# Prelimit Coupling and Steady-State Convergence of Constant-stepsize Nonsmooth Contractive SA

Yixuan Zhang,[‡] Dongyan (Lucy) Huo,[§] Yudong Chen,[†] Qiaomin Xie,[‡*]

[‡]Department of Industrial and Systems Engineering, University of Wisconsin-Madison
[§]School of Operations Research and Information Engineering, Cornell University
[†]Department of Computer Sciences, University of Wisconsin-Madison

**Abstract**

Motivated by Q-learning, we study nonsmooth contractive stochastic approximation (SA) with constant stepsize. We focus on two important classes of dynamics: 1) nonsmooth contractive SA with additive noise, and 2) synchronous and asynchronous Q-learning, which features both additive and multiplicative noise. For both dynamics, we establish weak convergence of the iterates to a stationary limit distribution in Wasserstein distance. Furthermore, we propose a prelimit coupling technique for establishing steady-state convergence and characterize the limit of the stationary distribution as the stepsize goes to zero. Using this result, we derive that the asymptotic bias of nonsmooth SA is proportional to the square root of the stepsize, which stands in sharp contrast to smooth SA. This bias characterization allows for the use of Richardson-Romberg extrapolation for bias reduction in nonsmooth SA.

## 1 Introduction

Stochastic Approximation (SA) is a fundamental algorithmic paradigm for solving fixed-point problems iteratively based on noisy observations. SA procedures have been widely used in many application domains, including reinforcement learning (RL), stochastic control and optimization [Ber19, SB18, KY03, MB11]. A typical SA algorithm is of the form

$$\theta_{t+1}^{(\alpha)} = \theta_t^{(\alpha)} + \alpha\big(\widetilde{\mathcal{H}}(\theta_t^{(\alpha)}, w_t) - \theta_t^{(\alpha)}\big), \tag{1}$$

where $\{w_t\}_{t\geq 0}$ represent the noise sequence and $\alpha > 0$ is a constant stepsize. The SA procedure (1) aims to approximately find the solution $\theta^*$ to the fixed-point equation $\mathcal{H}(\theta^*) = \theta^*$, where $\mathcal{H}(\cdot) := \mathbb{E}_w[\widetilde{\mathcal{H}}(\cdot, w)]$ is the expectation of the operator $\widetilde{\mathcal{H}}(\cdot, w)$ with respect to the noise. Equation (1) covers many popular algorithms, such as the prevalent stochastic gradient descent (SGD) algorithm for minimizing an objective function [Lan20], and variants of TD-learning algorithms for policy evaluation in RL [SB18].

In this work, we focus on *nonsmooth* contractive SA, where the operator $\widetilde{\mathcal{H}}(\cdot, w)$ may be nondifferentiable (in its first argument) and $\mathcal{H}(\cdot)$ is a contractive mapping with respect to a norm $\|\cdot\|_c$. One prominent example of nonsmooth contractive SA is the celebrated Q-learning algorithm for optimal control in RL [WD92], where $\widetilde{\mathcal{H}}$ corresponds to the noisy optimal Bellman operator involving a max function. Other common nonsmooth mappings include the largest eigenvalue of a matrix, $\ell_1$-norm regularized functions, and their composition with smooth functions [Sag13, Sha03]. It is of fundamental interest to gain a complete understanding of the evolution and long-run behavior of the iterates $\{\theta_t^{(\alpha)}\}_{t\geq 0}$ generated by nonsmooth contractive SA.

Under suitable conditions on the operator $\widetilde{\mathcal{H}}$ and the noise sequence $\{w_t\}_{t\geq 0}$, the SA iterates $\{\theta_t^{(\alpha)}\}_{t\geq 0}$ form a time-homogeneous Markov chain and quickly converge to some limit random variable $\theta^{(\alpha)}$ [DDB20, YBVE21]. Recent work has developed a suite of results for *smooth* SA [DDB20, HCX23b, DJMS21], including the

*Accepted for presentation at the ACM SIGMETRICS 2024. Emails: yzhang2554@wisc.edu, dh622@cornell.edu, yudong.chen@wisc.edu, qiaomin.xie@wisc.edu

geometric convergence of the chain, finite-time bounds on the higher moments, as well as properties of the limit $\theta^{(\alpha)}$. It has been observed that often $\mathbb{E}[\theta^{(\alpha)}] \neq \theta^*$, due to the use of constant stepsize. The difference $\mathbb{E}[\theta^{(\alpha)}] - \theta^*$ is referred to as the asymptotic bias. In particular, for SA with *differentiable* dynamic, the work [DDB20, HCX23b] makes use of Taylor expansion of $\widetilde{\mathcal{H}}$ to establish that the asymptotic bias is proportional to the stepsize $\alpha$ (up to a higher order term), i.e.,

$$\mathbb{E}[\theta^{(\alpha)}] - \theta^* = c\alpha + o(\alpha), \tag{2}$$

where $c$ is some vector independent of $\alpha$ and $o(\alpha)$ denotes a term that decays faster than $\alpha$. Such a fine-grained characterization of SA iterates gives rise to variance and bias reduction techniques that lead to improved estimation of the target solution $\theta^*$, as well as efficient statistical inference procedures [DDB20, HCX23b, HCX23a].

For nonsmooth SA, far little is known. Existing analysis based on the linearization / Taylor expansion of $\widetilde{\mathcal{H}}$ is no longer applicable. Hence, distributional convergence and bias characterization results like (2) have not been established for nonsmooth SA procedures like Q learning. In fact, it is not even clear whether equation (2) remains valid for nonsmooth SA, and if not, what is the correct characterization.

**Our Contributions:** To investigate the above questions, we consider two important classes of nonsmooth contractive SA algorithms:

1. Nonsmooth SA with additive noise, where $\widetilde{\mathcal{H}}(\theta, w) \equiv \mathcal{T}(\theta) + w$. Our results cover operators $\mathcal{T}$ that are $g \circ F$ decomposable, which is a rich class of smooth and nonsmooth functions [Sha03]. See Section 2 for the formal description of the model.

2. A general form of Q-learning dynamics, which are nonsmooth SA with both additive and multiplicative noise. The model covers both synchronous Q-learning and asynchronous Q-learning as special cases. See Section 3 for the formal description of the model.

