y_{k+1}^{n-1} - y_k^{n-1})\rangle \\ &= \langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\nabla y^{n,*}(\hat{y}_{k+1}^{n-1})^\top h^{n-1}(y_k^{n-2}, y_k^{n-1})\rangle \\ &\quad + \langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\nabla y^{n,*}(\hat{y}_{k+1}^{n-1})^\top\psi_k^{n-1}\rangle.\end{aligned} \tag{40}$$

The first term in the right-hand side (RHS) of (40) can be bounded as

$$\begin{aligned}&\langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\nabla y^{n,*}(\hat{y}_{k+1}^{n-1})^\top h^{n-1}(y_k^{n-2}, y_k^{n-1})\rangle \\ &\leq L_{y,n}\beta_{k,n-1}\|y_k^{n,*} - y_{k+1}^n\|\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\| \\ &\leq L_{y,n}L_{h,n-1}\beta_{k,n-1}\|y_k^{n,*} - y_{k+1}^n\|\|y_k^{n-1} - y_k^{n-1,*}\| \\ &\leq \frac{2L_{y,n}^2L_{h,n-1}^2}{\lambda_{n-1}}\beta_{k,n-1}\|y_k^{n,*} - y_{k+1}^n\|^2 + \frac{\lambda_{n-1}}{8}\beta_{k,n-1}\|y_k^{n-1} - y_k^{n-1,*}\|^2\end{aligned} \tag{41}$$

where the second inequality follows from

$$
\begin{aligned}
\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\| &= \|h^{n-1}(y_k^{n-2}, y_k^{n-1}) - h^{n-1}(y_k^{n-2}, y_k^{n-1,*})\| \\
&\leq L_{h,n-1}\|y_k^{n-1} - y_k^{n-1,*}\|.
\end{aligned} \tag{42}
$$

The second term in the RHS of (40) can be further decomposed into

$$
\begin{aligned}
&\langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\nabla y^{n,*}(\hat{y}_{k+1}^{n-1})^\top \psi_k^{n-1}\rangle \\
&= \langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}(\nabla y^{n,*}(\hat{y}_{k+1}^{n-1}) - \nabla y^{n,*}(y_k^{n-1}))^\top \psi_k^{n-1}\rangle \\
&\quad + \langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\nabla y^{n,*}(y_k^{n-1})^\top \psi_k^{n-1}\rangle.
\end{aligned} \tag{43}
$$

Taking expectation on the first term in the RHS of (43) leads to

$$
\begin{aligned}
&\mathbb{E}_k\langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\big(\nabla y^{n,*}(\hat{y}_{k+1}^{n-1}) - \nabla y^{n,*}(y_k^{n-1})\big)^\top \psi_k^{n-1}\rangle \\
&\leq L_{y',n}\beta_{k,n-1}\mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|\hat{y}_{k+1}^{n-1} - y_k^{n-1}\|\|\psi_k^{n-1}\|\big] \\
&\overset{(a)}{\leq} L_{y',n}\beta_{k,n-1}\mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|y_{k+1}^{n-1} - y_k^{n-1}\|\|\psi_k^{n-1}\|\big] \\
&\overset{(b)}{\leq} L_{y',n}\beta_{k,n-1}^2\Big(\mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\|\|\psi_k^{n-1}\|\big] + \mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|\psi_k^{n-1}\|^2\big]\Big) \\
&= L_{y',n}\beta_{k,n-1}^2\Big(\mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\|\mathbb{E}[\|\psi_k^{n-1}\||\mathcal{F}_k^n]\big] + \mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|\psi_k^{n-1}\|^2\big]\Big) \\
&\overset{(c)}{\leq} L_{y',n}\beta_{k,n-1}^2\Big(\sigma_{n-1}\mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\|\big] + \sigma_{n-1}^2\mathbb{E}_k\|y_k^{n,*} - y_{k+1}^n\|\Big) \\
&\leq L_{y',n}\beta_{k,n-1}^2\Big(\sigma_{n-1}\mathbb{E}_k\big[\|y_k^{n,*} - y_{k+1}^n\|\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\|\big] + \frac{\sigma_{n-1}^2}{2}\mathbb{E}_k\|y_k^{n,*} - y_{k+1}^n\|^2 + \frac{\sigma_{n-1}^2}{2}\Big) \\
&\overset{(d)}{\leq} L_{y',n}\sigma_{n-1}\beta_{k,n-1}^2\Big(\frac{L_{h,n-1} + \sigma_{n-1}}{2}\mathbb{E}_k\|y_{k+1}^n - y_k^{n,*}\|^2 + \frac{L_{h,n-1}}{2}\|y_k^{n-1} - y_k^{n-1,*}\|^2 + \frac{\sigma_{n-1}}{2}\Big),
\end{aligned} \tag{44}
$$

where (a) is due to

$$
\|\hat{y}_{k+1}^{n-1} - y_k^{n-1}\| = (1-a)\|y_k^{n-1} - y_{k+1}^{n-1}\| \leq \|y_k^{n-1} - y_{k+1}^{n-1}\|, \tag{45}
$$

then (b) is due to

$$
\|y_{k+1}^{n-1} - y_k^{n-1}\| \leq \beta_k^{n-1}\big(\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\| + \|\psi_k^{n-1}\|\big) \tag{46}
$$

and (c) follows from Assumption 3 and Jensen's inequality:

$$
\mathbb{E}[\|\psi_k^n\|] = \mathbb{E}[\sqrt{\|\psi_k^n\|^2}] \leq \sqrt{\mathbb{E}\|\psi_k^n\|^2} \leq \sigma_n, \tag{47}
$$

the (d) follows from (42) and one-step Young's inequality:

$$
\begin{aligned}
\|y_k^{n,*} - y_{k+1}^n\|\|h^{n-1}(y_k^{n-2}, y_k^{n-1})\| &\overset{(42)}{\leq} L_{h,n-1}\|y_k^{n,*} - y_{k+1}^n\|\|y_k^{n-1} - y_k^{n-1,*}\| \\
&\leq \frac{L_{h,n-1}}{2}\|y_k^{n,*} - y_{k+1}^n\|^2 + \frac{L_{h,n-1}}{2}\|y_k^{n-1} - y_k^{n-1,*}\|^2.
\end{aligned} \tag{48}
$$

The second term in (43) can be bounded as

$$
\begin{aligned}
\mathbb{E}_k\langle y_k^{n,*} - y_{k+1}^n, \beta_{k-1,n}\nabla y^{n,*}(y_k^{n-1})^\top \psi_k^{n-1}\rangle &= \mathbb{E}_k\big[\langle y_k^{n,*} - y_{k+1}^n, \beta_{k,n-1}\nabla y^{n,*}(x_k)^\top \mathbb{E}[\psi_k^{n-1}|\mathcal{F}_k^n]\rangle\big] \\
&\leq L_{y,n}\beta_{k,n-1}\mathbb{E}_k[\|y_k^{n,*} - y_{k+1}^n\|\|\mathbb{E}[\psi_k^{n-1}|\mathcal{F}_k^n]\|] \\
&\overset{(a)}{\leq} \frac{L_{y,n}c_{n-1}}{2}\beta_{k,n-1}\big(\mathbb{E}_k\|y_k^{n,*} - y_{k+1}^n\|^2 + \beta_{k,n-1}\big)
\end{aligned} \tag{49}
$$

where (a) follows from Assumption 3.

Collecting and substituting the upper bounds in (41), (44) and (49) into (40) yields

$$
\begin{aligned}
&\mathbb{E}_k\langle y_k^{n,*}-y_{k+1}^n, y_{k+1}^{n,*}-y_k^{n,*}\rangle \\
&\le \Big(\big(\frac{L_{y,n}c_{n-1}}{2}+\frac{2L_{y,n}^2L_{h,n-1}^2}{\lambda_{n-1}}\big)\beta_{k,n-1}+L_{y',n}\sigma_{n-1}\frac{L_{h,n-1}+\sigma_{n-1}}{2}\beta_{k,n-1}^2\Big)\mathbb{E}_k\|y_{k+1}^n-y_k^{n,*}\|^2 \\
&\quad+\big(\frac{\lambda_{n-1}}{8}\beta_{k,n-1}+\frac{L_{y',n}\sigma_{n-1}L_{h,n-1}}{2}\beta_{k,n-1}^2\big)\|y_k^{n-1}-y_k^{n-1,*}\|^2+\frac{L_{y',n}\sigma_{n-1}^2+L_{y,n}c_{n-1}}{2}\beta_{k,n-1}^2.
\end{aligned}
\tag{50}
$$

The last term in (39) can be bounded as

$$
\begin{aligned}
\mathbb{E}_k\|y_k^{n,*}-y_{k+1}^{n,*}\|^2 &\le L_{y,n}^2\beta_{k,n-1}^2\mathbb{E}_k\|h^{n-1}(y_k^{n-2},y_k^{n-1})+\psi_k^{n-1}\|^2 \\
&\le 2L_{y,n}^2\beta_{k,n-1}^2\|h^{n-1}(y_k^{n-2},y_k^{n-1})\|^2+2L_{y,n}^2\sigma_{n-1}^2\beta_{k,n-1}^2 \\
&\overset{(42)}{\le} 2L_{y,n}^2L_{h,n-1}^2\beta_{k,n-1}^2\|y_k^{n-1}-y_k^{n-1,*}\|^2+2L_{y,n}^2\sigma_{n-1}^2\beta_{k,n-1}^2.
\end{aligned}
\tag{51}
$$

Substituting the upper bounds in (50) and (51) into (39) yields (for $2\le n\le N$)

$$
\begin{aligned}
&\mathbb{E}_k\|y_{k+1}^n-y_{k+1}^{n,*}\|^2 \\
&\le \Big(1+\big(L_{y,n}c_{n-1}+\frac{4L_{y,n}^2L_{h,n-1}^2}{\lambda_{n-1}}\big)\beta_{k,n-1}+L_{y',n}\sigma_{n-1}(L_{h,n-1}+\sigma_{n-1})\beta_{k,n-1}^2\Big)\mathbb{E}_k\|y_{k+1}^n-y_k^{n,*}\|^2 \\
&\quad+\frac{\lambda_{n-1}}{2}\beta_{k,n-1}\|y_k^{n-1}-y_k^{n-1,*}\|^2+(L_{y',n}\sigma_{n-1}^2+L_{y,n}c_{n-1}+2L_{y,n}^2\sigma_{n-1}^2)\beta_{k,n-1}^2
\end{aligned}
\tag{52}
$$

where we have used the following condition of the step size to simplify the inequality:

$$
(L_{y',n}\sigma_{n-1}L_{h,n-1}+2L_{y,n}^2L_{h,n-1}^2)\beta_{k,n-1}^2\le\frac{\lambda_{n-1}}{4}\beta_{k,n-1},\quad 2\le n\le N.
\tag{53}
$$

(2) When $n=1$**.** The update of y_k^1 is correlated with its upper level variable x_k instead of y_k^{n-1} when $n\ge 2$. And since the update of x_k depends on all variables while the update of y_k^{n-1} $(n\ge 2)$ only depends on y_k^{n-2}, the analysis of y_k^1 is different from that of y_k^n $(n>2)$. The difference therefore lies in analyzing (39), which captures the dependence of lower level variable to its upper level variable.

By the mean-value theorem, for some $\hat{x}_{k+1}=ax_k+(1-a)x_{k+1}, a\in[0,1]$, the second term in (39) can be rewritten as

$$
\begin{aligned}
\langle y_k^{1,*}-y_{k+1}^1, y_{k+1}^{1,*}-y_k^{1,*}\rangle &= \langle y_k^{1,*}-y_{k+1}^1, \nabla y^{1,*}(\hat{x}_{k+1})^\top(x_{k+1}-x_k)\rangle \\
&= \langle y_k^{1,*}-y_{k+1}^1, \alpha_k\nabla y^{1,*}(\hat{x}_{k+1})^\top v(x_k,y_k^{1:N})\rangle \\
&\quad+\langle y_k^{1,*}-y_{k+1}^1, \alpha_k\nabla y^{1,*}(\hat{x}_{k+1})^\top\xi_k\rangle.
\end{aligned}
\tag{54}
$$

The first term in the RHS of (54) can be bounded as

$$
\begin{aligned}
&\langle y_k^{1,*}-y_{k+1}^1, \alpha_k\nabla y^{1,*}(\hat{x}_{k+1})^\top v(x_k,y_k^{1:N})\rangle \\
&\le L_{y,1}\alpha_k\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\| \qquad (55)\\
&\overset{(a)}{\le} L_{y,1}\alpha_k\big(L_{v,y}\|y_k^{1,*}-y_{k+1}^1\|\sum_{n=1}^N L_y(n)\|y_k^n-y_k^{n,*}\|+L_v\|y_k^{1,*}-y_{k+1}^1\|\|x_k-x^*\|\big) \\
&\overset{(b)}{\le} L_{y,1}\alpha_k\big(\frac{L_{v,y}}{2}\|y_k^{1,*}-y_{k+1}^1\|^2+\frac{L_{v,y}N}{2}\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2+\frac{L_{y,1}L_v^2}{\lambda_0}\|y_k^{1,*}-y_{k+1}^1\|^2+\frac{\lambda_0}{4L_{y,1}}\|x_k-x^*\|^2\big) \\
&=\big(\frac{L_{y,1}L_{v,y}}{2}+\frac{L_{y,1}^2L_v^2}{\lambda_0}\big)\alpha_k\|y_{k+1}^1-y_k^{1,*}\|^2+\frac{L_{y,1}L_{v,y}N}{2}\alpha_k\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2+\frac{\lambda_0}{4}\alpha_k\|x_k-x^*\|^2
\end{aligned}
\tag{56}
$$

and (a) follows from

$$\begin{aligned}\|v(x_k,y_k^{1:N})\| &= \|v(x_k,y_k^{1:N})-v(x_k)+v(x_k)-\underbrace{v(x^*)}_{=0}\| \\ &\le \|v(x_k,y_k^{1:N})-v(x_k)\|+L_v\|x_k-x^*\| \\ &\le L_{v,y}\sum_{n=1}^N L_y(n)\|y_k^n-y_k^{n,*}\|+L_v\|x_k-x^*\|,\end{aligned} \tag{57}$$

where the first inequality follows from Assumption 2 and the last inequality follows from Lemma 10; and (b) follows from Young's inequality:

$$\begin{aligned}\|y_k^{1,*}-y_{k+1}^1\|\sum_{n=1}^N L_y(n)\|y_k^n-y_k^{n,*}\| &\le \frac{1}{2}\|y_k^{1,*}-y_{k+1}^1\|^2+\frac{1}{2}\Big(\sum_{n=1}^N L_y(n)\|y_k^n-y_k^{n,*}\|\Big)^2 \\ &\le \frac{1}{2}\|y_k^{1,*}-y_{k+1}^1\|^2+\sum_{n=1}^N \frac{N}{2}L_y^2(n)\|y_k^n-y_k^{n,*}\|^2,\end{aligned} \tag{58a}$$

and

$$L_v\|y_k^{1,*}-y_{k+1}^1\|\|x_k-x^*\| \le \frac{L_{y,1}L_v^2}{\lambda_0}\|y_k^{1,*}-y_{k+1}^1\|^2+\frac{\lambda_0}{4L_{y,1}}\|x_k-x^*\|^2. \tag{58b}$$

The second term in the RHS of (54) can be further decomposed as

$$\begin{aligned}&\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(\hat{x}_{k+1})^\top\xi_k\rangle \\ &=\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\big(\nabla y^{1,*}(\hat{x}_{k+1})-\nabla y^{1,*}(x_k)\big)^\top\xi_k\rangle+\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(x_k)^\top\xi_k\rangle.\end{aligned} \tag{59}$$

The first term in the RHS of (59) can be bounded similarly to (44), with the upper level update term $\|x_{k+1}-x_k\|$ in place of $\|y_{k+1}^{n-1}-y_k^{n-1}\|$ $(n>2)$, that is

$$\begin{aligned}&\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\big(\nabla y^{1,*}(\hat{x}_{k+1})-\nabla y^{1,*}(x_k)\big)^\top\xi_k\rangle \\ &=\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\big(\nabla y^{1,*}(\hat{x}_{k+1})-\nabla y^{1,*}(x_k)\big)^\top\xi_k\rangle \\ &\le L_{y',1}\alpha_k\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|\hat{x}_{k+1}-x_k\|\|\xi_k\|\big] \\ &\le L_{y',1}\alpha_k\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|x_{k+1}-x_k\|\|\xi_k\|\big] \\ &\le L_{y',1}\alpha_k^2\Big(\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|\|\xi_k\|\big]+\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|\xi_k\|^2\big]\Big) \\ &=L_{y',1}\alpha_k^2\Big(\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|\mathbb{E}[\|\xi_k\||\mathcal{F}_k^1]\big]+\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\mathbb{E}[\|\xi_k\|^2|\mathcal{F}_k^1]\big]\Big) \\ &\le L_{y',1}\alpha_k^2\Big(\sigma_0\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|\big]+\sigma_0^2\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^1\|\Big) \\ &\le L_{y',1}\alpha_k^2\Big(\sigma_0\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|\big]+\frac{\sigma_0^2}{2}\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^1\|^2+\frac{\sigma_0^2}{2}\Big) \qquad (60) \\ &\le L_{y',1}\sigma_0\alpha_k^2\Big(\frac{\sigma_0+L_{v,y}+L_v}{2}\mathbb{E}_k\|y_{k+1}^1-y_k^{1,*}\|^2+\frac{L_{v,y}N}{2}\sum_{n=1}^N\|y_k^n-y_k^{n,*}\|^2+\frac{L_v}{2}\|x_k-x^*\|^2+\frac{\sigma_0}{2}\Big),\end{aligned} \tag{61}$$

where the fourth inequality follows from Assumption 3 and the last inequality follows from similar derivations of the upper bound of $\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|$ shown in (55)–(56).

The second term in the RHS of (59) can be bounded as

$$\begin{aligned}\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(x_k)^\top\xi_k\rangle &= \mathbb{E}_k\big[\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(x_k)^\top\mathbb{E}[\xi_k|\mathcal{F}_k^1]\rangle\big] \\ &\le L_{y,1}\alpha_k\mathbb{E}_k[\|y_k^{1,*}-y_{k+1}^1\|\|\mathbb{E}[\xi_k|\mathcal{F}_k^1]\|] \\ &\le \frac{L_{y,1}c_0}{2}\alpha_k\big(\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^1\|^2+\alpha_k\big).\end{aligned} \tag{62}$$

Substituting the upper bounds in (56), (61) and (62) into (54) yields

$$
\begin{aligned}
&\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1, y_{k+1}^{1,*}-y_k^{1,*}\rangle \\
&\leq \Big(\big(\frac{L_{y,1}L_{v,y}}{2}+\frac{L_{y,1}^2L_v^2}{\lambda_0}+\frac{L_{y,1}c_0}{2}\big)\alpha_k+L_{y',1}\frac{\sigma_0^2+(L_{v,y}+L_v)\sigma_0}{2}\alpha_k^2\Big)\mathbb{E}_k\|y_{k+1}^1-y_k^{1,*}\|^2 \\
&\quad+\big(\frac{L_{y,1}L_{v,y}N}{2}\alpha_k+\frac{L_{y',1}\sigma_0L_{v,y}N}{2}\alpha_k^2\big)N\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2 \\
&\quad+\big(\frac{\lambda_0}{4}\alpha_k+\frac{L_{y',1}L_v\sigma_0}{2}\alpha_k^2\big)\|x_k-x^*\|^2+\frac{L_{y',1}\sigma_0^2+L_{y,1}c_0}{2}\alpha_k^2.
\end{aligned}
\tag{63}
$$

The last term in (39) can be bounded as

$$
\begin{aligned}
&\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^1\|^2 \\
&\leq L_{y,1}^2\alpha_k^2\mathbb{E}_k\|v(x_k,y_k^{1:N})+\xi_k\|^2 \\
&\leq 2L_{y,1}^2\alpha_k^2\|v(x_k,y_k^{1:N})\|^2+2L_{y,1}^2\sigma_0^2\alpha_k^2 \\
&\overset{(57)}{\leq} 4L_{y,1}^2\alpha_k^2\big(L_{v,y}^2N\sum_{n=1}^N L_y(n)^2\|y_k^n-y_k^{n,*}\|^2+L_v^2\|x_k-x^*\|^2\big)+2L_{y,1}^2\sigma_0^2\alpha_k^2.
\end{aligned}
\tag{64}
$$

Substituting the upper bounds in (63) and (64) into (39) yields

$$
\begin{aligned}
&\mathbb{E}_k\|y_{k+1}^1-y_{k+1}^{1,*}\|^2 \\
&\leq \big(1+L_{y,1}\big(L_{v,y}+2L_{y,1}L_v^2\lambda_0^{-1}+c_0\big)\alpha_k+L_{y',1}\sigma_0(L_{v,y}+L_v+\sigma_0)\alpha_k^2\big)\mathbb{E}_k\|y_{k+1}^1-y_k^{1,*}\|^2 \\
&\quad+\big(L_{y,1}L_{v,y}N\alpha_k+(L_{y',1}\sigma_0L_{v,y}+4L_{y,1}^2L_{v,y}^2)N\alpha_k^2\big)\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2 \\
&\quad+\big(\frac{\lambda_0}{2}\alpha_k+(L_{y',1}L_v\sigma_0+4L_{y,1}^2L_v^2)\alpha_k^2\big)\|x_k-x^*\|^2+(L_{y',1}\sigma_0^2+L_{y,1}c_0+2L_{y,1}^2\sigma_0^2)\alpha_k^2.
\end{aligned}
\tag{65}
$$

This completes the analysis of lower-level sequences.

B.2 Analysis of the main sequence

Recall that we defined the shorthand notations $y_k^{n,*}=y^{n,*}(y_k^{n-1})$ with $y_k^{1,*}=y^{1,*}(x_k)$; $y_k^{1:N}=(y_k^1,y_k^2,\dots,y_k^N)$. For convenience, we write $\mathbb{E}[\cdot|\mathcal{F}_k]$ as $\mathbb{E}_k[\cdot]$. In this section, we will analyze the main sequence and then establish the convergence rate.

First we have

$$
\begin{aligned}
&\mathbb{E}_k\|x_{k+1}-x^*\|^2 \\
&=\|x_k-x^*\|^2+2\alpha_k\mathbb{E}_k\langle x_k-x^*, v(x_k,y_k^{1:N})+\xi_k\rangle+\mathbb{E}_k\|x_{k+1}-x_k\|^2 \\
&=\|x_k-x^*\|^2+2\alpha_k\mathbb{E}_k\langle x_k-x^*, v(x_k,y_k^{1:N})-v(x_k)\rangle+2\alpha_k\langle x_k-x^*, v(x_k)\rangle \\
&\quad+2\alpha_k\langle x_k-x^*,\mathbb{E}_k[\xi_k]\rangle+\alpha_k^2\mathbb{E}_k\|v(x_k,y_k^{1:N})+\xi_k\|^2.
\end{aligned}
\tag{66}
$$

By Lemma 10, the second term in (66) can be bounded as

$$
\begin{aligned}
\langle x_k-x^*, v(x_k,y_k^{1:N})-v(x_k)\rangle &\leq L_{v,y}\|x_k-x^*\|\sum_{n=1}^N L_y(n)\|y_k^n-y_k^{n,*}\| \\
&\leq \frac{\lambda_0}{8}\|x_k-x^*\|^2+\frac{2L_{v,y}^2N}{\lambda_0}\sum_{n=1}^N L_y(n)^2\|y_k^n-y_k^{n,*}\|^2.
\end{aligned}
\tag{67}
$$

By the strong monotonicity of $v(x,y^*(x))$ in Assumption 5, the third term in (66) can be bounded as

$$
\langle x_k-x^*, v(x_k)\rangle \leq -\lambda_0\|x_k-x^*\|^2. \tag{68}
$$

Using Assumption 3, the fourth term in (66) can be bounded as

$$\langle x_k - x^*, \mathbb{E}_k[\xi_k] \rangle \le \frac{\lambda_0}{8} \|x_k - x^*\|^2 + \frac{2c_0^2}{\lambda_0} \alpha_k. \tag{69}$$

The last term in (66) can be bounded as

$$\begin{aligned} \mathbb{E}_k \|v(x_k, y_k^{1:N}) + \xi_k\|^2 &\le 2\|v(x_k, y_k^{1:N})\|^2 + 2\sigma_0^2 \\ &\overset{(57)}{\le} 4L_v^2 \|x_k - x^*\|^2 + 4NL_{v,y}^2 \sum_{n=1}^N L_y(n)^2 \|y_k^n - y_k^{n,*}\|^2 + 2\sigma_0^2. \end{aligned} \tag{70}$$

Substituting the upper bounds in (67)–(70) into (66) yields

$$\begin{aligned} \mathbb{E}_k \|x_{k+1} - x^*\|^2 \le \big(1 - \frac{3}{2}\lambda_0 \alpha_k + 4L_v^2 \alpha_k^2\big) \|x_k - x^*\|^2 + 4\big(\frac{L_{v,y}^2}{\lambda_0} \alpha_k + L_{v,y}^2 \alpha_k^2\big) N \sum_{n=1}^N L_y^2(n) \|y_k^n - y_k^{n,*}\|^2 \\ + 2\big(\sigma_0^2 + \frac{2c_0^2}{\lambda_0}\big) \alpha_k^2. \end{aligned} \tag{71}$$

Establishing convergence. For brevity, we fist define the following series

$$\begin{aligned} &C_0(1) := L_{y,1}\big(L_{v,y} + 2L_{y,1} L_v^2 \lambda_0^{-1} + c_0\big),\ C_1(1) := L_{y',1} \sigma_0 (L_v + L_{v,y} + \sigma_0); \\ &C_0(n) := L_{y,n} c_{n-1} + \frac{4L_{y,n}^2 L_{h,n-1}^2}{\lambda_{n-1}},\ C_1(n) := L_{y',n} \sigma_{n-1} (L_{h,n-1} + \sigma_{n-1}), 2 \le n \le N; \\ &C_2(n) := (4\frac{L_{v,y}^2}{\lambda_0} + \frac{L_{y,1} L_{v,y}}{2}) N L_y^2(n),\ C_3(n) := (L_{v,y}^2 + \frac{L_{y',1} \sigma_0 L_{v,y}}{2}) N L_y^2(n), \forall n. \end{aligned} \tag{72}$$

Define a Lyapunov function $\mathcal{J}_k := \|x_k - x^*\|^2 + \sum_{n=1}^N \|y_k^n - y_k^{n,*}\|^2$. Then we have

$$\mathbb{E}_k[\mathcal{J}_{k+1}] - \mathcal{J}_k = \mathbb{E}_k \|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + \sum_{n=1}^N \|y_{k+1}^n - y_{k+1}^{n,*}\|^2 - \|y_k^n - y_k^{n,*}\|^2. \tag{73}$$

Substituting (52), (65) and (71) into (73), and then applying (38) yields

$$\begin{aligned} &\mathbb{E}_k[\mathcal{J}_{k+1}] - \mathcal{J}_k \\ &\le \big(-\lambda_0 \alpha_k + \big(L_{y',1} L_v \sigma_0 + 4L_{y,1}^2 L_v^2 + 4L_v^2\big) \alpha_k^2 \big) \|x_k - x^*\|^2 \\ &\quad + \sum_{n=1}^{N-1} \big((1 + C_0(n) \beta_{k,n-1} + C_1(n) \beta_{k,n-1}^2)(1 - \lambda_n \beta_{k,n}) - 1 + \frac{\lambda_n}{2} \beta_{k,n} + C_2(n) \alpha_k + C_3(n) \alpha_k^2\big) \|y_k^n - y_k^{n,*}\|^2 \\ &\quad + \big((1 + C_0(N) \beta_{k,N-1} + C_1(N) \beta_{k,N-1}^2)(1 - \lambda_N \beta_{k,N}) - 1 + C_2(N) \alpha_k + C_3(N) \alpha_k^2\big) \|y_k^N - y_k^{N,*}\|^2 \\ &\quad + \Theta(\alpha_k^2) + \Theta\big(\sum_{n=1}^N (1 + \beta_{k,n-1} + \beta_{k,n-1}^2) \beta_{k,n}^2\big), \end{aligned} \tag{74}$$

where we define $\beta_{k,0} := \alpha_k$ to simplify the result. As a clarification, the second term in the last inequality disappears when $N \le 1$. Let the step sizes satisfy

$$-\lambda_0 \alpha_k + \big(L_{y',1} L_v \sigma_0 + 4L_{y,1}^2 L_v^2 + 4L_v^2\big) \alpha_k^2 \le -\frac{\lambda_0}{2} \alpha_k, \tag{75}$$

$$(1 + C_0(n) \beta_{k,n-1} + C_1(n) \beta_{k,n-1}^2)(1 - \lambda_n \beta_{k,n}) - 1 + \frac{\lambda_n}{2} \beta_{k,n} + C_2(n) \alpha_k + C_3(n) \alpha_k^2 \le -\frac{\lambda_0}{2} \alpha_k, 1 \le n \le N-1, \tag{76}$$

$$(1 + C_0(N) \beta_{k,n-1} + C_1(N) \beta_{k,n-1}^2)(1 - \lambda_N \beta_{k,N}) - 1 + C_2(N) \alpha_k + C_3(N) \alpha_k^2 \le -\frac{\lambda_0}{2} \alpha_k, \tag{77}$$

Note that (75) always admits solution for small enough α_1. Given $\beta_{k,N}$, applying Lemma 11 for $n = N, \ldots, 1$ to (77) and (76) implies that there exist solutions for $\beta_{k,n} (\forall n)$.

Then by (75)–(77), we have from (74) that

$$\mathbb{E}_k[\mathcal{J}_{k+1}] \leq \big(1 - \frac{\lambda_0}{2}\alpha_k\big)\mathcal{J}_k + \Theta(\alpha_k^2) + \Theta\big(\sum_{n=1}^{N}(1+\beta_{k,n-1}+\beta_{k,n-1}^2)\beta_{k,n}^2\big). \tag{78}$$

Note that (78) implies a finite-time convergence rate of $\frac{1}{k}$ with the choice of step size. Applying Robbins-Siegmund's theorem stated in Lemma 12 to (78) gives $\sum_{k=1}^{\infty}\alpha_k\mathcal{J}_k < \infty$ and $\lim_{k\to\infty}\mathcal{J}_k < \infty$ almost surely, which along with the fact that $\sum_{k=1}^{\infty}\alpha_k = \infty$ implies $\lim_{k\to\infty}\mathcal{J}_k = 0$, i.e. for any $n \in [N]$

$$\lim_{k\to\infty}\|x_k - x^*\|^2 = 0, \qquad \lim_{k\to\infty}\|y_k^n - y_k^{n,*}\|^2 = 0,\ a.s. \tag{79}$$

Finally, as a direct result of Lemma 13, we can directly obtain the same convergence theorem for the alternative error metric $\|x_k - x^*\|^2 + \sum_{n=1}^{N}\|y_k^n - y^{n,*}\|^2$. This completes the proof.

C Proof of Theorem 2

C.1 Analysis of the lower-level sequences

In this section, we provide a bound of the lower-level optimality gaps. Recall that we defined the shorthand notations $y_k^{n,*} = y^{n,*}(y_k^{n-1})$ with $y_k^{1,*} = y^{1,*}(x_k)$; $y_k^{1:N} = (y_k^1, y_k^2, \dots, y_k^N)$. For convenience, we write $\mathbb{E}[\cdot|\mathcal{F}_k]$ as $\mathbb{E}_k[\cdot]$.