The first main result of the paper establishes the weak convergence of the Markov chain $\{\theta_t^{(\alpha)}\}_{t\geq 0}$ to a unique stationary distribution in $W_2$ — the Wasserstein distance of order 2 with respect to the contraction norm $\|\cdot\|_c$ — for both the additive noise setting and Q-learning. Moreover, we characterize the geometric convergence rate. As a by-product of our analysis, we derive finite-time upper bounds on $\mathbb{E}\|\theta_t^{(\alpha)} - \theta^*\|_c^{2n}$, the $2n$-th moments of the estimation errors, generalizing the mean-square error (MSE) bound (i.e. $\mathbb{E}\|\theta_t^{(\alpha)} - \theta^*\|_c^2 \leq \dots$) in [CMSS20, CMSS23] to higher moments and the smooth SA results in [DDB20, SY19] to nonsmooth SA.

We next turn to the characterization of the stationary distribution of $\{\theta_t^{(\alpha)}\}_{t\geq 0}$. As existing techniques, which are based on linearizing $\widetilde{\mathcal{H}}(\theta, w)$ as $\theta \to \theta^*$, are not applicable for nonsmooth SA, we take an alternative approach by studying the limiting behavior of the properly rescaled iterates as the constant stepsize $\alpha$ approaches 0. Since the MSE of $\theta_t^{(\alpha)}$ is of order $\mathcal{O}(\alpha)$ [CMSS20], the proper choice of scaling is by $\sqrt{\alpha}$, the diffusion scaling. In particular, we consider the centered and $\sqrt{\alpha}$-scaled iterates $Y_t^{(\alpha)} := \frac{\theta_t^{(\alpha)} - \theta^*}{\sqrt{\alpha}}$ so that the MSE of $Y_t^{(\alpha)}$ is $\mathcal{O}(1)$. The weak convergence of $\theta_t^{(\alpha)}$ to a limit $\theta^{(\alpha)}$ implies that $Y_t^{(\alpha)}$ converges weakly to the limit $Y^{(\alpha)} := \frac{\theta^{(\alpha)} - \theta^*}{\sqrt{\alpha}}$ as $t \to \infty$. Therefore, to understand the stationary distribution $\theta^{(\alpha)}$ and its scaled version $Y^{(\alpha)}$, we are interested in characterizing *steady-state convergence*, i.e., the convergence of $Y^{(\alpha)}$ as $\alpha \to 0$ and the limit $Y$ (if exists). This limit is illustrated by the red solid path in Fig. 1.

As we argue in Section 1.1, existing approaches to steady-state convergence face severe challenges in the nonsmooth SA setting. In this work, we develop a new *prelimit coupling technique*, which allows us to establish the weak convergence of $Y^{(\alpha)}$ in $W_2$ to a unique limiting random variable $Y$ as $\alpha \to 0$. Importantly, our technique can handle both additive noise and multiplicative noise, and provide an explicit rate of convergence. An overview of our technique is provided in Section 1.2. We remark that our technique can be potentially applied to the study of steady-state convergence in other stochastic dynamical systems and hence may be of its own interest.

Since convergence in $W_2$ implies convergence of the first two moments of $Y^{(\alpha)}$, we obtain the following characterization of the asymptotic bias of the SA iterates:

$$\mathbb{E}[\theta^{(\alpha)}] - \theta^* = \mathbb{E}[Y] \cdot \sqrt{\alpha} + o(\sqrt{\alpha}). \tag{3}$$

$$
\begin{tikzcd}
Y_t^{(\alpha)} \arrow[r, "t \to \infty"] \arrow[d, dashed, "\alpha t = \bar{t}"', "\alpha \downarrow 0"] & Y^{(\alpha)} \arrow[d, "\alpha \downarrow 0"] \\
\overline{Y}_{\bar{t}} \arrow[r, dashed, "\bar{t} \to \infty"'] & Y
\end{tikzcd}
$$

Figure 1: Steady-state convergence.

We further provide a fine-grained characterization of the expectation $\mathbb{E}[Y]$, which appears above in (3), and relate it to the structure of the SA update (1). Our results show that $\mathbb{E}[Y] \neq 0$ precisely when the operator $\widetilde{\mathcal{H}}$ is truly nonsmooth, in which case the asymptotic bias is of order $\sqrt{\alpha}$. This result stands in sharp contrast to the $\alpha$-order bias of smooth SA in equation (2).

Finally, we explore the implications of the above results for iterate averaging and extrapolation. In particular, we consider applying Polyak-Ruppert (PR) tail averaging [Rup88, PJ92, JKK+18] and Richardson-Romberg (RR) extrapolation [Hil87] to the iterates generated by contractive SA algorithms. We investigate the resulting estimation errors and biases in the presence of nonsmoothness. In particular, thanks to the bias characterization in (3), we can employ the RR extrapolation technique to eliminate the leading term $\mathbb{E}[Y] \cdot \sqrt{\alpha}$ and reduce the asymptotic bias to a higher order of $\sqrt{\alpha}$.

## 1.1 Challenges of Applying Existing Techniques to Nonsmooth SA

Steady-state convergence, i.e., showing $Y^{(\alpha)} \to Y$ in Figure 1, is a problem of fundamental interest in stochastic dynamical systems, such as queueing networks [GZ06]. One well-known approach to proving steady-state convergence in queueing networks is via justifying the *interchange of limits*, i.e., equivalence of the solid and dashed paths in Figure 1 [GZ06, Gur14, YY16, YY18]. Doing so is well recognized to be technically challenging, often requiring sophisticated "hydro-dynamic limits" methodology [Bra98] as well as a well-defined stochastic differential equation (SDE) $\bar{Y}_{\bar{t}}$ with a stationary distribution. In our setting, it is unclear whether nonsmooth SA is associated with such an SDE, let alone the validity of interchanging the limits.

An alternative approach to the steady-state convergence is based on the *Basic Adjoint Relationship* (BAR) for the generator of the stochastic process. By using BAR with an exponential test function, one may be able to prove convergence of the moment generating function and in turn weak convergence of the corresponding random variables [BDM17, BDM24, CMM22]. In our setting, however, the BAR of moment generating functions does not always lead to a straightforward solution. In fact, even for *smooth* SA dynamics with only additive noise (i.e., $\widetilde{\mathcal{H}}(\theta,w) = \mathcal{H}(\theta) + w$), steady-state convergence is proved in the work [CMM22] only when the limit random variable $Y$ is Gaussian and under the assumption that the following equation from BAR has a unique solution in $Y$:

$$
\mathbb{E}\big[\big(\varphi^\top \operatorname{Var}(w)\varphi - 2i\varphi^\top \nabla\mathcal{H}(\theta^*)Y\big)e^{i\varphi^\top Y}\big] = 0, \quad \forall \varphi \in \mathbb{R}^d. \tag{4}
$$

Verifying this assumption is challenging in general; in [CMM22] this is done only when $d = 1$ or under some restricted conditions when $d \geq 2$. This difficulty is only exacerbated in the broader nonsmooth contractive SA setting, which covers the smooth SA setting considered in [CMM22].