It follows from (38) that

$$\mathbb{E}_k\|y_{k+1}^n - y_k^{n,*}\|^2 \leq (1-\lambda_n\beta_{k,n})\|y_k^n - y_k^{n,*}\|^2 + 2(\sigma_n^2 + c_n^2\lambda_n^{-1})\beta_{k,n}^2. \tag{80}$$

Bounding the drifting optimality gap. For any $n \geq 1$, we have

$$\|y_{k+1}^n - y_{k+1}^{n,*}\|^2 = \|y_{k+1}^n - y_k^{n,*}\|^2 + 2\langle y_k^{n,*} - y_{k+1}^n, y_{k+1}^{n,*} - y_k^{n,*}\rangle + \|y_k^{n,*} - y_{k+1}^{n,*}\|^2. \tag{81}$$

(1) When $n = 1$. By the mean-value theorem, for some $\hat{x}_{k+1} = ax_k + (1-a)x_{k+1}, a \in [0,1]$, the second term in (81) can be rewritten as

$$\begin{aligned}
\langle y_k^{1,*} - y_{k+1}^1, y_{k+1}^{1,*} - y_k^{1,*}\rangle &= \langle y_k^{1,*} - y_{k+1}^1, \nabla y^{1,*}(\hat{x}_{k+1})^\top (x_{k+1} - x_k)\rangle \\
&= \langle y_k^{1,*} - y_{k+1}^1, \alpha_k \nabla y^{1,*}(\hat{x}_{k+1})^\top v(x_k, y_k^{1:N})\rangle \\
&\quad + \langle y_k^{1,*} - y_{k+1}^1, \alpha_k \nabla y^{1,*}(\hat{x}_{k+1})^\top \xi_k\rangle.
\end{aligned} \tag{82}$$

The first term in (82) can be bounded as

$$\begin{aligned}
&\langle y_k^{1,*} - y_{k+1}^1, \alpha_k \nabla y^{1,*}(\hat{x}_{k+1})^\top v(x_k, y_k^{1:N})\rangle \\
&\leq L_{y,1}\alpha_k\|y_k^{1,*} - y_{k+1}^1\|\|v(x_k, y_k^{1:N})\| \qquad (83)\\
&\leq L_{y,1}\alpha_k\Big(L_{v,y}\|y_k^{1,*} - y_{k+1}^1\|\sum_{n=1}^{N}L_y(n)\|y_k^n - y_k^{n,*}\| + \|y_k^{1,*} - y_{k+1}^1\|\|v(x_k)\|\Big) \\
&\leq L_{y,1}\alpha_k\Big(\frac{L_{v,y}}{2}\|y_k^{1,*} - y_{k+1}^1\|^2 + \frac{L_{v,y}N}{2}\sum_{n=1}^{N}L_y^2(n)\|y_k^n - y_k^{n,*}\|^2 \\
&\qquad + 2L_{y,1}\|y_k^{1,*} - y_{k+1}^1\|^2 + \frac{1}{8L_{y,1}}\|v(x_k)\|^2\Big) \\
&= L_{y,1}\big(\frac{L_{v,y}}{2} + 2L_{y,1}\big)\alpha_k\|y_{k+1}^1 - y_k^{1,*}\|^2 + \frac{L_{y,1}L_{v,y}N}{2}\alpha_k\sum_{n=1}^{N}L_y^2(n)\|y_k^n - y_k^{n,*}\|^2 + \frac{1}{8}\alpha_k\|v(x_k)\|^2,
\end{aligned} \tag{84}$$

where the second inequality follows from Lemma 10:

$$\begin{aligned}
\|v(x_k, y_k^{1:N})\| &\leq \|v(x_k, y_k^{1:N}) - v(x_k)\| + \|v(x_k)\| \\
&\leq L_{v,y}\sum_{n=1}^{N}L_y(n)\|y_k^n - y_k^{n,*}\| + \|v(x_k)\|.
\end{aligned} \tag{85}$$

The second term in (82) can be further decomposed as

$$
\begin{aligned}
&\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(\hat{x}_{k+1})^\top\xi_k\rangle\\
&=\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\big(\nabla y^{1,*}(\hat{x}_{k+1})-\nabla y^{1,*}(x_k)\big)^\top\xi_k\rangle+\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(x_k)^\top\xi_k\rangle.
\end{aligned}
\tag{86}
$$

The first term in (86) can be bounded as

$$
\begin{aligned}
&\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\big(\nabla y^{1,*}(\hat{x}_{k+1})-\nabla y^{1,*}(x_k)\big)^\top\xi_k\rangle &(87)\\
&\overset{(60)}{\le} L_{y',1}\sigma_0\alpha_k^2\Big(\mathbb{E}_k\big[\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|\big]+\frac{\sigma_0}{2}\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^1\|^2+\frac{\sigma_0}{2}\Big)\\
&\le L_{y',1}\sigma_0\alpha_k^2\Big(\frac{L_{v,y}+\sigma_0+1}{2}\mathbb{E}_k\|y_{k+1}^1-y_k^{1,*}\|^2+\frac{L_{v,y}N}{2}\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2+\frac{1}{2}\|v(x_k)\|^2+\frac{\sigma_0}{2}\Big)
\end{aligned}
$$

where the last inequality follows from similar derivations of the upper bound of $\|y_k^{1,*}-y_{k+1}^1\|\|v(x_k,y_k^{1:N})\|$ shown in (83)–(84).

The second term in (86) can be bounded as

$$
\mathbb{E}_k\langle y_k^{1,*}-y_{k+1}^1,\alpha_k\nabla y^{1,*}(x_k)^\top\xi_k\rangle\overset{(62)}{\le}\frac{L_{y,1}c_0}{2}\alpha_k\big(\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^1\|^2+\alpha_k\big).\tag{88}
$$

Substituting the upper bounds in (84), (87) and (88) into (82) yields

$$
\begin{aligned}
\langle y_k^{1,*}-y_{k+1}^1,y_{k+1}^{1,*}-y_k^{1,*}\rangle\le{}&\big(L_{y,1}\big(\frac{L_{v,y}+c_0}{2}+2L_{y,1}\big)\alpha_k+L_{y',1}\sigma_0\frac{L_{v,y}+\sigma_0+1}{2}\alpha_k^2\big)\|y_{k+1}^1-y_k^{1,*}\|^2\\
&+\frac{1}{2}\big(L_{y,1}L_{v,y}N\alpha_k+L_{y',1}\sigma_0L_{v,y}N\alpha_k^2\big)\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2\\
&+\big(\frac{1}{8}\alpha_k+\frac{L_{y',1}\sigma_0}{2}\alpha_k^2\big)\|v(x_k)\|^2+\frac{L_{y',1}\sigma_0^2+L_{y,1}c_0}{2}\alpha_k^2.
\end{aligned}
\tag{89}
$$

The last term in (81) can be bounded as

$$
\begin{aligned}
&\mathbb{E}_k\|y_k^{1,*}-y_{k+1}^{1,*}\|^2\\
&\le L_{y,1}^2\alpha_k^2\mathbb{E}_k\|v(x_k,y_k^1)+\xi_k\|^2\le 2L_{y,1}^2\alpha_k^2\|v(x_k,y_k^1)\|^2+2L_{y,1}^2\sigma_0^2\alpha_k^2\\
&\overset{(85)}{\le} 4L_{v,y}^2L_{y,1}^2N\alpha_k^2\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2+4L_{y,1}^2\alpha_k^2\|v(x_k)\|^2+2L_{y,1}^2\sigma_0^2\alpha_k^2.
\end{aligned}
\tag{90}
$$

Substituting the upper bounds in (89) and (90) into (81) yields

$$
\begin{aligned}
&\mathbb{E}_k\|y_{k+1}^1-y_{k+1}^{1,*}\|^2\\
&\le\big(1+L_{y,1}\big(L_{v,y}+c_0+4L_{y,1}\big)\alpha_k+L_{y',1}\sigma_0(L_{v,y}+\sigma_0+1)\alpha_k^2\big)\mathbb{E}_k\|y_{k+1}^1-y_k^{1,*}\|^2\\
&\quad+\big(L_{y,1}L_{v,y}N\alpha_k+(L_{y',1}\sigma_0L_{v,y}+4L_{v,y}^2L_{y,1}^2)N\alpha_k^2\big)\sum_{n=1}^N L_y^2(n)\|y_k^n-y_k^{n,*}\|^2\\
&\quad+\big(\frac{1}{4}\alpha_k+(L_{y',1}\sigma_0+4L_{y,1}^2)\alpha_k^2\big)\|v(x_k)\|^2+(L_{y',1}\sigma_0^2+L_{y,1}c_0+2L_{y,1}^2\sigma_0^2)\alpha_k^2.
\end{aligned}
\tag{91}
$$

(2) When $n\ge 2$**.** The update of y_k^n $(n\ge2)$ has no direct dependence on x_k, therefore the analysis is identical to that of Theorem 1. It directly follows from (51) that

$$
\begin{aligned}
&\mathbb{E}_k\|y_{k+1}^n-y_{k+1}^{n,*}\|^2\\
&\le\big(1+\big(L_{y,n}c_{n-1}+\frac{4L_{y,n}^2L_{h,n-1}^2}{\lambda_{n-1}}\big)\beta_{k,n-1}+L_{y',n}\sigma_{n-1}(L_{h,n-1}+\sigma_{n-1})\beta_{k,n-1}^2\big)\mathbb{E}_k\|y_{k+1}^n-y_k^{n,*}\|^2\\
&\quad+\frac{\lambda_{n-1}}{2}\beta_{k,n-1}\|y_k^{n-1}-y_k^{n-1,*}\|^2+(L_{y',n}\sigma_{n-1}^2+L_{y,n}c_{n-1}+2L_{y,n}^2\sigma_{n-1}^2)\beta_{k,n-1}^2
\end{aligned}
\tag{92}
$$

where we have imposed the following condition on the step size

$$
(L_{y',n}\sigma_{n-1}L_{h,n-1}+2L_{y,n}^2L_{h,n-1}^2)\beta_{k,n-1}^2\le\frac{\lambda_{n-1}}{4}\beta_{k,n-1},\ 2\le n\le N.\tag{93}
$$

This completes the analysis of the lower-level sequences.

C.2 Analysis of the main sequence

In this section, we provide an analysis of the main sequence update, and then establish the finite-time convergence rate. Recall the shorthand notations $y_k^{n,*} = y^{n,*}(y_k^{n-1})$ with $y_k^{1,*} = y^{1,*}(x_k)$.

By the L_v-smoothness of $F(x)$, we have

$$\begin{aligned}
&\mathbb{E}_k[F(x_{k+1})] - F(x_k) \\
&\geq \mathbb{E}_k\big\langle v(x_k), x_{k+1} - x_k\big\rangle - \frac{L_v}{2}\mathbb{E}_k\|x_{k+1} - x_k\|^2 \\
&= \mathbb{E}_k\big\langle v(x_k), \alpha_k v(x_k, y_k^{1:N})\big\rangle + \mathbb{E}_k\big\langle v(x_k), \alpha_k \xi_k\big\rangle - \frac{L_v}{2}\mathbb{E}_k\|x_{k+1} - x_k\|^2.
\end{aligned} \tag{94}$$

Define $L_y(n) \coloneqq \sum_{i=n}^N L_{y,i-1}L_{y,i-2}\dots L_{y,n}$ with $L_{y,n-1}L_{y,i-2}\dots L_{y,n} \coloneqq 1$. Using Lemma 10, the first term in (94) can be bounded as

$$\begin{aligned}
\big\langle v(x_k), \alpha_k v(x_k, y_k^{1:N})\big\rangle &= \big\langle v(x_k), \alpha_k(v(x_k, y_k^{1:N}) - v(x_k))\big\rangle + \alpha_k\|v(x_k)\|^2 \\
&\geq -L_{v,y}\alpha_k\Big[\big\|v(x_k)\big\|\sum_{n=1}^N L_y(n)\|y_k^n - y_k^{n,*}\|\Big] + \alpha_k\|v(x_k)\|^2 \\
&\geq -\frac{\alpha_k}{4}\|v(x_k)\|^2 - L_{v,y}^2 N\alpha_k\sum_{n=1}^N L_y^2(n)\|y_k^n - y_k^{n,*}\|^2 + \alpha_k\|v(x_k)\|^2 \\
&= \frac{3\alpha_k}{4}\|v(x_k)\|^2 - L_{v,y}^2 N\alpha_k\sum_{n=1}^N L_y^2(n)\|y_k^n - y_k^{n,*}\|^2.
\end{aligned} \tag{95}$$

The second term in (94) can be bounded as

$$\begin{aligned}
\mathbb{E}_k\big\langle v(x_k), \alpha_k\xi_k\big\rangle &= \big\langle v(x_k), \alpha_k\mathbb{E}_k[\xi_k]\big\rangle \\
&\geq -\frac{\alpha_k}{4}\|v(x_k)\|^2 - \alpha_k\|\mathbb{E}_k[\xi_k]\|^2 \\
&\geq -\frac{\alpha_k}{4}\|v(x_k)\|^2 - c_0^2\alpha_k^2.
\end{aligned} \tag{96}$$

The last term in (94) can be bounded as

$$\begin{aligned}
\mathbb{E}_k\|x_{k+1} - x_k\|^2 &\leq 2\alpha_k^2\big(\|v(x_k, y_k^{1:N})\|^2 + \mathbb{E}_k\|\xi_k\|^2\big) \\
&\overset{(85)}{\leq} 4\alpha_k^2\|v(x_k)\|^2 + 4L_{v,y}^2 N\alpha_k^2\sum_{n=1}^N L_y^2(n)\|y_k^n - y_k^{n,*}\|^2 + 2\sigma_0^2\alpha_k^2.
\end{aligned} \tag{97}$$

Substituting the bounds in (95), 96 and (97) into (94) yields

$$\begin{aligned}
&\mathbb{E}_k[F(x_{k+1})] - F(x_k) \\
&\geq (\frac{\alpha_k}{2} - 2L_v\alpha_k^2)\|v(x_k)\|^2 - N(L_{v,y}^2\alpha_k + 2L_vL_{v,y}^2\alpha_k^2)\sum_{n=1}^N L_y^2(n)\|y_k^n - y_k^{n,*}\|^2 - (L_v\sigma_0^2 + c_0^2)\alpha_k^2.
\end{aligned} \tag{98}$$

Establishing convergence. For brevity, we fist define the following series

$$\begin{aligned}
&C_4(1) \coloneqq L_{y,1}\big(L_{v,y} + c_0 + 4L_{y,1}\big),\ C_5(1) \coloneqq L_{y',1}\sigma_0\big(L_{v,y} + \sigma_0 + 1\big); \\
&C_4(n) \coloneqq L_{y,n}c_{n-1} + 4L_{y,n}^2 L_{h,n-1}^2\lambda_{n-1}^{-1},\ C_5(n) \coloneqq L_{y',n}\sigma_{n-1}(L_{h,n-1} + \sigma_{n-1}), 2 \leq n \leq N; \\
&C_6(n) \coloneqq \big(L_{y,1}L_{v,y} + L_{v,y}^2\big)NL_y^2(n),\ C_7(n) \coloneqq \big(L_{y',1}\sigma_0 L_{v,y} + 4L_{v,y}^2L_{y,1}^2 + 2L_vL_{v,y}^2\big)NL_y^2(n), \forall n.
\end{aligned} \tag{99}$$