The $\sqrt{\alpha}$-scaling in our problem suggests a potential connection to the Langevin diffusion SDE and the literature on the Unadjusted Langevin Algorithm (ULA) [DM17, DM19]. ULA corresponds to the Euler-Maruyama discretization of the Langevin diffusion and is given by

$$
Y_{t+1}^{(\alpha)} = Y_t^{(\alpha)} - \alpha\nabla U(Y_t^{(\alpha)}) + \sqrt{\alpha}w_t,
$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is a potential function and $\{w_t\}_{t\geq 0}$ are i.i.d. Gaussian noise. However, by comparing ULA with the SA update (1) scaled by $\sqrt{\alpha}$, one sees that the latter reduces to ULA only when the noise is additive and Gaussian and $\mathcal{H}$ is a gradient field and positive homogeneous at $\theta^*$.

We complement the above discussion with a simple example of nonsmooth contractive SA:

$$\theta_{t+1}^{(\alpha)} = \theta_t^{(\alpha)} + \alpha\Big( -\frac{1}{2}|\theta_t^{(\alpha)}| - b - \theta_t^{(\alpha)} + w_t\Big), \tag{5}$$

where $w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $b \in \mathbb{R}$. The above dynamic is a special case of (1) with $\widetilde{\mathcal{H}}(\theta, w) = -|\theta|/2 - b + w$, which is nondifferentiable at $\theta = 0$. Despite its apparent simplicity, this example already demonstrates some of the complexity in understanding the steady-state behavior of nonsmooth SA. For example, it is unclear how to follow the BAR approach to obtain a functional equation like (4) for the limit $Y$. The derivation of (4) in [CMM22] relies on the continuous differentiability of the contraction operator $\widetilde{\mathcal{H}}(\theta, w)$.[1] Also, incidentally, when $b = 0$, the dynamic (5) becomes an ULA update. The results in [DM17, DM19] on ULA suggest that the limit $Y$ is not Gaussian as its density function $e^{U(x)}$ involves a non-quadratic $U$. This contrasts to smooth SA, for which the BAR approach shows that $Y$ is Gaussian [CMM22]. As we soon see, the techniques in this paper bypass the need of working directly with and imposing assumptions on equations like (4).

## 1.2 Prelimit Coupling Technique

To overcome the challenges discussed in the previous subsection, we develop a prelimit coupling technique that can be used to establish the desired steady-state convergence without restrictive assumptions. We establish this result by proving convergence in Wasserstein distance $W_2$, i.e.,

$$\lim_{\alpha \to 0} W_2(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)) = 0$$

for a random variable $Y$, where $\mathcal{L}(Y)$ denotes the distribution of $Y$. Our approach applies coupling arguments to the prelimit random variables $Y_t^{(\alpha)}$ with $\alpha > 0, t < \infty$ and consists of three steps.

**Step 1: Gaussian Noise and Rational Stepsize** First, we assume that the noise sequence $\{w_t\}_{t \geq 0}$ is i.i.d. Gaussian. Consider two stepsizes $\alpha$ and $\alpha' = \alpha/k$, where $k \in \mathbb{N}^+$. We have the corresponding scaled iterates $Y_t^{(\alpha)}$ and $Y_t^{(\alpha')}$ generated by equation (1). The main idea is to couple these two sequences in such a way that one step of $Y_t^{(\alpha)}$ corresponds to $k$ steps of $Y_t^{(\alpha')}$:

$$\begin{aligned} Y_{t+1}^{(\alpha)} &= (1-\alpha)Y_t^{(\alpha)} + \sqrt{\alpha}\Big[\widetilde{\mathcal{H}}\Big(\sqrt{\alpha}Y_t^{(\alpha)} + \theta^*, \frac{w_{kt} + \cdots + w_{kt+k-1}}{\sqrt{k}}\Big) - \theta^*\Big], \\ Y_{t+1}^{(\alpha')} &= (1-\alpha')Y_t^{(\alpha')} + \sqrt{\alpha'}\Big[\widetilde{\mathcal{H}}\big(\sqrt{\alpha'}Y_t^{(\alpha')} + \theta^*, w_t\big) - \theta^*\Big]. \end{aligned}$$

Note that $(w_{kt} + \cdots + w_{kt+k-1})/\sqrt{k}$ and $w_t$ are identically distributed under the Gaussian noise assumption. Under this coupling, we establish convergence of the squared distance $\mathbb{E}\|Y_t^{(\alpha)} - Y_{kt}^{(\alpha')}\|_c^2$ under some appropriate norm $\|\cdot\|_c$. Sending $t$ to infinity gives $W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha')})\big) \in o(1)$. Generalizing this argument to rational stepsizes $\alpha$ and $\alpha'$, we conclude that $(Y^{(\alpha)})_{\alpha \to 0, \alpha \in \mathbb{Q}^+}$ is a Cauchy sequence with respect to $W_2$. Consequently, there exists a limit $Y$ such that

$$\lim_{\alpha \to 0, \alpha \in \mathbb{Q}^+} W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\right) = 0.$$

**Step 2: General Stepsize** Still assuming Gaussian noise, we prove that $\mathcal{L}(Y^{(\alpha)})$ is continuous in $\alpha$ with respect to $W_2$. To this end, we consider two real-valued stepsizes $\alpha$ and $\alpha'$ and couple the sequences $Y_t^{(\alpha)}$ and $Y_t^{(\alpha')}$, this time by letting them share the same noise:

$$\begin{aligned} Y_{t+1}^{(\alpha)} &= (1-\alpha)Y_t^{(\alpha)} + \sqrt{\alpha}\left(\widetilde{\mathcal{H}}\big(\sqrt{\alpha}Y_t^{(\alpha)} + \theta^*, w_t\big) - \theta^*\right), \\ Y_{t+1}^{(\alpha')} &= (1-\alpha')Y_t^{(\alpha')} + \sqrt{\alpha'}\left(\widetilde{\mathcal{H}}\big(\sqrt{\alpha'}Y_t^{(\alpha')} + \theta^*, w_t\big) - \theta^*\right). \end{aligned}$$

[1] In particular, for deriving the limit of the $T_2$ term in [CMM22, Page 15].

We again control the squared distance $\mathbb{E}\|Y_t^{(\alpha)} - Y_t^{(\alpha')}\|_c^2$, and then set $t \to \infty$ followed by $\alpha' \to \alpha$, thus establishing the continuity property $\lim_{\alpha' \to \alpha} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha')})\big) = 0$. Since $\mathbb{Q}^+$ is dense in $\mathbb{R}^+$, together with the result from step 1, we obtain

$$\lim_{\alpha \to 0} W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\right) = 0.$$

**Step 3: General Noise** In this step, we relax the Gaussian noise assumption. Suppose the sequence $Y_t'^{(\alpha)}$ is driven by some general noise $w_t'$, and let $Y_t^{(\alpha)}$ be driven by Gaussian noise $w_t$ with matching first two moments. Setting $\kappa = \lfloor \alpha^{-1/2} \rfloor$, we use a multivariate Berry-Esseen bound in Wasserstein distance [Bon20] to show that there exists a coupling between $w_t'$ and $w_t$ such that

$$\mathbb{E}\Big\| \frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t - \frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t' \Big\|_2^2 = W_2^2\Big(\mathcal{L}\big(\frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t\big), \mathcal{L}\big(\frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t'\big)\Big) \in \mathcal{O}\Big(\frac{1}{\kappa}\Big),$$

Under this noise coupling, we bound $\mathbb{E}\|Y_{\kappa t}'^{(\alpha)} - Y_{\kappa t}^{(\alpha)}\|_c^2$, which in turn bounds $W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}((Y')^{(\alpha)})\right)$, thereby establishing that $Y'^{(\alpha)}$ and $Y^{(\alpha)}$ have the same distributional limit as $\alpha \to 0$.

Following the above three-step procedure, the majority of the technical work goes into obtaining tight estimates for squared distances of the form $\mathbb{E}\|Y_t^{(\alpha)} - Y_{t'}^{(\alpha')}\|_c^2$, with potentially mismatched stepsizes $(\alpha, \alpha')$ and time indices $(t, t')$. Doing so under the nonsmooth SA dynamics requires carefully analyzing the multi-step dynamics and leveraging the contractive property via a generalized Moreau envelope argument.

## 1.3 Notations

We use $B^d(\theta, \epsilon)$ to denote an open ball centering at $\theta$ with radius $\epsilon > 0$ with respect to $\ell_2$-norm. A function is $C^k$ if it is $k$ times continuously differentiable. An operator $\mathcal{T} : \mathbb{R}^d \to \mathbb{R}^d$ is said to be $\gamma$-contractive w.r.t. the norm $\|\cdot\|_c$ if for some $\gamma \in (0, 1)$,

$$\|\mathcal{T}(\theta) - \mathcal{T}(\theta')\|_c \leq \gamma \|\theta - \theta'\|_c, \quad \forall \theta, \theta' \in \mathbb{R}^d. \tag{6}$$

A function $h$ is called $L$-smooth w.r.t. some norm $\|\cdot\|$ if $\|\nabla h(x) - \nabla h(y)\|_* \leq L\|x - y\|, \forall x, y$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Let $\mathcal{P}_2(\mathbb{R}^d)$ denote the space of square-integrable distributions on $\mathbb{R}^d$. For a random vector $\theta$, let $\mathcal{L}(\theta)$ denote the distribution of $\theta$ and $\mathrm{Var}(\theta)$ its covariance matrix. The Wasserstein 2-distance between two distributions $\mu$ and $\nu$ in $\mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$W_2(\mu, \nu) = \inf_{\xi \in \Pi(\mu, \nu)} \left( \int_{\mathbb{R}^d} \|u - v\|_c^2 \, \mathrm{d}\xi(u, v) \right)^{\frac{1}{2}} = \inf \left\{ \left( \mathbb{E}\left[ \|\theta - \theta'\|_c^2 \right] \right)^{\frac{1}{2}} : \mathcal{L}(\theta) = \mu, \mathcal{L}(\theta') = \nu \right\},$$

where $\Pi(\mu, \nu)$ is the set of joint distributions in $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ with marginal distributions $\mu$ and $\nu$.

For a finite set $\mathcal{S}$, we use $\Delta(\mathcal{S})$ to denote the probability simplex over $\mathcal{S}$. Given $\pi \in \Delta(\mathcal{S})$, we denote by $\mathrm{Multi}(\pi, n)$ the multinomial distribution with event probabilities $\pi$ and number of trials $n$.

For two real valued functions $f(x), g(x) : \mathbb{R}^+ \to \mathbb{R}$, we write $f(x) \in o(g(x))$ if $\lim_{x \to 0} \frac{f(x)}{g(x)} = 0$, and we write $f(x) \in \mathcal{O}(g(x))$ if there exist $x_0, M > 0$ such that $|f(x)| \leq M g(x), \forall x \leq x_0$. We say that $f(x)$ is superpolynomial if $f(x) \in o(x^n), \forall n \in \mathbb{N}^+$.

**Paper organization.**

The remainder of the paper is organized as follows. In Section 2 we present the model and the main results for SA with additive noise. In Section 3 we extend these results to Q-learning. In Section 4 we explore the implications of our results for Polyak-Ruppert averaging and Richardson-Romberg extrapolation. We outline the proofs of our main results in Section 5. We provide numerical experiments that corroborate our theoretical results in Section 6. We discuss additional related work in Section 7.

## 2 Nonsmooth Stochastic Approximation with Additive Noise

In this section, we consider contractive nonsmooth stochastic approximation with additive noise.

### 2.1 Model Setup

We consider the following stochastic approximation iteration with additive noise:

$$\theta_{t+1}^{(\alpha)} = \theta_t^{(\alpha)} + \alpha\big(\mathcal{T}(\theta_t^{(\alpha)}) - \theta_t^{(\alpha)} + w_t\big), \tag{7}$$

where $\mathcal{T} : \mathbb{R}^d \to \mathbb{R}^d$ is an operator, $\alpha > 0$ is a constant stepsize and $\{w_t\}_{t\geq 0}$ is a sequence of i.i.d zero-mean noise.