Define a Lyapunov function $\mathcal{L}_k \coloneqq -F(x_k) + \sum_{n=1}^N \|y_k^n - y_k^{n,*}\|^2$. Then we have

$$\mathbb{E}_k[\mathcal{L}_{k+1}] - \mathcal{L}_k = F(x_k) - \mathbb{E}_k[F(x_{k+1})] + \sum_{n=1}^N \mathbb{E}_k\|y_{k+1}^n - y_{k+1}^{n,*}\|^2 - \|y_k^n - y_k^{n,*}\|^2. \tag{100}$$

Substituting (91), (92) and (98) into (100), and then applying (38) yields

$$
\begin{aligned}
&\mathbb{E}_k[\mathcal{L}_{k+1}] - \mathcal{L}_k \\
&\leq \big(-\frac{1}{4}\alpha_k + (L_{y',1}\sigma_0 + 4L_{y,1}^2 + 2L_v)\alpha_k^2\big)\|v(x_k)\|^2 \\
&\quad + \sum_{n=1}^{N-1}\big((1+C_4(n)\beta_{k,n-1}+C_5(n)\beta_{k,n-1}^2)(1-\lambda_n\beta_{k,n}) - 1 + \frac{\lambda_n}{2}\beta_{k,n} + C_6(n)\alpha_k + C_7(n)\alpha_k^2\big)\|y_k^n - y_k^{n,*}\|^2 \\
&\quad + \big((1+C_4(N)\beta_{k,N-1}+C_5(N)\beta_{k,N-1}^2)(1-\lambda_N\beta_{k,N}) - 1 + C_6(N)\alpha_k + C_7(N)\alpha_k^2\big)\|y_k^N - y_k^{N,*}\|^2 \\
&\quad + \Theta(\alpha_k^2) + \Theta\Big(\sum_{n=1}^{N}(1+\beta_{k,n-1}+\beta_{k,n-1}^2)\beta_{k,n}^2\Big).
\end{aligned} \tag{101}
$$

As a clarification, the second term in the last inequality is 0 when $N = 1$. We have also used $\beta_{k,0} = \alpha_k$. Consider the following choice of step sizes

$$
-\frac{1}{4}\alpha_k + (L_{y',1}\sigma_0 + 4L_{y,1}^2 + 2L_v)\alpha_k^2 \leq -\frac{1}{8}\alpha_k, \tag{102}
$$

$$
(1+C_1(n)\beta_{k,n-1}+C_2(n)\beta_{k,n-1}^2)(1-\lambda_n\beta_{k,n}) - 1 + \frac{\lambda_n}{2}\beta_{k,n} + C_3(n)\alpha_k + C_4(n)\alpha_k^2 \leq -\lambda_n\alpha_k, n \leq N-1, \tag{103}
$$

$$
(1+C_1(N)\beta_{k,N-1}+C_2(N)\beta_{k,N-1}^2)(1-\lambda_N\beta_{k,N}) - 1 + C_3(N)\alpha_k + C_4(N)\alpha_k^2 \leq -\lambda_N\alpha_k. \tag{104}
$$

Note that (102) always admits solution for small enough α_1. Given $\beta_{k,N}$, applying Lemma 11 for $n = N, \ldots, 1$ to (104) and (103) tells that there exist solutions for $\beta_{k,n}(\forall n)$.

With (102)–(104), it follows from (101) that

$$
\begin{aligned}
&\mathbb{E}_k[\mathcal{L}_{k+1}] - \mathcal{L}_k \\
&\leq -\frac{\alpha_k}{8}\|v(x_k)\|^2 - \sum_{n=1}^{N}\lambda_n\alpha_k\|y_k^n - y_k^{n,*}\|^2 + \Theta(\alpha_k^2) + \Theta\Big(\sum_{n=1}^{N}(1+\beta_{k,n-1}+\beta_{k,n-1}^2)\beta_{k,n}^2\Big).
\end{aligned} \tag{105}
$$

Furthermore, taking expectation on both sides of (105) then summing over $k = 1, \ldots, K$ yields

$$
\begin{aligned}
&\sum_{k=1}^{K}\alpha_k\mathbb{E}\Big[\frac{1}{8}\|v(x_k)\|^2 + \lambda_n\|y_k^n - y_k^{n,*}\|^2\Big] \\
&\leq \mathcal{L}_1 - \mathbb{E}[\mathcal{L}_{K+1}] + \Theta\Big(\sum_{k=1}^{K}\alpha_k^2\Big) + \Theta\Big(\sum_{k=1}^{K}\sum_{n=1}^{N}(1+\beta_{k,n-1}+\beta_{k,n-1}^2)\beta_{k,n}^2\Big) \\
&\leq \mathcal{L}_1 + C_F + \Theta\Big(\sum_{k=1}^{K}\alpha_k^2\Big) + \Theta\Big(\sum_{k=1}^{K}\sum_{n=1}^{N}(1+\beta_{k,n-1}+\beta_{k,n-1}^2)\beta_{k,n}^2\Big).
\end{aligned} \tag{106}
$$

The inequality (106) implies a convergence rate of $\mathcal{O}(\frac{1}{\sqrt{K}})$ with step sizes $\alpha_k = \Theta(\frac{1}{\sqrt{K}})$ and $\beta_k = \Theta(\frac{1}{\sqrt{K}})$. This completes the proof.

D Proof of Lemma 1 and Corollary 1

To prove the corollary, it suffices to prove Lemma 1 and then directly apply Theorem 1 and 2. We direct the readers interested in why we can relax the assumptions in [8] to the proof of Theorem 1 and 2. In particular, we provide a refined technique on bounding the drifting optimality gap in (39) and (81), which is crucial in alleviating the assumption.

Proof. We start to verify the Assumptions by order.

(1) Conditions (a) and (b) $\Rightarrow$ Assumption 1. Since $g(x, y)$ is strongly-convex w.r.t. y, there exists a unique $y^*(x)$ such that $h(x, y^*(x)) = -\nabla_y g(x, y^*(x)) = 0$.

By [21, Lemma 2.2], we have

$$\|y^*(x) - y^*(x')\| \leq L_y \|x - x'\|, \ L_y = \frac{L_{xy}}{\lambda_1}. \tag{107}$$

By [8, Lemma 2], we have

$$\|\nabla y^*(x) - \nabla y^*(x')\| \leq L_{y'} \|x - x'\|, \ L_{y'} = \frac{l_{xy} + l_{xy} L_y}{\lambda_1} + \frac{L_{xy}(l_{yy} + l_{yy} L_y)}{\lambda_1^2}. \tag{108}$$

(2) Conditions (a)–(c) ⇒ Assumption 2. By [21, Lemma 2.2], we have

$$\|v(x,y) - v(x,y')\| \leq L_{v,y} \|y - y'\|, \ L_{v,y} = l_{fx} + \frac{l_{fy} L_{xy}}{\lambda_1} + l_y \Big(\frac{l_{xy}}{\lambda_1} + \frac{L_h L_{xy}}{\lambda_1^2}\Big) \tag{109a}$$

$$\|v(x) - v(x')\| \leq L_v \|y - y'\|, \ L_v = \frac{L_{xy}(L_{v,y} + l'_{fy})}{\lambda_1} + l_{fx} + l_y \Big(\frac{l_{xy} l_y}{\lambda_1} + \frac{l_{yy} L_{xy}}{\lambda_1^2}\Big). \tag{109b}$$

Lastly, it follows from condition (b) that $\|h(x,y) - h(x,y')\| \leq L_h \|y - y'\|$.

(3) Condition (e) ⇒ Assumption 3; (a) ⇒ Assumption 4; (d) ⇒ Assumption 5; (f) ⇒ Assumption 6. These conditions directly imply their corresponding Assumption 3 when $N = 1$. ■

E Proof of Theorem 3

In this section, we will provide a proof of theorem 3. We omit all the index n since $N = 1$. We also write $y^*(x_k)$ in short as y_k^*. With $A_x := \mathbb{E}_{s \sim \mu_{\pi_x}, a \sim \pi_x, s' \sim \mathcal{P}}[\phi(s)(\gamma\phi(s') - \phi(s))^\top], b_x := \mathbb{E}_{s \sim \mu_{\pi_x}, a \sim \pi_x}[r(s,a)\phi(s)]$, we list the conditions we need as follow. These conditions are also adopted in [57].

Lemma 4 (Verification of assumptions) *In the context of the AC update* (21) *and* (22). *Consider the following conditions*

- *(l) For any $s \in \mathcal{S}$, $\|\phi(s)\| \leq 1$. For any $x \in \mathbb{R}^{d_0}$, there exists a constant $\lambda_1 > 0$ such that $\langle y - y', A_x(y - y')\rangle \leq -\lambda_1 \|y - y'\|^2$ for any $y, y' \in \mathbb{R}^{d_1}$. The smallest singular value of A_x is lower bounded by $\sigma > 0$.*
- *(m) There exist constants L_π, L'_π and C_π such that for any $s \in \mathcal{S}$ and $a \in \mathcal{A}$ and $x, x' \in \mathbb{R}^{d_0}$, the following inequalities hold: i) $\|\pi_x(a|s) - \pi_{x'}(a|s)\| \leq L_\pi \|x - x'\|$. ii) $\|\nabla \log \pi_x(a|s) - \nabla \log \pi_{x'}(a|s)\| \leq L'_\pi \|x - x'\|$. iii) $\|\nabla \log \pi_x(a|s)\| \leq C_\pi$.*
- *(n) For any $x \in \mathbb{R}^{d_0}$, the Markov chain induced by the policy π_x and transition kernel $\mathcal{P}$ is ergodic. There exist positive constants κ and $\rho < 1$ such that*

  $$\|\mathbb{P}_{\pi_x}(s_t \in \cdot | s_0 = s, a_0 = a) - \mu_{\pi_x}(\cdot)\|_{TV} \leq \kappa \rho^t, \forall (s,a) \in \mathcal{S} \times \mathcal{A}, \tag{110}$$

  where $\mathbb{P}_{\pi_x}(s_t \in \cdot | s_0, a_0)$ is the probability measure of the tth state s_t on the Markov chain induced by policy π_x and transition kernel $\mathcal{P}$, given the initial state and action s_0, a_0.
- *(o) The sampling protocol is: $s_k, a_k \sim d_{\pi_x}$, $s'_k \sim \mathcal{P}(\cdot | s_k, a_k)$; $\bar{s}_k \sim \mu_x$, $\bar{a}_k \sim \pi_x(\cdot | \bar{s}_k)$ and $\bar{s}'_k \sim \mathcal{P}(\cdot | \bar{s}_k, \bar{a}_k)$.*

Consider the actor critic update defined in (21) *and* (22). *Then we have:*

$$\textit{(l)–(n)} \Rightarrow \textit{Assumption 1}; \quad \textit{(l)\&(m)} \Rightarrow \textit{Assumption 2\&4}; \quad \textit{Assumption 6 holds}. \tag{111}$$

Moreover, a slightly more generalized version of Assumption 3 holds under condition (l)&(o):

$$\begin{aligned} &\mathbb{E}[\xi_k | \mathcal{F}_k^1] = 0, \ \mathbb{E}[\psi_k | \mathcal{F}_k] = 0, \\ &\|\xi_k\|^2 \leq \sigma_0^2 + \bar{\sigma}_0^2 \|y_k - y^*(x_k)\|^2, \|\psi_k\|^2 \leq \sigma_1^2 + \bar{\sigma}_1^2 \|y_k - y^*(x_k)\|^2, \end{aligned} \tag{112}$$

where $\sigma_0^2 = 8C_\pi^2(1 + 4\sigma^{-2})$, $\bar{\sigma}_0^2 = 32C_\pi^2$, $\sigma_1^2 = 32\sigma^{-2} + 8$ and $\bar{\sigma}_1^2 = 32$.

Proof. We will check the assumptions by order.

(1) Condition (l)–(n) $\Rightarrow$Assumption 1. This is shown in Lemma 5.

(2) Condition (l)&(m) $\Rightarrow$Assumption 2&4. We first check Assumption 2. In actor critic, we have $v(x) = v(x, y^*(x)) = \nabla F(x)$. By [64, Lemma 3.2], there exists a constant $L_v \coloneqq \frac{L'_\pi}{(1-\gamma)^2} + \frac{(1+\gamma)C_\pi}{(1-\gamma)^2}$ such that

$$\|\nabla F(x) - \nabla F(x')\| \leq L_v \|x - x'\|. \tag{113}$$

Then we have

$$\begin{aligned}
&\|v(x,y) - v(x,y')\| = \|\mathbb{E}[(\gamma\phi(s') - \phi(s))^\top (y - y')\nabla \log \pi_x(a|s)]\| \leq 2C_\pi \|y - y'\|,\\
&\|h(x,y) - h(x,y')\| = \|A_x(y - y')\| \leq 2\|y - y'\|.
\end{aligned} \tag{114}$$

This completes the verification of Assumption 2. Lastly, Assumption 4 is directly implied by the inequality $\langle y - y', A_x(y - y')\rangle \leq -\lambda_1 \|y - y'\|^2$ in condition (l).

(3) Assumption 6 holds. It is clear that $|F(x)| \leq \frac{1}{1-\gamma}$.

(4) Proving (112)**.** It is easy to check that $\mathbb{E}[\xi_k|\mathcal{F}_k^1] = 0,\ \mathbb{E}[\psi_k|\mathcal{F}_k] = 0$. Next we have

$$\begin{aligned}
\|\xi_k\|^2 &\leq 2\mathbb{E}\|(r(s,a) + (\gamma\phi(s') - \phi(s))^\top y_k)\nabla \log \pi_{x_k}(a|s)]\|^2\\
&\quad + 2\|(r(\bar{s}_k, \bar{a}_k) + (\gamma\phi(\bar{s}_k) - \phi(\bar{s}_k))^\top y_k)\nabla \log \pi_{x_k}(\bar{a}_k|\bar{s}_k)\|^2\\
&\leq 8C_\pi^2 + 16C_\pi^2 \|y_k\|^2\\
&\leq 8C_\pi^2 + 32C_\pi^2 \|y_k^*\|^2 + 32C_\pi^2 \|y_k - y_k^*\|^2\\
&\leq 8C_\pi^2(1 + 4\sigma^{-2}) + 32C_\pi^2 \|y_k - y_k^*\|^2 \coloneqq \sigma_0^2 + \bar{\sigma}_0^2 \|y_k - y_k^*\|^2,
\end{aligned} \tag{115}$$

where to get the last inequality we have used $\|y_k^*\| = \|A_{x_k}^{-1} b_{x_k}\| \leq \sigma^{-1}$. Similarly we have

$$\begin{aligned}
\|\psi_k\|^2 &\leq 2\mathbb{E}\|\phi(s)(\gamma\phi(s') - \phi(s))^\top y_k + r(s,a)\phi(s)\|^2\\
&\quad + 2\|\phi(s_k)(\gamma\phi(s'_k) - \phi(s_k))^\top y_k + r(s_k, a_k)\phi(s_k)\|^2\\
&\leq 16\|y_k\|^2 + 8\\
&\leq 32\|y_k - y_k^*\|^2 + 32\sigma^{-2} + 8 \coloneqq \sigma_1^2 + \bar{\sigma}_1^2 \|y_k - y_k^*\|^2.
\end{aligned} \tag{116}$$

This completes the proof. ■

We restate Theorem 3 as follows.

Theorem 5 (Restatement of Theorem 3) *Consider the sequences generated by* (21) *and* (22) *for $k = [K]$. Under conditions (l)–(o), Theorem 2 holds; that is, with $\alpha_k = \Theta(\frac{1}{\sqrt{K}})$ and $\beta_k = \Theta(\frac{1}{\sqrt{K}})$, we have*

$$\frac{1}{K}\sum_{k=1}^{K}\Big(\mathbb{E}\|\nabla F(x_k)\|^2 + \mathbb{E}\|y_k - y^*(x_k)\|^2\Big) = \mathcal{O}\Big(\frac{1}{\sqrt{K}}\Big). \tag{117}$$

We have verified the necessary assumptions for Theorem 2 to hold in Lemma 4, except that Assumption 3 needs a slight adaptation in AC. Thus the proof will be similar to that of Theorem 2, and only the steps that are different due to the adaptation of Assumption 3 will be shown here.

E.1 Analysis of the critic optimality gap

Contraction of the critic optimality gap. First we have

$$\mathbb{E}_k\|y_{k+1} - y_k^*\|^2 = \|y_k - y_k^*\|^2 + 2\beta_k \mathbb{E}_k\langle y_k - y_k^*, h(x_k, y_k) + \psi_k\rangle + \mathbb{E}_k\|y_{k+1} - y_k\|^2, \tag{118}$$

The second term in (118) can be bounded as

$$\begin{aligned}
\mathbb{E}_k\langle y_k - y_k^*, h(x_k, y_k) + \psi_k\rangle &= \langle y_k - y_k^*, h(x_k, y_k)\rangle + \langle y_k - y_k^*, \mathbb{E}_k[\psi_k]\rangle\\
&\leq -\lambda_1 \|y_k - y_k^*\|^2.
\end{aligned} \tag{119}$$

where the last inequality follows from the strong monotonicity of $h(x,y)$ and $\mathbb{E}_k[\psi_k]=0$ verified in Lemma 4.

The third term in (118) can be bounded as

$$\begin{aligned}
\mathbb{E}_k\|y_{k+1}-y_k\|^2 &= \beta_k^2\mathbb{E}_k\|h(x_k,y_k)+\psi_k\|^2\\
&= \beta_k^2(\|h(x_k,y_k)\|^2+\mathbb{E}_k\|\psi_k\|^2)\\
&\le \beta_k^2(\|h(x_k,y_k)\|^2+\sigma_1^2+\bar{\sigma}_1^2\|y_k-y_k^*\|^2)\\
&\le (L_h^2+\bar{\sigma}_1^2)\beta_k^2\|y_k-y_k^*\|^2+\sigma_1^2\beta_k^2, \qquad (120)
\end{aligned}$$

where the second last inequality follows from (112) and the last inequality follows from Assumption 2 which gives

$$\|h(x,y)\| = \|h(x,y)-\underbrace{h(x,y^*(x))}_{=0}\| \le L_h\|y-y^*(x)\|. \qquad (121)$$

Collecting the upper bounds in (119) and (120) yields

$$\begin{aligned}
\mathbb{E}_k\|y_{k+1}-y_k^*\|^2 &\le (1-2\lambda_1\beta_k+(L_h^2+\bar{\sigma}_1^2)\beta_k^2)\|y_k-y_k^*\|^2+\sigma_1^2\beta_k^2\\
&\le (1-\lambda_1\beta_k)\|y_k-y_k^*\|^2+\sigma_1^2\beta_k^2, \qquad (122)
\end{aligned}$$

where the last inequality is due to the choice of step size that satisfies $(L_h^2+\bar{\sigma}_1^2)\beta_k^2\le\lambda_1\beta_k$.

Bounding the drifting optimality gap. Next we start to bound the second term in (82) as follows

$$\begin{aligned}
&\mathbb{E}_k\langle y_k^*-y_{k+1},\alpha_k\nabla y^*(\hat{x}_{k+1})^\top\xi_k\rangle\\
&=\mathbb{E}_k\langle y_k^*-y_{k+1},\alpha_k\big(\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\big)^\top\xi_k\rangle+\mathbb{E}_k\langle y_k^*-y_{k+1},\alpha_k\nabla y^*(x_k)^\top\mathbb{E}_k[\xi_k|\mathcal{F}_k^1]\rangle\\
&=\mathbb{E}_k\langle y_k^*-y_{k+1},\alpha_k\big(\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\big)^\top\xi_k\rangle\\
&\le\alpha_k\mathbb{E}_k\big[\|y_k^*-y_{k+1}\|\|\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\|\|\xi_k\|\big]\\
&\le\sigma_0\alpha_k\mathbb{E}_k\big[\|y_k^*-y_{k+1}\|\|\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\|\big]\\
&\quad+\bar{\sigma}_0\alpha_k\mathbb{E}_k\big[\|y_k^*-y_{k+1}\|\|\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\|\|y_k-y_k^*\|\big] \qquad (123)
\end{aligned}$$

where the second inequality follows from $\mathbb{E}_k[\xi_k|\mathcal{F}_k^1]=0$ shown in Lemma 4, and the last inequality follows from (112).

The first term in the RHS of (123) can be bounded as

$$\begin{aligned}
&\mathbb{E}_k\big[\|y_k^*-y_{k+1}\|\|\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\|\big]\\
&\le L_{y'}\mathbb{E}_k[\|y_k^*-y_{k+1}\|\|x_{k+1}-x_k\|]\\
&\le L_{y'}\alpha_k\big(\mathbb{E}_k[\|y_k^*-y_{k+1}\|\|v(x_k,y_k)\|]+\mathbb{E}_k[\|y_k^*-y_{k+1}\|\|\xi_k\|]\big)\\
&\le\frac{1}{2}L_{y'}\alpha_k\big(\mathbb{E}_k\|y_k^*-y_{k+1}\|^2+\|v(x_k,y_k)\|^2+\mathbb{E}_k\|y_k^*-y_{k+1}\|^2+\|\xi_k\|^2\big)\\
&\le\frac{1}{2}L_{y'}\alpha_k\big(2\mathbb{E}_k\|y_k^*-y_{k+1}\|^2+\|v(x_k,y_k)\|^2+\sigma_1^2+\bar{\sigma}_1^2\|y_k-y_k^*\|^2\big) \qquad (124)
\end{aligned}$$

where the first inequality follows from Lemma 5 and the last inequality follows from (112).

The second term in (123) can be bounded as

$$\begin{aligned}
\mathbb{E}_k\big[\|y_k^*-y_{k+1}\|\|\nabla y^*(\hat{x}_{k+1})-\nabla y^*(x_k)\|\|y_k-y_k^*\|\big]&\le 2L_y\mathbb{E}_k\big[\|y_k^*-y_{k+1}\|\|y_k-y_k^*\|\big]\\
&\le L_y\mathbb{E}_k\|y_k^*-y_{k+1}\|^2+L_y\|y_k-y_k^*\|^2. \qquad (125)
\end{aligned}$$

Substituting (124) and (125) into (123), then substituting (123) and (83) into (82) gives

$$\begin{aligned}
\mathbb{E}_k\langle y_k^*-y_{k+1},y_{k+1}^*-y_k^*\rangle &\le \big(L_y\big(\frac{L_{v,y}}{2}+2L_y+\bar{\sigma}_0\big)\alpha_k+L_{y'}\sigma_0\alpha_k^2\big)\mathbb{E}_k\|y_{k+1}-y_k^*\|^2\\
&\quad+\frac{1}{2}\big(L_y(L_{v,y}+\bar{\sigma}_0)\alpha_k+L_{y'}\sigma_0\bar{\sigma}_1^2\alpha_k^2\big)\|y_k-y_k^*\|^2\\
&\quad+\big(\frac{1}{8}\alpha_k+\frac{1}{2}L_{y'}\sigma_0\alpha_k^2\big)\|v(x_k)\|^2+\frac{1}{2}L_{y'}\sigma_0\sigma_1^2\alpha_k^2. \qquad (126)
\end{aligned}$$

The last term in (81) can be bounded as

$$\mathbb{E}_k\|y^*_{k+1}-y^*_k\|^2 \le L_y^2\alpha_k^2\mathbb{E}_k\|v(x_k,y_k)+\xi_k\|^2 = L_y^2\alpha_k^2(\|v(x_k,y_k)\|^2+\mathbb{E}_k\|\xi_k\|^2)$$
$$\le L_y^2\alpha_k^2(\|v(x_k,y_k)\|^2+\sigma_0^2+\bar{\sigma}_0^2\|y_k-y_k^*\|^2), \tag{127}$$

where the last inequality follows from (112). Substituting (126) and (127) into (81) gives

$$\mathbb{E}_k\|y_{k+1}-y^*_{k+1}\|^2 \le \big(1+L_y\big(L_{v,y}+4L_y+2\bar{\sigma}_0\big)\alpha_k+2L_{y'}\sigma_0\alpha_k^2\big)\mathbb{E}_k\|y_{k+1}-y_k^*\|^2$$
$$+\big(L_y(L_{v,y}+\bar{\sigma}_0)\alpha_k+(L_{y'}\sigma_0\bar{\sigma}_1^2+L_y^2\bar{\sigma}_0^2)\alpha_k^2\big)\|y_k-y_k^*\|^2$$
$$+\big(\frac{1}{4}\alpha_k+(L_{y'}\sigma_0+L_y^2)\alpha_k^2\big)\|v(x_k,y_k*)\|^2+(\sigma_0\sigma_1^2+L_y^2\sigma_0^2)\alpha_k^2. \tag{128}$$

E.2 Analysis of the actor sequence

Analysis of main sequence. The second term in (94) is instead bounded as

$$\mathbb{E}_k\langle v(x_k,y_k),\alpha_k\xi_k\rangle = \langle v(x_k,y_k),\alpha_k\mathbb{E}_k[\xi_k]\rangle = 0. \tag{129}$$

Then the last term in (94) is instead bounded as

$$\mathbb{E}_k\|x_{k+1}-x_k\|^2 = \alpha_k^2(\|v(x_k,y_k)\|^2+\mathbb{E}_k\|\xi_k\|^2) \le \alpha_k^2(\|v(x_k,y_k)\|^2+\sigma_0^2+\bar{\sigma}_0^2\|y_k-y_k^*\|^2). \tag{130}$$

Substituting the bounds in (95), (129) and (130) into (94) yields

$$\mathbb{E}_k[F(x_{k+1})]-F(x_k) \ge (\frac{3\alpha_k}{4}-\frac{L_v}{2}\alpha_k^2)\|v(x_k,y_k^*)\|^2-(L_{v,y}^2\alpha_k+\frac{L_{v,y}}{2}\bar{\sigma}_0^2\alpha_k^2)\|y_k-y_k^*\|^2-\frac{L_v\sigma_0^2}{2}\alpha_k^2. \tag{131}$$

Establishing convergence. Recall that the Lyapunov function $\mathcal{L}_k = -F(x_k)+\|y_k-y_k^*\|^2$. With the bounds in (122), (128) and (131), we have

$$\mathbb{E}_k[\mathcal{L}_{k+1}]-\mathcal{L}_k \le \big(-\frac{1}{2}\alpha_k+(\frac{L_v}{2}+L_{y'}\sigma_0L_y^2)\alpha_k^2\big)\|v(x_k,y_k^*)\|^2$$
$$+\big((1+C_0'\alpha_k+C_1'\alpha_k^2)(1-\lambda_1\beta_k)-1+C_2'\alpha_k+C_3'\alpha_k^2\big)\|y_k-y_k^*\|^2$$
$$+\Theta(\alpha_k^2+(1+\alpha_k+\alpha_k^2)\beta_k^2), \tag{132}$$

where $C_0' \coloneqq L_y\big(L_{v,y}+4L_y+2\bar{\sigma}_0\big)$, $C_1' \coloneqq 2L_{y'}\sigma_0$, $C_2' \coloneqq L_y(L_{v,y}+\bar{\sigma}_0)+L_{v,y}^2$, $C_3' \coloneqq L_{y'}\sigma_0\bar{\sigma}_1^2+\frac{L_{v,y}}{2}\bar{\sigma}_0^2$. Notice that (132) takes a similar form to that of (101) ($N=1$).

If the step sizes are chosen such that

$$-\frac{1}{2}\alpha_k+(\frac{L_v}{2}+L_{y'}\sigma_0L_y^2)\alpha_k^2 \le -\frac{1}{4}\alpha_k,$$
$$(1+C_0'\alpha_k+C_1'\alpha_k^2)(1-\lambda_1\beta_k)-1+C_2'\alpha_k+C_3'\alpha_k^2 \le -\lambda_1\alpha_k, \tag{133}$$

then it follows from the derivation after (101) that Theorem 2 holds for AC update.

E.3 Supporting lemmas for Theorem 3

Lemma 5 (Complete version of Lemma 2) *Consider the AC update in* (21)-(22). *Under conditions (l)–(n), there exist constants $L_y, L_{y'}$ such that*

$$\|y^*(x)-y^*(x')\| \le L_y\|x-x'\|,$$
$$\|\nabla y^*(x)-\nabla y^*(x')\| \le L_{y'}\|x-x'\|. \tag{134}$$

Proof. Under condition (l), we have $y^*(x) = -A_x^{-1}b_x$. With Lemma 7 and $\|A_x^{-1}\| \le \sigma^{-1}$, $\|b_x\| \le 1$, applying Lemma 14 to $y^*(x)$ implies that it is Lipschitz continuous with modulus

$$L_y \coloneqq (\sigma^{-1}+2\sigma^{-2})L'_\mu.$$

We next verify the Lipschitz continuity of $\nabla y^*(x)$. For $x \in \mathbb{R}^d$ and $f : \mathbb{R}^d \mapsto \mathbb{R}^{d_1 \times d_2}$, we denote $[x]_i$ as the ith element of x and we use $\nabla_i f(x) \coloneqq \frac{\partial f(x)}{\partial [x]_i}$. Then we have

$$\nabla_i y^*(x) = A_x^{-1} \nabla_i A_x A_x^{-1} b_x - A_x^{-1} \nabla_i b_x = -A_x^{-1} \nabla_i A_x y^*(x) - A_x^{-1} \nabla_i b_x. \tag{135}$$

By Lemma 14, to prove $\nabla_i y^*(x)$ is Lipschitz continuous w.r.t. x, it suffices to prove $\nabla_i A_x$, $y^*(x)$, $\nabla_i b_x$ and A_x^{-1} are bounded (in norm) and Lipschitz continuous. First we have

$$\|A_x^{-1}\| \leq \sigma^{-1}, \|y^*(x)\| \leq \|A_x^{-1}\| \|b_x\| \leq \sigma^{-1}. \tag{136}$$

And by Lemma 7, we have

$$\|A_x^{-1} - A_{x'}^{-1}\| \leq 2\sigma^{-2} L'_\mu \|x - x'\|. \tag{137}$$

Thus it suffices to prove $\nabla_i A_x$ and $\nabla_i b_x$ are bounded in norm and Lipschitz continuous.