Stochastic approximation subsumes many important iterative algorithms. For example, if $\mathcal{T}(\theta) = -\nabla U(\theta) + \theta$ for some function $U : \mathbb{R}^d \to \mathbb{R}$ that is twice continuously differentiable, $L$-smooth and $\sigma$-strongly convex, then the update (7) corresponds to Stochastic Gradient Descent (SGD) for minimizing $U$ [Lan20]. If $\mathcal{T}(\theta) = A\theta + b$, where $A \in \mathbb{R}^{d\times d}$ is a Hurwitz matrix, then (7) becomes Linear SA, which in turn covers the TD-learning algorithm in reinforcement learning [HCX23b, SY19]. In both examples, the operator $\mathcal{T}$ is at least $C^1$-smooth and contractive in $\|\cdot\|_2$ (or its weighted version).

In this work, we consider a more general class of SA algorithms with a potentially nonsmooth operator $\mathcal{T}$. We only assume that $\mathcal{T}$ is contractive with respect to an arbitrary norm.

**Assumption 1** (Contractive $\mathcal{T}$)**.** *The operator $\mathcal{T} : \mathbb{R}^d \to \mathbb{R}^d$ is $\gamma$-contractive for some $\gamma \in (0, 1)$ with respect to some norm $\|\cdot\|_c$.*

By Banach fixed point theorem, the fixed point equation $\mathcal{T}(\theta) = \theta$ has a unique solution $\theta^* \in \mathbb{R}^d$.

We consider the following moment assumption for the additive noise $w_t$, indexed by $n \geq 1$.

**Assumption 2** ($\boldsymbol{n}$)**.** *The random variables $\{w_t\}_{t\geq 0}$ have finite $(2n)$-th moments.*

Such moment assumptions, for example with $n = 1$ or 2, are standard in prior work on the analysis of SGD and SA[DDB20, KY03, SY19]. In general, under Assumption 2($\boldsymbol{n}$) we can control the $2n$-th moment of the SA iterates $\{\theta_t^{(\alpha)}\}_{t\geq 0}$.

### 2.2 Moments Bounds and Convergence to Stationary Distribution

We first derive finite-time upper bounds on $\mathbb{E}\|\theta_t^{(\alpha)} - \theta^*\|_c^{2n}$, the $2n$-th moments of the estimation errors, generalizing the results in [CMSS20, CMSS23] to higher moments and those in [DDB20] to nonsmooth SA.

**Proposition 1** (Moment Bounds)**.** *For each integer $n \geq 1$, under Assumption 1 and Assumption 2($\boldsymbol{n}$), there exists $\bar{\alpha} > 0$ such that for any $\alpha \leq \bar{\alpha}$, there exists $t_{\alpha,n} > 0$ and*

$$\mathbb{E}[\|\theta_t^{(\alpha)} - \theta^*\|_c^{2n}] \leq c_n \mathbb{E}[\|\theta_{t_{\alpha,n}}^{(\alpha)} - \theta^*\|_c^{2n}](1 - \alpha(1 - \sqrt{\gamma}))^{t - t_{\alpha,n}} + c_n' \alpha^n, \quad \forall t \geq t_{\alpha,n}, \tag{8}$$

*where $c_n$ and $c_n'$ are constants that are independent with $\alpha$ and $t$. Moreover, $t_{\alpha,1} = 0$.*

In subsequent analysis, we mostly use Assumption 2($\boldsymbol{n}$) and Proposition 1 with $n \in \{1, 2\}$. In particular, Proposition 1 with $n = 1$ provides a finite-time mean-square error (MSE) bound. Using this bound, we can establish our first main theorem, which proves the weak convergence of the stochastic process $\{\theta_t^{(\alpha)}\}_{t\geq 0}$ to a unique stationary distribution in $W_2$; moreover, we characterize its geometric convergence rate.

**Theorem 1** (Distributional Convergence)**.** *Under Assumption 1 and Assumption 2(1), there exists $\bar{\alpha}' > 0$ such that for any stepsize $\alpha \leq \bar{\alpha}'$ and any initial distribution of $\theta_0^{(\alpha)}$, the sequence $\{\theta_t^{(\alpha)}\}_{t\geq 0}$ converges geometrically in $W_2$ to a random variable $\theta^{(\alpha)}$ with*

$$W_2^2(\mathcal{L}(\theta_t^{(\alpha)}), \mathcal{L}(\theta^{(\alpha)})) \leq c \cdot (1 - \alpha(1 - \sqrt{\gamma}))^t, \quad \forall t \geq 0,$$

*where $c$ is a constant that is independent of $\alpha$ and $t$. Moreover, $\mathbb{E}[\|\theta^{(\alpha)} - \theta^*\|_2^2] \in \mathcal{O}(\alpha)$.*

A key step in proving Proposition 1 and Theorem 1 is to construct a proper *smooth* Lyapunov function for the nonsmooth dynamics. Previous works on higher moments bounds and convergence in $W_2$ focus on linear SA and smooth SGD [DDB20, HCX23b]. These dynamics are smooth and contractive in the $\ell_2$ norm $\|\cdot\|_c = \|\cdot\|_2$, the square of which can be used as a smooth Lyapunov function. However, for general contractive SA, the norm $\|\cdot\|_c$ may be nondifferentiable, e.g., $\|\cdot\|_\infty$. To handle this general setting, we make use of the generalized Moreau envelope of $\|\cdot\|_c$, a technique that has been used in [CMSS20, CMSS23] to study the MSE (i.e., $n = 1$) of contractive SA. To further establish the weak convergence result in Theorem 1, we develop a careful coupling argument using the Moreau envelope, going beyond the $\ell_2$ norm based anaysis in [HCX23b, DDB20]. The proofs of Proposition 1 and Theorem 1 are outlined in Section 5 and given in full in Appendices A and B.

## 2.3 Steady-State Convergence and Bias Characterization

Sometimes we restrict to a more specific but still quite general class of SA dynamics. In particular, we consider operators $\mathcal{T}$ that are defined by the so-called $g \circ F$ decomposable functions, a class of nonsmooth functions first introduced in the work [Sha03]. We extend the definition in [Sha03] to multi-dimensional functions.