We start by

$$\nabla_i b_x = \mathbb{E}_{s \sim \mu_{\pi_x}, a \sim \pi_x(\cdot|s)} [\nabla_i \log \pi_x(a|s) G_x(s,a)], \tag{138}$$

where $G_x(s,a) \coloneqq \mathbb{E}_{\pi_x}[\sum_{t=0}^{\infty} \big(r(s_t,a_t)\phi(s_t) - b_x\big)|s_0 = s, a_0 = a]$. By letting $\hat{r}(s,a,s') = r(s,a)\phi(s)$ in Lemma 8, we have

$$\|G_x(s,a)\| \leq C_G,\ \|G_x(s,a) - G_{x'}(s,a)\| \leq L_G \|x - x'\|, \tag{139}$$

where $C_G \coloneqq 2 + \frac{\rho\kappa}{1-\rho}$ and $L_G \coloneqq L'_\mu + \frac{\rho\kappa L_\pi |\mathcal{A}|}{1-\rho} + \big(\frac{\kappa}{1-\rho} + 1\big)^2 (L_\pi |\mathcal{A}| + L_\mu) + L_\mu$. Then we have $\|\nabla_i b_x\|$ can be bounded as

$$\|\nabla_i b_x\| \leq C_\pi \mathbb{E}_{s \sim \mu_x, a \sim \pi_x(\cdot|s)}[\|G_x(s,a)\|] \leq C_\pi C_G, \tag{140}$$

Now we start to prove the Lipschitz continuity of $\nabla_i b_x$. First we have

$$\begin{aligned}
&\|\nabla_i b_x - \nabla_i b_{x'}\| \\
&\leq \big\|\mathbb{E}_{s \sim \mu_{\pi_x}, a \sim \pi_x}[\nabla_i \log \pi_x(a|s) G_x(s,a)] - \mathbb{E}_{s \sim \mu_{\pi_{x'}}, a \sim \pi_{x'}}[\nabla_i \log \pi_x(a|s) G_x(s,a)]\big\| \\
&\quad + \mathbb{E}_{s \sim \mu_{\pi_{x'}}, a \sim \pi_{x'}} \big\|\nabla_i \log \pi_x(a|s) G_x(s,a) - \nabla_i \log \pi_{x'}(a|s) G_{x'}(s,a)\big\| \\
&\leq \|\mu_{\pi_x} \cdot \pi_x - \mu_{\pi_{x'}} \cdot \pi_{x'}\|_{TV} \sup \|\nabla \log \pi_x(a|s) G_x(s,a)\| \\
&\quad + \mathbb{E}_{s \sim \mu_{\pi_{x'}}, a \sim \pi_{x'}} \|\nabla_i \log \pi_x(a|s) G_x(s,a) - \nabla_i \log \pi_{x'}(a|s) G_{x'}(s,a)\| \\
&\leq C_G L'_\mu \|x - x'\| + (C_\pi L_G + L_\pi C_G)\|x - x'\| \coloneqq L'_b \|x - x'\|,
\end{aligned} \tag{141}$$

where the $\mu_{\pi_x} \cdot \pi_x$ denotes the probability measure specified by the probability function $(\mu_{\pi_x} \cdot \pi_x)(s,a) = \mu_{\pi_x}(s)\pi_x(a|s)$. In the second inequality, we apply Lemma 6 to the first term; and for the second term, we apply Lemma 14 along with (139) and condition (m).

For $\nabla_i A_x$, we have

$$\nabla_i A_x = \mathbb{E}_{s \sim \mu_{\pi_x}, a \sim \pi_x}[\nabla_i \log \pi_x(a|s) G_x(s,a)], \tag{142}$$

where we slightly abuse the notation and define $G_x(s,a) \coloneqq \mathbb{E}_{\pi_x}\big[\sum_{t=0}^{\infty} \big(\phi(s_t)\big(\gamma\phi(s_{t+1}) - \phi(s_t)\big)^\top - A_x\big)|s_0 = s, a_0 = a\big]$. Observing that $\nabla_i A_x$ has similar structure as that of $\nabla_i b_x$, we can apply the same technique and obtain

$$\begin{aligned}
\|\nabla_i A_x\| &\leq C_\pi C'_G, \\
\|\nabla_i A_x - \nabla_i A_{x'}\| &\leq C'_G L'_\mu \|x - x'\| + (C_\pi L'_G + L_\pi C'_G)\|x - x'\| \coloneqq L'_A \|x - x'\|,
\end{aligned} \tag{143}$$

where $C'_G \coloneqq 4 + \frac{2\rho\kappa}{1-\rho}$ and $L'_G \coloneqq 2L'_\mu + \frac{\rho\kappa L_\pi |\mathcal{A}|}{1-\rho} + \big(\frac{\kappa}{1-\rho} + 1\big)^2 (L_\pi|\mathcal{A}| + L_\mu) + L_\mu$.

Finally, applying Lemma 14 to (135) with (136), (137), (140), (141) and (143) yields

$$\|\nabla_i y^*(x) - \nabla_i y^*(x')\| \leq L_{y'} \|x - x'\|, \tag{144}$$

where $L_{y'} \coloneqq 2\sigma^{-3} L'_\mu C_\pi C'_G + L'_A \sigma^{-2} + L_y C_\pi C'_G \sigma^{-1} + 2\sigma^{-2} L'_\mu C_\pi C_G + \sigma^{-1} L'_b$. This completes the proof. ■

Lemma 6 *[66, Lemma 3] Define $(\mu_{\pi_x}\cdot\pi_x)(s,a) \coloneqq \mu_{\pi_x}(s)\pi_x(a|s)$. Under conditions (n) and (m), it holds that*

$$\|\mu_{\pi_x}-\mu_{\pi_{x'}}\|_{TV}\leq L_\mu\|x-x'\|,\ \|\mu_{\pi_x}\cdot\pi_x-\mu_{\pi_{x'}}\cdot\pi_{x'}\|_{TV}\leq L'_\mu\|x-x'\| \tag{145}$$

where $L_\mu \coloneqq 2L_\pi|\mathcal{A}|(\log_\rho\kappa^{-1}+\frac{1}{1-\rho})$ and $L'_\mu \coloneqq L_\mu+2L_\pi|\mathcal{A}|$.

Lemma 7 *Define $\mu_{\pi_x}\cdot\pi_x(s,a) \coloneqq \mu_{\pi_x}(s)\pi_x(a|s)$. Under conditions (n) and (m), the following inequalities hold*

$$\|A_x-A_{x'}\|\leq 2L'_\mu\|x-x'\|,\ \|A_x^{-1}-A_{x'}^{-1}\|\leq 2\sigma^{-2}L'_\mu\|x-x'\|,\ \|b_x-b_{x'}\|\leq L'_\mu\|x-x'\| \tag{146}$$

where $L'_\mu=2L_\pi|\mathcal{A}|(1+\log_\rho\kappa^{-1}+\frac{1}{1-\rho})$.

Proof. First we have

$$\|b_x-b_{x'}\|\leq\|\mu_{\pi_x}\cdot\pi_x-\mu_{\pi_{x'}}\cdot\pi_{x'}\|_{TV}\sup_{s,a}\|r(s,a)\phi(s)\|\leq L'_\mu\|x-x'\|, \tag{147}$$

where the last inequality follows from Lemma 6. And similarly, we have

$$\|A_x-A_{x'}\|\leq 2L'_\mu\|x-x'\|. \tag{148}$$

Finally, we have

$$\begin{aligned}&\|A_x^{-1}-A_{x'}^{-1}\|=\|A_{x'}^{-1}(A_x-A_{x'})A_x^{-1}\|\leq\sigma^{-2}\|A_x-A_{x'}\|\\&\leq\sigma^{-2}\|\mu_{\pi_x}\cdot\pi_x-\mu_{\pi_{x'}}\cdot\pi_{x'}\|_{TV}\sup_{s,s'}\|\phi(s)(\gamma\phi(s')-\phi(s))\|\leq 2\sigma^{-2}L'_\mu\|x-x'\|,\end{aligned} \tag{149}$$

where the last inequality follows from Lemma 6. This completes the proof. ■

Lemma 8 *Suppose conditions (l)–(n) hold. With mapping $\hat{r}:\mathcal{S}\times\mathcal{A}\times\mathcal{S}\mapsto\mathbb{R}^{d\times d'}$ such that $\|r(s,a,s')\|\leq C_r$ for any (s,a,s'), define*

$$\begin{aligned}G_x(s,a)&\coloneqq\mathbb{E}_{\substack{a_t\sim\pi_x(\cdot|s_t)\\ s_{t+1}\sim\mathcal{P}(\cdot|s_t,a_t)}}\Big[\sum_{t=0}^{\infty}\big(\hat{r}(s_t,a_t,s_{t+1})-\bar{r}_x\big)\big|s_0=s,a_0=a\Big],\\ \text{with}\quad\bar{r}_x&\coloneqq\mathbb{E}_{\substack{s\sim\mu_{\pi_x},a\sim\pi_x(\cdot|s)\\ s'\sim\mathcal{P}(\cdot|s,a)}}[\hat{r}(s,a,s')].\end{aligned} \tag{150}$$

Then there exists a constant L_G such that for any $(s,a)\in\mathcal{S}\times\mathcal{A}$ and $x,x'\in\mathbb{R}^d$, the following inequalities hold

$$\begin{aligned}\|G_x(s,a)-G_{x'}(s,a)\|&\leq L_G\|x-x'\|,\\ \|G_x(s,a)\|&\leq 2C_r+\frac{C_r\rho\kappa}{1-\rho}.\end{aligned} \tag{151}$$

Proof. We write $G_x(s_0,a_0)$ as:

$$\begin{aligned}G_x(s_0,a_0)=\mathbb{E}_{s_1\sim\mathcal{P}}[\hat{r}(s_0,a_0,s_1)]-\bar{r}_x+\sum_{t=1}^{\infty}\Big(&\sum_{(s,a)\in\mathcal{S}\times\mathcal{A}}\mathbf{Pr}_{\pi_x}(s_t=s|s_0,a_0)\pi_x(a|s)\mathbb{E}_{s'\sim\mathcal{P}}[\hat{r}(s,a,s')]\\&-\sum_{(s,a)\in\mathcal{S}\times\mathcal{A}}\mu_{\pi_x}(s)\pi_x(a|s)\mathbb{E}_{s'\sim\mathcal{P}}[\hat{r}(s,a,s')]\Big).\end{aligned} \tag{152}$$

Given (s_0,a_0), define the vector $p_1\coloneqq[\mathcal{P}(s^{(0)}|s_0,a_0),\mathcal{P}(s^{(1)}|s_0,a_0),...,\mathcal{P}(s^{(|\mathcal{S}|)}|s_0,a_0)]$ where $s^{(0)},...,s^{(|\mathcal{S}|)}$ are states in $\mathcal{S}$. Given π_x, define the following state transition matrix

$$P_{\pi_x}\coloneqq\begin{bmatrix}\mathcal{P}_{\pi_x}(s^{(0)}|s^{(0)}) & \mathcal{P}_{\pi_x}(s^{(1)}|s^{(0)}) & \dots & \mathcal{P}_{\pi_x}(s^{(|\mathcal{S}|)}|s^{(0)})\\ \vdots & & & \\ \mathcal{P}_{\pi_x}(s^{(0)}|s^{(|\mathcal{S}|)}) & \mathcal{P}_{\pi_x}(s^{(1)}|s^{(|\mathcal{S}|)}) & \dots & \mathcal{P}_{\pi_x}(s^{(|\mathcal{S}|)}|s^{(|\mathcal{S}|)})\end{bmatrix}, \tag{153}$$

where $\mathcal{P}_{\pi_x}(s'|s) = \sum_{a\in\mathcal{A}} \mathcal{P}(s'|s,a)\pi_x(a|s)$. Then it is clear that we can write the probability function $\mathbf{Pr}_{\pi_x}(s_t = \cdot|s_0, a_0)$ as its vector form $p_1 P_{\pi_x}^{t-1}$. We slightly abuse the notation and use $[p_1 P_{\pi_x}^t]_s = \mathbf{Pr}_{\pi_x}(s_t = s|s_0, a_0)$. Then (152) can be rewritten as

$$
\begin{aligned}
G_x(s_0,a_0) = \mathbb{E}_{s_1\sim\mathcal{P}}[\hat{r}(s_0,a_0,s_1)] - \bar{r}_x + \sum_{t=0}^{\infty}\Big(&\sum_{(s,a)\in\mathcal{S}\times\mathcal{A}} [p_1 P_{\pi_x}^t]_s \pi_x(a|s)\mathbb{E}_{s'\sim\mathcal{P}}[\hat{r}(s,a,s')] \\
&- \sum_{s,a}[p_1 P_{\pi_x}^{\infty}]_s \pi_x(a|s)\mathbb{E}_{s'\sim\mathcal{P}}[\hat{r}(s,a,s')]\Big) \\
= \mathbb{E}_{s_1\sim\mathcal{P}}[\hat{r}(s_0,a_0,s_1)] - \bar{r}_x + \sum_{t=0}^{\infty}\sum_{(s,a)} &([p_1 P_{\pi_x}^t]_s - [p_1 P_{\pi_x}^{\infty}]_s)\pi_x(a|s)\mathbb{E}_{s'\sim\mathcal{P}}[\hat{r}(s,a,s')] \\
= \mathbb{E}_{s_1\sim\mathcal{P}}[\hat{r}(s_0,a_0,s_1)] - \bar{r}_x + \sum_{(s,a)} &[p_1 Y_x]_s \pi_x(a|s)\mathbb{E}_{s'\sim\mathcal{P}}[\hat{r}(s,a,s')], \qquad (154)
\end{aligned}
$$

where $Y_x := \sum_{t=0}^{\infty}(P_{\pi_x}^t - P_{\pi_x}^{\infty})$. Then $\|G_x(s,a)\|$ can be bounded as follows

$$
\begin{aligned}
\|G_x(s,a)\| &\leq 2C_r + C_r\sum_{s,a}|[p_1 Y_x]_s|\pi_x(a|s) \\
&\leq 2C_r + C_r\sum_{s}|[p_1 Y_x]_s| \\
&\leq 2C_r + \frac{C_r\rho\kappa}{1-\rho} := C_G, \qquad (155)
\end{aligned}
$$

where the last inequality follows from condition (n) and

$$
\begin{aligned}
\sum_s |[p_1 Y_x]_s| &\leq \sum_{t=1}^{\infty}\sum_s |\mathbf{Pr}_{\pi_x}(s_t = s|s_0,a_0) - \mu_{\pi_x}(s)| \\
&= \sum_{t=1}^{\infty}\|\mathbb{P}_{\pi_x}(s_t\in\cdot|s_0,a_0) - \mu_{\pi_x}(\cdot)\|_{TV} \leq \frac{\rho\kappa}{1-\rho}. \qquad (156)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\|G_x(s,a) - G_{x'}(s,a)\| \\
&\leq \|\bar{r}_x - \bar{r}_{x'}\| + C_r\sum_{s,a}\big|[p_1 Y_x]_s\big|\big\|\pi_x(a|s) - \pi_{x'}(a|s)\big\| + C_r\sum_{s,a}\big|[p_1(Y_x - Y_{x'})]_s\big|\pi_{x'}(a|s) \\
&\leq \|\bar{r}_x - \bar{r}_{x'}\| + \sum_s\big|[p_1 Y_x]_s\big| L_\pi|\mathcal{A}|\|x - x'\| + \|p_1(Y_x - Y_{x'})\|_1 \\
&\leq \|\bar{r}_x - \bar{r}_{x'}\| + \frac{\rho\kappa L_\pi|\mathcal{A}|}{1-\rho}\|x - x'\| + \|Y_x - Y_{x'}\|_\infty \qquad (157)
\end{aligned}
$$

where the last inequality follows from (156). The first term in (157) can be bounded as

$$
\|\bar{r}_x - \bar{r}_{x'}\| \leq \|\mu_x\cdot\pi_x - \mu_{x'}\cdot\pi_{x'}\|_{TV}\sup_{s,a,s'}\|r(s,a,s')\| \leq C_r L'_\mu, \qquad (158)
$$

where the last inequality follows from Lemma 6. By [40, Theorem 2.5], we have $Y_x + P_{\pi_x}^{\infty} = (I - P_{\pi_x} + P_{\pi_x}^{\infty})^{-1}$. First note that

$$
\begin{aligned}
\|(I - P_{\pi_x} + P_{\pi_x}^{\infty})^{-1}\|_\infty &\leq \|Y_x\|_\infty + \|P_{\pi_x}^{\infty}\|_\infty \\
&\leq \sum_{t=0}^{\infty}\|P_{\pi_x}^t - P_{\pi_x}^{\infty}\|_\infty + 1 \\
&= \sum_{t=0}^{\infty}\max_{s_0\in\mathcal{S}}\sum_s|\mathbf{Pr}_{\pi_x}(s_t = s|s_0) - \mu_{\pi_x}(s)| + 1 \leq \frac{\kappa}{1-\rho} + 1, \qquad (159)
\end{aligned}
$$

where the last inequality follows from condition (n).

We also have

$$
\begin{aligned}
&\|(I-P_{\pi_x}+P^\infty_{\pi_x})^{-1}-(I-P_{\pi_{x'}}+P^\infty_{\pi_{x'}})^{-1}\|_\infty\\
&\leq \|(I-P_{\pi_x}+P^\infty_{\pi_x})^{-1}\|_\infty\|P_{\pi_x}-P_{\pi_{x'}}+P^\infty_{\pi_{x'}}-P^\infty_{\pi_x}\|_\infty\|(I-P_{\pi_{x'}}+P^\infty_{\pi_{x'}})^{-1}\|_\infty\\
&\overset{(159)}{\leq} \big(\frac{\kappa}{1-\rho}+1\big)^2\big(\|P_{\pi_x}-P_{\pi_{x'}}\|_\infty+\|P^\infty_{\pi_{x'}}-P^\infty_{\pi_x}\|_\infty\big)\\
&\leq \big(\frac{\kappa}{1-\rho}+1\big)^2(L_\pi|\mathcal{A}|+L_\mu)\|x-x'\| \qquad (160)
\end{aligned}
$$

where in the last inequality we have used

$$
\begin{aligned}
\|P_{\pi_x}-P_{\pi_{x'}}\|_\infty &= \max_s\sum_{s'}|\sum_a\pi_x(a|s)\mathcal{P}(s'|s,a)-\sum_a\pi_{x'}(a|s)\mathcal{P}(s'|s,a)|\\
&= \max_s|\sum_a\pi_x(a|s)-\sum_a\pi_{x'}(a|s)|\sum_{s'}\mathcal{P}(s'|s,a)\\
&\leq \max_s\sum_a|\pi_x(a|s)-\pi_{x'}(a|s)|\leq L_\pi|\mathcal{A}|\|x-x'\|,\\
\|P^\infty_{\pi_{x'}}-P^\infty_{\pi_{x'}}\|_\infty &= \|\mu_{\pi_x}-\mu_{\pi_{x'}}\|_{TV}\leq L_\mu\|x-x'\| \text{ (Lemma 6).} \qquad (161)
\end{aligned}
$$

With (160) and (161), we can write

$$
\begin{aligned}
\|Y_x-Y_{x'}\|_\infty &\leq \|P^\infty_{\pi_x}-P^\infty_{\pi_{x'}}\|_\infty+\|(I-P_{\pi_x}+P^\infty_{\pi_x})^{-1}-(I-P_{\pi_{x'}}+P^\infty_{\pi_{x'}})^{-1}\|_\infty\\
&\leq \Big(\big(\frac{\kappa}{1-\rho}+1\big)^2(L_\pi|\mathcal{A}|+L_\mu)+L_\mu\Big)\|x-x'\|. \qquad (162)
\end{aligned}
$$

Substituting (158) and (162) into (157) gives

$$
\|G_x(s,a)-G_{x'}(s,a)\|\leq\Big(C_rL'_\mu+\frac{\rho\kappa L_\pi|\mathcal{A}|}{1-\rho}+\big(\frac{\kappa}{1-\rho}+1\big)^2(L_\pi|\mathcal{A}|+L_\mu)+L_\mu\Big)\|x-x'\|. \qquad (163)
$$

This completes the proof. ■

F Proof of Lemma 3 and Corollary 2

Here we prove Lemma 3 which along with the generic Theorem 1 and 2 implies Corollary 2.

Proof. We will verify the assumptions by order.

(1) Condition (g) ⇒ Assumption 1. Note that $y^{n,*}(y^{n-1})=f^{n-1}(y^{n-1})$, then (g) directly implies Assumption 1 holds.

(2) Condition (g) ⇒ Assumption 2. First note

$$
\begin{aligned}
v(x) &= v\big(x,y^{1,*}(x),y^{2,*}(y^{1,*}(x)),\ldots,y^{N,*}(\ldots y^{2,*}(y^{1,*}(x))\ldots)\big)\\
&= v\big(x,f^0(x),f^1(f^0(x)),\ldots,f^{N-1}(\ldots f^1(f^0(x))\ldots)\big)\\
&= \nabla f^0(x)\nabla f^1(f^0(x))\cdots\nabla f^N(f^{N-1}(\ldots f^1(f^0(x))\ldots)). \qquad (164)
\end{aligned}
$$

By Lemma 14, in order for $v(x)$ to be Lipschitz continuous, it suffices to let $\nabla f^n(x)$ be bounded and Lipschitz continuous for every $n=0,1,\ldots,N$. This is satisfied under condition (g).

Now in order for $v(x,y^1,y^2,\ldots,y^N)$ be Lipschitz continuous w.r.t. $y^1,y^2,\ldots,y^N$, it again suffices to let $\nabla f^n(x)$ be bounded and Lipschitz continuous for every $n=0,1,\ldots,N$, which is satisfied under condition (g).

Finally, the Lipschitz continuity of $h^n(y^{n-1},y^n)$ w.r.t. y^n is directly implied by condition (g).

(3) Condition (h) and (i)⇒ Assumption 3. First we have

$$
\begin{aligned}
\mathbb{E}[\xi_k|\mathcal{F}_k^1] &= -v(x_k,y_k^1,\ldots,y_k^N)+\mathbb{E}[\nabla f^0(x_k;\hat{\zeta}_k^0)\cdots\nabla f^N(y_k^N;\zeta_k^N)|\mathcal{F}_k^1]\\
&= -v(x_k,y_k^1,\ldots,y_k^N)+\mathbb{E}[\nabla f^0(x_k;\hat{\zeta}_k^0)\cdots\nabla f^N(y_k^N;\zeta_k^N)|\mathcal{F}_k]\\
&= -v(x_k,y_k^1,\ldots,y_k^N)+\nabla f^0(x_k)\cdots\nabla f^N(y_k^N)=0, \qquad (165)
\end{aligned}
$$

where we have used the condition that $\hat{\zeta}_k^0, \zeta_k^0, \zeta_k^1, \ldots, \zeta_k^N$ are conditionally independent of each other given $\mathcal{F}_k$. The same goes for ψ_k^n that

$$\mathbb{E}[\psi_k^n|\mathcal{F}_k^{n+1}] = -h^n(y_k^{n-1}, y_k^n) + \mathbb{E}[f^{n-1}(y_k^{n-1}, \zeta_k^{n-1})|\mathcal{F}_k] - y_k^n = 0. \tag{166}$$

The bounded variance condition directly implies that $\mathbb{E}[\|\psi_k^n\|^2|\mathcal{F}_k^{n+1}] < \infty$. Now for ξ_k we have

$$\begin{aligned}
&\mathbb{E}[\|\xi_k\|^2|\mathcal{F}_k^1]\\
&= \mathbb{E}[\|\xi_k\|^2|\mathcal{F}_k]\\
&= \mathbb{E}_k\|\nabla f^0(x)\nabla f^1(y^1)\cdots\nabla f^N(y^N) - \nabla f^0(x_k;\hat{\zeta}_k^0)\cdots\nabla f^N(y_k^N;\zeta_k^N)\|^2\\
&= \mathbb{E}_k\|\nabla f^0(x_k;\hat{\zeta}_k^0)\|^2 ... \mathbb{E}_k\|\nabla f^N(y_k^N;\zeta_k^N)\|^2 - \|\nabla f^0(x)\|^2\|\nabla f^1(y^1)\|^2 ... \|\nabla f^N(y^N)\|^2, \quad (167)
\end{aligned}$$

which is bounded by a constant since under contion (h), we have $\mathbb{E}_k\|\nabla f^n(x_k;\zeta_k^n)\|^2 < \infty$ for any n.

(4) Condition (j) $\Rightarrow$ Assumption 5, (k) $\Rightarrow$ Assumption 6. These assumptions are directly implied by the conditions.

(5) Verifying Assumption 4. By plugging in $y^{n,*}(y^{n-1}) = f^{n-1}(y^{n-1})$, it is immediate that Assumption 4 holds with $\lambda_n = 1$ for any $n \in [N]$. ■

G Proof of Theorem 4

Before we prove the result, we first give a lemma that establishes the connection between Theorem 4 and the generic Theorem 2.

Lemma 9 *In the context of the MAMPG update in* (32) *and* (33)*. Consider the following conditions:*

(p) There exist constants L_π, L'_π, L''_π and C_π such that for any $(s,a) \in \mathcal{S}\times\mathcal{A}$ and $x, x' \in \mathbb{R}^{d_0}$, we have: i) $\|\pi_x(a|s) - \pi_{x'}(a|s)\| \leq L_\pi\|x - x'\|$; ii) $\|\nabla\log\pi_x(a|s) - \nabla\log\pi_{x'}(a|s)\| \leq L'_\pi\|x - x'\|$; iii) $\|\nabla^2\log\pi_x(a|s) - \nabla^2\log\pi_{x'}(a|s)\| \leq L''_\pi\|x - x'\|$ and iv) $\|\nabla\log\pi_x(a|s)\| \leq C_\pi$.

(q) Given $\mathcal{F}_k$, we have for any $n \in \{1,\ldots,N\}$ and $i \in \{1,2,...,M\}$: $f_i^n(y_{k,i}^{n-1};\zeta_{k,i}^n)$ and $\nabla f_i^n(y_{k,i}^n;\zeta_{k,i}^n)$ are respectively the unbiased estimators of $f_i^n(y_{k,i}^{n-1})$ and $\nabla f_i^n(y_{k,i}^n)$ with bounded variance. Likewise, $f_i^0(x_k;\zeta_{k,i}^0)$ and $\nabla f_i^0(x_k;\hat{\zeta}_{k,i}^0)$ are respectively unbiased estimators of $f_i^0(x_k)$ and $\nabla f_i^0(x_k)$ with bounded variance.

(r) Given $\mathcal{F}_k$, $\hat{\zeta}_{k,i}^0, \zeta_{k,i}^0, \zeta_{k,i}^1, \ldots, \zeta_{k,i}^N$ are conditionally independent for $i = 1,2,...,M$.

We use $a \Rightarrow b$ to indicate that a is a sufficient condition of b. Then we have

$$\begin{gathered}
(p) \Rightarrow \textit{Assumption 1\& 2}; \quad (q)\&(r) \Rightarrow \textit{Assumption 3};\\
\textit{Assumption 4 holds naturally for } (32); \quad \textit{Assumption 6 holds under bounded reward}.
\end{gathered} \tag{168}$$

Condition (p) is a standard assumption commonly adopted in the literature; see e.g., [14]. It is satisfied with certain popular policy parameterization such as the softmax policy. Conditions (q)&(r) can be satisfied with certain choice of the estimators and a simple sampling protocol.

Proof. We now check the assumptions by order.

(1) (p)$\Rightarrow$ Assumption 1. First we have $y^{n,*}(y^{n-1}) = f^{n-1}(y^{n-1})$. In order for the concatenation $f^{n-1}(y^{n-1})$ to be Lipschitz continuous and smooth, we only need each block $f_i^{n-1}(y_i^{n-1})$ to be Lipschitz continuous and smooth. Recall that $f_i^{n-1}(y_i^{n-1}) = y_i^{n-1} + \eta\nabla F_i(y_i^{n-1})$. The Lipschitz continuity of $f_i^{n-1}(y_i^{n-1})$ is guaranteed by the Lipschitz smoothness of $F_i(\cdot)$, which is well established in the literature [64]. Thus we only need to check the Lipschitz smoothness of $f_i^{n-1}(y_i^{n-1})$, that is, the Lipschitz continuity of $\nabla^2 F_i(y_i^{n-1})$. By [14], the policy hessian is given by

$$\nabla^2 F(x) = \mathbb{E}_{\zeta\sim p(\cdot|x)}\Big[\underbrace{g(x;\zeta)\sum_{t=0}^{H}\nabla\log\pi_x(a_t|s_t)^\top + \nabla g(x;\zeta)}_{H(x;\zeta)}\Big], \tag{169}$$

where $\zeta = (s_0, a_0, ..., s_H, a_H)$ and $p(\zeta|x) = \rho(s_0)\pi_x(a_0|s_0)\Pi_{t=0}^{H-1}\mathcal{P}(s_{t+1}|a_t, s_t)\pi_x(a_{t+1}|s_{t+1})$; $g(x;\zeta) := \sum_{h=0}^{H}\nabla \log \pi_x(a_h|s_h)\sum_{t=h}^{H}\gamma^t r(s_t, a_t)$, and we omit i since the result holds for all i.

For any $x, x' \in \mathbb{R}^{d_0}$, we have

$$
\begin{aligned}
&\|\nabla^2 F(x) - \nabla^2 F(x')\| \\
&\leq \big\|\mathbb{E}_{\zeta\sim p(\cdot|x)}[H(x;\zeta)] - \mathbb{E}_{\zeta\sim p(\cdot|x')}[H(x;\zeta)]\big\| + \big\|\mathbb{E}_{\zeta\sim p(\cdot|x')}[H(x;\zeta)] - \mathbb{E}_{\zeta\sim p(\cdot|x')}[H(x';\zeta)]\big\| \\
&\leq \big\|\mathbb{E}_{\zeta\sim p(\cdot|x)}[H(x;\zeta)] - \mathbb{E}_{\zeta\sim p(\cdot|x')}[H(x;\zeta)]\big\| + \mathbb{E}_{\zeta\sim p(\cdot|x')}\big\|H(x;\zeta) - H(x';\zeta)\big\|. \qquad (170)
\end{aligned}
$$

We consider the second term first. By Lemma 14, in order for $H(x;\zeta)$ to be Lipschitz continuous w.r.t. x, it suffices to prove: i) $\nabla \log \pi_x(a|s)$ can be bounded and Lipschitz continuous; ii) $g(x;\zeta)$ can be bounded and Lipschitz continuous; iii) $\nabla g(x;\zeta)$ is Lipschitz continuous. First, i) is directly implied by condition (p). We then prove ii) as follows

$$
\|g(x;\zeta)\| \leq \sum_{h=0}^{H}\big\|\nabla \log \pi_x(a_h|s_h)\big\|\big|\sum_{t=h}^{H}\gamma^t r(s_t, a_t)\big| \leq \frac{C_\pi}{(1-\gamma)^2} \qquad (171)
$$

$$
\begin{aligned}
\|g(x;\zeta) - g(x';\zeta)\| &= \sum_{h=0}^{H}\big\|\nabla \log \pi_x(a_h|s_h) - \nabla \log \pi_{x'}(a_h|s_h)\big\|\sum_{t=h}^{H}|\gamma^t r(s_t, a_t)| \\
&\leq \frac{L'_\pi}{(1-\gamma)^2}\|x - x'\|. \qquad (172)
\end{aligned}
$$

Next we prove iii) as follows

$$
\begin{aligned}
\|\nabla g(x;\zeta) - \nabla g(x';\zeta)\| &\leq \sum_{h=0}^{H}\big\|\nabla^2 \log \pi_x(a_h|s_h) - \nabla^2 \log \pi_{x'}(a_h|s_h)\big\|\sum_{t=h}^{H}|\gamma^t r(s_t, a_t)| \\
&\leq \frac{L''_\pi}{(1-\gamma)^2}\|x - x'\|. \qquad (173)
\end{aligned}
$$

By Lemma 14, we know i), ii) and iii) imply the Lipschitz continuity of $H(x;\zeta)$, i.e. it holds that

$$
\|H(x;\zeta) - H(x';\zeta)\| \leq \frac{L''_\pi + 2HC_\pi L'_\pi}{(1-\gamma)^2}\|x - x'\|. \qquad (174)
$$

The first term in (170) can be bounded as

$$
\begin{aligned}
\big\|\mathbb{E}_{\zeta\sim p(\cdot|x)}[H(x;\zeta)] - \mathbb{E}_{\zeta\sim p(\cdot|x')}[H(x;\zeta)]\big\| &\leq \sup_{\zeta}\big\|H(x;\zeta)\big\|\sum_{\zeta}|p(\zeta|x) - p(\zeta|x')| \\
&\overset{(176)}{\leq} \frac{HC_\pi^2 + L'_\pi}{(1-\gamma)^2}\sum_{\zeta}|p(\zeta|x) - p(\zeta|x')| \\
&\leq \frac{HC_\pi^2 + L'_\pi}{(1-\gamma)^2}(H+1)|\mathcal{A}|L_\pi\|x - x'\| \qquad (175)
\end{aligned}
$$

where the second inequality follows from

$$
\big\|H(x;\zeta)\big\| \leq \|g(x;\zeta)\|\sum_{t=0}^{H}\|\nabla \log \pi_x(a_t|s_t)\| + \|\nabla g(x;\zeta)\| \leq \frac{HC_\pi^2}{(1-\gamma)^2} + \frac{L'_\pi}{(1-\gamma)^2}. \qquad (176)
$$

Substituting (174) and (175) into (170) yields

$$
\|\nabla^2 F(x) - \nabla^2 F(x')\| \leq \Big(\frac{L''_\pi + 2HC_\pi L'_\pi}{(1-\gamma)^2} + \frac{HC_\pi^2 + L'_\pi}{(1-\gamma)^2}(H+1)|\mathcal{A}|L_\pi\Big)\|x - x'\|. \qquad (177)
$$

This implies that $f_i^{n-1}(\cdot)$ is $\eta\big(\frac{L''_\pi + 2HC_\pi L'_\pi}{(1-\gamma)^2} + \frac{HC_\pi^2 + L'_\pi}{(1-\gamma)^2}(H+1)|\mathcal{A}|L_\pi\big)$-Lipschitz smooth for $n \in [N]$.