**Definition 1.** *We say that the function $f : \mathbb{R}^d \to \mathbb{R}^d$ is $g \circ F$ decomposable at $\bar{\theta}$ if it admits the following local representation*

$$f(\theta) = f(\bar{\theta}) + g(F(\theta - \bar{\theta})), \quad \forall \theta \in B^d(\bar{\theta}, \epsilon)$$

*for some mappings $g : \mathbb{R}^m \to \mathbb{R}^d$ and $F : B^d(0, \epsilon) \to \mathbb{R}^m$ that satisfy: (i) $g$ is positively homogeneous[2] of degree 1 and continuous; (ii) $F$ is differentiable at $B^d(0, \epsilon)$, $\nabla F$ is continuous at 0 and $F(0) = 0$.*

The $g \circ F$ decomposable function class is a rich class that contains max-functions, largest eigenvalue functions, and $\ell_1$-norm regularized functions, as well as their composition with smooth functions. See [Sha03, Sag13] for other special cases of $g \circ F$ decomposable functions and their connection to other nonsmooth classes [Mif77, Wri93, LOS00, Lew02, DL14, DDJ23]. Note that the requirement of $\nabla F(\cdot)$ continuous at 0 is used for the steady-state convergence result.

With Definition 1, we consider potentially nonsmooth SA updates (7) with an operator $\mathcal{T}$ satisfying the following assumption.

**Assumption 3** (Nonsmooth Class)**.** *The operator $\mathcal{T}$ is $g \circ F$ decomposable at its fixed point $\theta^*$. Explicitly, there exists $\epsilon > 0$ such that*

$$\mathcal{T}(\theta) = \theta^* + g(F(\theta - \theta^*)), \quad \forall \theta \in B^d(\theta^*, \epsilon)$$

*for some mappings $g : \mathbb{R}^m \to \mathbb{R}^d$ and $F : B^d(0, \epsilon) \to \mathbb{R}^m$ satisfying the requirements in Definition 1.*

Under Assumption 3, the operator $\mathcal{T}$ is at least locally $C^0$ at $\theta^*$. By setting $m = d$ and $g$ as the identity mapping, this assumption covers all locally $C^1$ and contractive $\mathcal{T}$, including SGD and Linear SA discussed earlier. In addition, this model covers operators $\mathcal{T}$ that are not differentiable at $\theta^*$, such as the example in (5) with $b = 0$ (corresponding to $g(\theta) = -\frac{|\theta|}{2}$ and $F(\theta) = \theta$), as well as the optimal Bellman operator that defines the Q learning algorithms (see Section 3).

Define the centered and rescaled iterate $Y_t^{(\alpha)} = \frac{\theta_t^{(\alpha)} - \theta^*}{\sqrt{\alpha}}$. Theorem 1 implies that $Y_t^{(\alpha)}$ converges weakly to a steady-state random variable $Y^{(\alpha)} := \frac{\theta^{(\alpha)} - \theta^*}{\sqrt{\alpha}}$ as $t \to \infty$. Focusing on SA satisfying the $g \circ F$ decomposability Assumption 3, our next theorem establishes steady-state convergence, that is, the convergence of $\{Y^{(\alpha)}\}_{\alpha \in (0, \bar{\alpha}')}$ as $\alpha \to 0$.

**Theorem 2** (Steady-State Convergence)**.** *Suppose that Assumption 1, Assumption 2(2) and Assumption 3 hold. There exists a unique random variable $Y$, depending only on $\mathcal{T}$ and $\mathrm{Var}(w_0)$, such that*

$$\lim_{\alpha \to 0} W_2 \left( \mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y) \right) = 0.$$

*Consequently, we have*

$$\mathbb{E}[\theta^{(\alpha)}] = \theta^* + \sqrt{\alpha}\mathbb{E}[Y] + o(\sqrt{\alpha}). \tag{9}$$

[2]A function $g : \mathbb{R}^m \to \mathbb{R}^d$ is homogeneous of degree 1 if $g(cx) = cg(x)$ for all $c \geq 0$ and $x \in \mathbb{R}^m$

Among other consequences, Theorem 2 implies that the steady-state bias, $\mathbb{E}[\theta^{(\alpha)}] - \theta^*$, is generally on the order of $O(\sqrt{\alpha})$ for small stepsizes $\alpha$. This result stands in sharp contrast to existing work on smooth SA, which has an order-wise smaller bias linear in $\alpha$. This $\sqrt{\alpha}$-bias property, which arises precisely due to the nonsmoothness of the SA dynamic, is further characterized in our next theorem. We highlight that Theorem 2 is a universality result: the limit $Y$ depends on the (zero-mean) noise $\{w_t\}_{t\geq 0}$ only through its variance and is otherwise independent of the noise distribution.

Note that Theorem 2 applies to any contractive SA within the $g \circ F$ decomposable class. In this generality, the convergence result in the theorem is asymptotic. The convergence rate and the specific order of the $o(\sqrt{\alpha})$ term depend on how fast $\nabla F(\theta)$ converges to $\nabla F(0)$; see equation (45) in our proof. It is possible to obtain explicit, nonasymptotic bounds on the convergence rate for specific SA dynamics and $\mathcal{T}$ operators. For example, in the next section, we establish an $\mathcal{O}(\alpha^{\frac{1}{4}})$ convergence rate for Q-learning.

The work [CMM22] also provides a steady-state convergence result but requires a strong uniqueness assumption, which is difficult to verify in most cases. Our results are established using a different technique, by directly proving the weak convergence of $Y^{(\alpha)}$ in $W_2$ using prelimit coupling. We outline the proof of Theorem 2 in Section 5.2, deferring the complete proof to Appendix C.

The following theorem provides a more fine-grained characterization of the expectation of the limit $Y$, which appears in the expression (9) for the steady-state bias.

**Theorem 3** (Bias Characterization). *Under the same setting as in Theorem 2, we have*

1. $\mathbb{E}[Y] = 0$ *if* $g$ *is continuously differentiable at* $0$ *or* $\nabla F(0) = 0$.
2. $\mathbb{E}[Y] \neq 0$ *if* $\mathrm{Var}(w_0)$ *is positive definite and there exists* $i \in [d]$ *such that the subdifferential or supdifferential of* $h_i(\theta) := g_i(\nabla F(0)\theta)$ *at 0 is not a singleton.*

Roughly speaking, the premise in Part (2) of the theorem implies that $\mathcal{T}$ is not differentiable at $\theta^*$ (otherwise its sub/supdifferential would be a singleton consisting of its gradient). In this case, provided that the noise $w_0$ is non-degenerate, we have $\mathbb{E}[Y] \neq 0$. Hence, equation (9) implies that the bias is on the order of $\Theta(\sqrt{\alpha})$. We conjecture that this result holds under more general settings of nonsmooth $\mathcal{T}$ where its sub/supdifferential may not exist. This $\sqrt{\alpha}$ order of the bias has important implications for bias reduction via the Richardson-Romberg extrapolation, which we discuss in Section 4.