(2) (q)&(r)⇒ Assumption 3. It is clear that conditions (q)&(r) imply condition (h)&(i). Thus by Lemma 3, Assumption 3 is satisfied.

Now we only need to specify the estimators that satisfy condition (q) as follow. First, it is known that the policy gradient takes the following form [14]:

$$\nabla F_i(x)=\mathbb{E}_{\zeta\sim\pi_x}\Big[\sum_{h=0}^{H}\nabla\log\pi_x(a_h|s_h)\sum_{t=h}^{H}\gamma^t r_i(s_t,a_t)|\rho_i,\mathcal{P}_i\Big]. \tag{178}$$

Then to estimate $f_i^n(y)$ ($n=0,1,...,N-1$), one can use:

$$f_i^n(y;\zeta_i^n)\coloneqq y+\eta\sum_{h=0}^{H}\nabla\log\pi_y(a_h|s_h)\sum_{t=h}^{H}\gamma^t r_i(s_t,a_t),\ n=0,1,...,N-1, \tag{179}$$

where $\zeta_i^n=(s_0,a_0,...,s_H,a_H)$ is generated under policy π_y, transition distribution $\mathcal{P}_i$ and initial distribution ρ_i. The estimator satisfies condition (q):

$$\begin{aligned}&\mathbb{E}_{\zeta_i^n}[f_i^n(y;\zeta_i^n)]=y+\eta\nabla F_i(y)=f_i^n(y),\\ &\mathbb{E}_{\zeta_i^n}[\|f_i^n(y;\zeta_i^n)-f_i^n(x)\|^2]\le\mathbb{E}_{\zeta_i^n}\|\sum_{h=0}^{H}\nabla\log\pi_y(a_h|s_h)\sum_{t=h}^{H}\gamma^t r_i(s_t,a_t)\|^2\le\frac{C_\pi^2}{(1-\gamma)^4}.\end{aligned} \tag{180}$$

To estimate $\nabla f_i^n(y)$ ($n=0,1,...,N-1$), one can use:

$$\nabla f_i^n(y;\zeta_i^n)\coloneqq I+\eta H(y;\zeta_i^n),\ n=0,1,...,N-1, \tag{181}$$

where $\zeta_i^n=(s_0,a_0,...,s_H,a_H)$ is generated under policy π_y, transition distribution $\mathcal{P}_i$ and initial distribution ρ_i. The estimator satisfies condition (q):

$$\begin{aligned}&\mathbb{E}_{\zeta_i^n}[\nabla f_i^n(y;\zeta_i^n)]=I+\eta\nabla^2 F_i(y)=\nabla f_i^n(y),\\ &\mathbb{E}_{\zeta_i^n}\|\nabla f_i^n(y;\zeta_i^n)-\nabla f_i^n(y)\|^2\le\mathbb{E}_{\zeta_i^n}\|\nabla f_i^n(y;\zeta_i^n)\|^2\overset{(176)}{\le}2+2\eta^2\frac{(HC_\pi^2+L_\pi')^2}{(1-\gamma)^4}.\end{aligned} \tag{182}$$

To estimate $\nabla f_i^N(x)$, one can use

$$\nabla f_i^N(x;\zeta_i^N)\coloneqq\sum_{h=0}^{H}\nabla\log\pi_x(a_h|s_h)\sum_{t=h}^{H}\gamma^t r_i(s_t,a_t), \tag{183}$$

where $\zeta_i^n=(s_0,a_0,...,s_H,a_H)$ is generated under policy π_y, transition kernel $\mathcal{P}_i$ and initial distribution ρ_i. This estimator satisfies the condition (q), following the similar lines in (180).

(3) Verifying Assumption 4 and 6. Assumption 4 is satisfied with $\lambda_n=1$ by directly plugging in $y^{n,*}(y^{n-1})=f^{n-1}(y^{n-1})$. Assumption 6 is satisfied by observing that

$$F(x)=\frac{1}{M}\sum_{i=1}^{M}F_i(\tilde{x}_i^N(x))\le\frac{1}{1-\gamma}, \tag{184}$$

where we have used the fact that $F_i(x)\le\frac{1}{1-\gamma}$ for any x. ■

Given the generic result in Theorem 2, Lemma 9 directly implies Theorem 4.

Theorem 6 (Restatement of Theorem 4) *Consider the sequences generated by the MAMPG update in* (32) *and* (33) *for $k=[K]$. Under conditions (p)–(r), we have Theorem 2 holds.*

H Technical Lemmas

Lemma 10 *Suppose Assumption 1 & 2 hold. Recall that $L_y(n)=\sum_{i=n}^{N}L_{y,i-1}L_{y,i-2}\dots L_{y,n}$ with $L_{y,n-1}L_{y,n-2}\dots L_{y,n}=1$ for any $n\in[N]$. Then it holds that*

$$\left\|v(x_k,y_k^{1:N})-v(x_k)\right\|\le L_{v,y}\sum_{n=1}^{N}L_y(n)\|y_k^n-y^{n,*}(y_k^{n-1})\|. \tag{185}$$

Proof. By the Lipschitz continuity of $v(x, y^1, \dots, y^N)$ w.r.t. $y^1, \dots, y^N$, we have

$$\|v(x_k, y_k^{1:N}) - v(x_k)\| \le L_{v,y} \sum_{n=1}^{N} \|y_k^n - y^{n,*}(\dots y^{2,*}(y^{1,*}(x_k))\dots)\|. \tag{186}$$

For any $n \ge 2$, we have

$$\begin{aligned}
&\|y_k^n - y^{n,*}(\dots y^{2,*}(y^{1,*}(x_k))\dots)\| \\
&\le \|y_k^n - y^{n,*}(y_k^{n-1})\| + \|y^{n,*}(y_k^{n-1}) - y^{n,*}(\dots y^{2,*}(y^{1,*}(x_k))\dots)\| \\
&\le \|y_k^n - y^{n,*}(y_k^{n-1})\| + L_{y,n-1}\|y_k^{n-1} - y^{n-1,*}(\dots y^{2,*}(y^{1,*}(x_k))\dots)\|.
\end{aligned} \tag{187}$$

Unraveling yields

$$\|y_k^n - y^{n,*}(\dots y^{2,*}(y^{1,*}(x_k))\dots)\| \le \sum_{j=1}^{n} L_{y,n-1}L_{y,n-2}\dots L_{y,j}\|y_k^j - y^{j,*}(y_k^{j-1})\|, \tag{188}$$

where $L_{y,n-1}L_{y,n-2}\dots L_{y,n} \coloneqq 1$. Substituting (188) into (186) completes the proof. ■

Lemma 11 *With any positive λ_1 and non-negative constants λ_0, $\lambda_2 < \lambda_1$ and $C_1, ..., C_4$, consider the following inequality about the step size $\beta_{k,n-1}$:*

$$(1 + C_1\beta_{k,n-1} + C_2\beta_{k,n-1}^2)(1 - \lambda_1\beta_{k,n}) - 1 + \lambda_2\beta_{k,n} + C_3\alpha_k + C_4\alpha_k^2 \le -\lambda_0\alpha_k. \tag{189}$$

Suppose all step sizes are in the same time-scale. Then given any $\beta_{k,n}$, if $\alpha_k \le \beta_{k,n-1} \le 1$, the above inequality always admits solutions for $\beta_{k,n-1}$.

Proof. First we have

$$C_2\beta_{k,n-1}^2 \le C_2\beta_{k,n-1}, \;\; C_4\alpha_k^2 \le C_4\alpha_k. \tag{190}$$

With the above inequality, we can simplify (189) to

$$(1 + (C_1 + C_2)\beta_{k,n-1})(1 - \lambda_1\beta_{k,n}) + \lambda_2\beta_{k,n} \le 1 - (\lambda_0 + C_3 + C_4)\alpha_k. \tag{191}$$

By $\lambda_2\beta_{k,n} \le (1+(C_1+C_2)\beta_{k,n-1})\lambda_2\beta_{k,n}$, the sufficient condition of (189) is

$$(1 + (C_1 + C_2)\beta_{k,n-1})(1 - \lambda'\beta_{k,n}) \le 1 - (\lambda_0 + C_3 + C_4)\alpha_k. \tag{192}$$

where $\lambda' = \lambda_1 - \lambda_2 > 0$. Next we show that (192) holds. With $\alpha_k \le \beta_{k,n-1}$, rearranging and simplifying (192) gives

$$\beta_{k,n-1} \le \lambda' \frac{\beta_{k,n}}{\lambda_0 + C_1 + C_2 + C_3 + C_4}, \tag{193}$$

which can be satisfied if $\beta_{k,n-1}, \beta_{k,n}$ are in the same scale, and $\beta_{1,n-1}$ is small relative to $\beta_{1,n}$. ■

Lemma 12 (Robbins-Siegmund [18, Theorem 2.3.5]) *Consider a sequence of σ-algebras $\{\mathcal{F}_k\}_{k\ge1}$ and four integrable non-negative sequences $\{U_k\}, \{V_k\}, \{\tau_k\}, \{\delta_k\}$ that satisfy*

i) $U_k, V_k, \tau_k, \delta_k$ are $\mathcal{F}_k$-measurable.

ii) $\Pi_{k\ge1}(1 + \tau_k) < \infty$ and $\sum_{k\ge1} \mathbb{E}[\beta_k] < \infty$.

iii) For $k \ge 1$, $\mathbb{E}[V_{k+1}|\mathcal{F}_k] \le V_k(1 + \tau_k) + \delta_k - U_{k+1}$.

Then it holds that

1) $V_k \xrightarrow{k\to\infty} V_\infty < \infty$ and $\sup_{k\ge1} \mathbb{E}[V_k] < \infty$.

2) $\sum_{k\ge1} \mathbb{E}[U_k] < \infty$ and $\sum_{k\ge1} U_k < \infty$ a.s.

Lemma 13 *Suppose Assumption 1 holds. Then there exists a positive constant C_N such that*

$$\|x_k - x^*\|^2 + \sum_{n=1}^{N} \|y_k^n - y^{n,*}\|^2 \le C_N\big(\|x_k - x^*\|^2 + \sum_{n=1}^{N} \|y_k^n - y^{n,*}(y_k^{n-1})\|^2\big). \tag{194}$$

Proof. First note that under Assumption 1, we have

$$\sum_{n=1}^{N} \|y_k^n - y^{n,*}\| = \sum_{n=1}^{N} \|y_k^n - y^{n,*}(\dots y^{2,*}(y^{1,*}(x^*))\|. \tag{195}$$

To bound the RHS of the above inequality, we can directly follow the derivation of (186)–(188) with $x_k = x^*$ and obtain

$$\begin{aligned}\sum_{n=1}^{N} \|y_k^n - y^{n,*}\| &= \sum_{n=1}^{N} \|y_k^n - y^{n,*}(\dots y^{2,*}(y^{1,*}(x^*))\| \\ &\leq L_y(1)\|y_k^1 - y^{1,*}(x^*)\| + \sum_{n=2}^{N} L_y(n)\|y_k^n - y^{n,*}(y_k^{n-1})\|, \end{aligned}\tag{196}$$

where $\{L_y(n)\}_{n=1}^N$ is a series of constants specified in Lemma 10.

Continuing from the last inequality, we have

$$\begin{aligned}\sum_{n=1}^{N} \|y_k^n - y^{n,*}\| &\leq L_y(1)\|y^{1,*}(x_k) - y^{1,*}(x^*)\| + L_y(1)\|y_k^1 - y^{1,*}(x_k)\| + \sum_{n=2}^{N} L_y(n)\|y_k^n - y^{n,*}(y_k^{n-1})\| \\ &\leq L_y(1)L_{y,1}\|x_k - x^*\| + \sum_{n=1}^{N} L_y(n)\|y_k^n - y^{n,*}(y_k^{n-1})\|. \end{aligned}\tag{197}$$

Then we have

$$\begin{aligned}&\|x_k - x^*\|^2 + \sum_{n=1}^{N} \|y_k^n - y^{n,*}\|^2 \\ &\leq \Big(\|x_k - x^*\| + \sum_{n=1}^{N} \|y_k^n - y^{n,*}\|\Big)^2 \\ &\overset{(197)}{\leq} 2(1 + L_y(1)L_{y,1})^2\|x_k - x^*\|^2 + 2N\sum_{n=1}^{N} L_y^2(n)\|y_k^n - y^{n,*}(y_k^{n-1})\|^2. \end{aligned}\tag{198}$$

With the above inequality, choosing $C_N = 2\max\{(1 + L_y(1)L_{y,1})^2, NL_y^2(1), ..., NL_y^2(n)\}$ completes the proof. ■

Lemma 14 (Lipschitz continuity of a product.) *Define $f_i : \mathbb{R}^d \mapsto \mathbb{R}^{d_i \times d_{i+1}}$. If there exist positive constants $L_1, L_2, ..., L_n$ and $C_1, C_2, ..., C_n$ such that for any $x, x' \in \mathbb{R}^d$ it holds that*

$$\|f_i(x) - f_i(x')\| \leq L_i\|x - x'\|,\ \|f_i(x)\| \leq C_i,\ \forall i \in [n]. \tag{199}$$

Then it holds that

$$\|f_1(x)f_2(x)...f_n(x) - f_1(x')f_2(x')...f_n(x')\| \leq \sum_{j=1}^{n} C_1C_2...L_j...C_n\|x - x'\|. \tag{200}$$

Proof. We can decompose the product as

$$\begin{aligned}&\|f_1(x)f_2(x)...f_n(x) - f_1(x')f_2(x')...f_n(x')\| \\ &= \|f_1(x)f_2(x)...f_n(x) - f_1(x')f_2(x)...f_n(x) + f_1(x')f_2(x)...f_n(x) - f_1(x')f_2(x')...f_n(x) \\ &\quad + \cdots + f_1(x')f_2(x')...f_n(x) - f_1(x')f_2(x')...f_n(x')\| \\ &\leq C_2...C_n\|f_1(x) - f_1(x')\| + C_1C_3..C_n\|f_2(x) - f_2(x')\| + \cdots + C_1C_2..C_{n-1}\|f_n(x) - f_n(x')\| \\ &\leq \sum_{j=1}^{n} C_1C_2...L_j...C_n\|x - x'\|. \end{aligned}\tag{201}$$

This completes the proof. ■