Part (1) of Theorem 3, on the other hand, implies that for any smooth SA where $\mathcal{T}$ is continuously differentiable at $\theta^*$, the asymptotic bias is order-wise smaller than $\sqrt{\alpha}$. This result is consistent with those in [DDB20, HCX23b], which show that the asymptotic biases of SGD and Linear SA with i.i.d. noise are of order $\Theta(\alpha)$ and 0, respectively.

# 3 Q-learning: Nonsmooth Stochastic Approximation with Multiplicative Noise

In this section, we extend our results to Q-learning algorithms, which are nonsmooth SA procedures with multiplicative noise.

## 3.1 Model Setup

Consider a discounted Markov decision process (MDP) defined by the tuple $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$, where $\mathcal{S}$ and $\mathcal{A}$ are respectively the (finite) state and action spaces, $P : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is the transition kernel, $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the stochastic reward function, and $\gamma \in (0, 1)$ is the discount factor. Given a policy $\pi : \mathcal{S} \to \Delta(\mathcal{A})$, the Q-function $q^\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is defined as $q^\pi(s, a) = \mathbb{E}_\pi\big[\sum_{k=0}^{\infty} \gamma^k r_k(s_k, a_k) \mid s_0 = s, a_0 = a\big]$, where $a_k \sim \pi(\cdot|s_k), s_{k+1} \sim P(\cdot|s_k, a_k)$ and $r_k$ is an independent copy of $r$. The goal is to find an optimal policy $\pi^*$ that maximizes the Q-function. Below we often view $P$ as an $|\mathcal{S}||\mathcal{A}|$-by-$|\mathcal{A}|$ matrix, $r$ as a random vector in $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, and $q^\pi$ as a vector in $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$.

**Q-learning** [WD92] is a popular class of reinforcement learning methods that approximate the optimal Q-function $q^* = q^{\pi^*}$, from which one can recover the optimal policy as $\pi^*(s) \in \arg\max_{a\in\mathcal{A}} q^*(s, a)$, $\forall s \in \mathcal{S}$.

We consider a general form of Q-learning that iteratively generates a sequence of Q-function estimates, $\{q_t : \mathcal{S} \times \mathcal{A} \to \mathbb{R}\}_{t\geq 0}$, according to the following recursion:

$$q_{t+1}^{(\alpha)} = q_t^{(\alpha)} + \alpha D_t\big(\gamma P_t f(q_t^{(\alpha)}) - q_t^{(\alpha)} + r_t\big), \tag{10}$$

where the function $f : \mathbb{R}^{|\mathcal{S}||\mathcal{A}|} \to \mathbb{R}^{|\mathcal{S}|}$ is given by

$$f_s(q) := \max_{a\in\mathcal{A}} q(s,a), \quad \forall s \in \mathcal{S},$$

and $\{(D_t, P_t, r_t)\}_{t\geq 0}$ are i.i.d. random matrices/vectors satisfying: (i) $D = \mathbb{E}[D_0]$ is a $|\mathcal{S}||\mathcal{A}|$-by-$|\mathcal{S}||\mathcal{A}|$ diagonal matrix with $D_{ii} \in (0,1], \forall i \in \mathcal{S} \times \mathcal{A}$; (ii) $\mathbb{E}[P_0] = P$; (iii) $\{r_t\}_{t\geq 0}$ are independent copies of $r$. Here $D_t, P_t$ and $r_t$ correspond to the empirical state-action distribution, empirical transition and empirical reward function, respectively, observed at the $t$-th iteration.

We discuss two important special cases of the above model.

- **Synchronous Q-learning [Wai19]:** At each time step $t$ and for each state-action pair $(s,a)$, we observe a reward $r_t(s,a) \overset{\mathrm{d}}{=} r(s,a)$ and a next state $x_t(s,a)$ drawn from the transition kernel $P(\cdot|s,a)$. The Q-function estimates are updated as

$$q_{t+1}^{(\alpha)}(s,a) = q_t^{(\alpha)}(s,a) + \alpha\Big(\gamma \max_{a'\in\mathcal{A}} q_t^{(\alpha)}(x_t(s,a),a') - q_t^{(\alpha)}(s,a) + r_t(s,a)\Big), \;\; \forall (s,a) \in \mathcal{S}\times\mathcal{A}.$$

  Synchronous Q-learning corresponds to the update rule (10) where $D_t \equiv I$ and $P_t$ is a binary random matrix whose $(s,a)$-th row is independently distributed as $\mathrm{Multi}\,(P\,(\cdot|s,a)\,,1)$.

- **Asynchronous Q-learning [CMZ23]:** At each time step $t$, we observe a state-action pair $(s_t, a_t) \sim \kappa_b$, where the distribution $\kappa_b \in \Delta(\mathcal{S}\times\mathcal{A})$ can be the stationary state-action distribution of some behavior policy. Conditioned on $(s_t,a_t)$, we observe the reward $r_t(s_t,a_t) \overset{\mathrm{d}}{=} r(s_t,a_t)$ and the next state $s'_{t+1}$ drawn according to $P(\cdot|s_t,a_t)$. The Q-function estimates are updated as

$$\begin{aligned} q_{t+1}^{(\alpha)}(s_t,a_t) &= q_t^{(\alpha)}(s_t,a_t) + \alpha\Big(\gamma \max_{a'\in\mathcal{A}} q_t^{(\alpha)}(s'_{t+1},a') - q_t^{(\alpha)}(s_t,a_t) + r_t(s_t,a_t)\Big), \\ q_{t+1}^{(\alpha)}(s,a) &= q_t^{(\alpha)}(s,a), \;\; \forall (s,a) \neq (s_t,a_t). \end{aligned}$$

  Asynchronous Q-learning corresponds to the update rule (10) with $\mathrm{diag}(D_t) \sim \mathrm{Multi}\,(\kappa_b, 1)$ and the same $P_t$ before. Note that only the $(s_t,a_t)$ entry of $q_t^{(\alpha)}$ is updated at iteration $t$, with $D_t$ acting as the corresponding mask matrix.

With other choices of $(D_t, P_t, r_t)$, the update rule (10) can capture other forms of Q-learning with different sampling models.

The Q-learning update (10) can be cast as contractive SA. To this end, define a random operator $\widetilde{\mathcal{H}}$ by

$$\widetilde{\mathcal{H}}(q; \{D_0, P_0, r_0\}) = \gamma D_0 P_0 f(q) + (I - D_0)q + D_0 r_0, \qquad \forall q \in \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}.$$

Denote by $\mathcal{H} : \mathbb{R}^{|\mathcal{S}||\mathcal{A}|} \to \mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$ the expected operator, where

$$\mathcal{H}(q) := \mathbb{E}_{\{D_0,P_0,r_0\}}\big[\widetilde{\mathcal{H}}(q, \{D_0,P_0,r_0\})\big] = \gamma DPf(q) + (I-D)q + D\bar{r},$$

with $\bar{r} := \mathbb{E}[r_0]$. It can be verified that $\mathcal{H}$ is a $\gamma_0$-contractive operator with respect to the infinity norm $\|\cdot\|_\infty$, where $\gamma_0 = 1 - (1-\gamma)\min_{i\in\mathcal{S}\times\mathcal{A}} D_{ii} \in (0,1)$ [CMZ23, Proposition 3.3]. Moreover, the optimal Q-function $q^*$ is the unique solution to the fixed point equation $\mathcal{H}(q^*) = q^*$, which can be seen to be equivalent to the optimal Bellman equation. To be consistent with the additive noise setting, below we use $\|\cdot\|_c$ to denote $\|\cdot\|_\infty$.

With the above notations, the Q-learning update (10) can rewritten as a contractive SA iteration:

$$q_{t+1}^{(\alpha)} = q_t^{(\alpha)} + \alpha\Big(\widetilde{\mathcal{H}}\big(q_t^{(\alpha)}; \{D_t, P_t, r_t\}\big) - q_t^{(\alpha)}\Big). \tag{11}$$

Note that the iteration (11) is nonsmooth due to the max operation in the function $f$ in (10); moreover, it involves *multiplicative* noise due to multiplication with the random matrices $D_t$ and $P_t$, which are viewed as noisy versions of $D$ and $P$.

For the noise we consider the following moment assumption, indexed by an integer $n \geq 1$:

**Assumption 4** ($\boldsymbol{n}$)**.** *The random variables* $\{(D_t, P_t, r_t)\}_{t\geq 0}$ *have finite* $(2n)$*-th moments.*

Below we analyze Q-learning. Our results parallel those in the additive noise setting, but the analysis is significantly more challenging because of the multiplicative noise.

### 3.2 Moments Bounds and Convergence to Stationary Distribution

We first derive finite-time upper bounds on $\mathbb{E}[\|q_t^{(\alpha)} - q^*\|_c^{2n}]$, the $2n$-th moments of the estimation errors.

**Proposition 2** (Moment Bounds)**.** *For each integer $n \geq 1$, under Assumption 4($\boldsymbol{n}$), there exists $\alpha_n > 0$ such that for any $\alpha \leq \alpha_n$, there exists $t_{\alpha,n} \geq 0$ such that*

$$\mathbb{E}[\|q_t^{(\alpha)} - q^*\|_c^{2n}] \leq c_n \mathbb{E}[\|q_{t_{\alpha,n}}^{(\alpha)} - q^*\|_c^{2n}](1-\alpha(1-\sqrt{\gamma_0}))^{t-t_{\alpha,n}} + c_n' \alpha^n, \quad t \geq t_{\alpha,n}, \tag{12}$$

*where $c_n$ and $c_n'$ are constants that are independent of $\alpha$ and $t$. Moreover, $t_{\alpha,1} = 0$.*

Similarly to the additive noise setting, we mostly use Proposition 2 with $n \in \{1, 2\}$ for the subsequent analysis. In particular, using Proposition 2 with $n = 1$, we can establish the weak convergence in $W_2$ of the stochastic process $\{q_t^{(\alpha)}\}_{t\geq 0}$ to a unique stationary distribution, and further characterize its geometric convergence rate. This is done in the following theorem.

**Theorem 4** (Distributional Convergence)**.** *Under Assumption 4(1), there exists $\bar{\alpha}_0' > 0$ such that for $\forall \alpha \leq \bar{\alpha}_0'$ and all initial distribution of $q_0^{(\alpha)}$, the sequence $\{q_t^{(\alpha)}\}_{t\geq 0}$ converges geometrically fast in $W_2$ to a random variable $q^{(\alpha)}$ with*

$$W_2^2\big(\mathcal{L}(q_t^{(\alpha)}), \mathcal{L}(q^{(\alpha)})\big) \leq c \cdot (1-\alpha(1-\sqrt{\gamma_0}))^t, \quad \forall t \geq 0,$$

*where $c$ is a constant independent of $\alpha$ and $t$. Moreover, $\mathbb{E}[\|q^{(\alpha)} - q^*\|_2^2] \in \mathcal{O}(\alpha)$.*

The proofs of Proposition 2 and Theorem 4 use the generalized Moreau envelop of the contraction norm $\|\cdot\|_c$, similarly to those of Proposition 1 and Theorem 1 for the additive noise setting. However, the multiplicative noise makes the analysis more involved. We discuss the key difference in Section 5.1. The complete proofs of Proposition 2 and Theorem 4 can be found in Appendix E and Appendix F, respectively.

### 3.3 Steady-State Convergence and Bias Characterization

Consider the centered/rescaled iterate $Y_t^{(\alpha)} = (q_t^{(\alpha)} - q^*)/\sqrt{\alpha}$. Theorem 4 implies that the sequence $\{Y_t^{(\alpha)}\}_{t\geq 0}$ converges weakly to a steady-state random variable $Y^{(\alpha)} = (q^{(\alpha)} - q^*)/\sqrt{\alpha}$. In the following theorem, we establish the steady-state convergence for $\{Y^{(\alpha)}\}$ as $\alpha \to 0$.

**Theorem 5** (Steady-State Convergence)**.** *Suppose Assumption 4(2) holds. There exists a unique random variable $Y$ such that*

$$\lim_{\alpha\to 0} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) = 0.$$

*Furthermore, we have $W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) \in \mathcal{O}(\alpha^{\frac{1}{4}})$, which implies that*

$$\mathbb{E}[q^{(\alpha)}] = q^* + \sqrt{\alpha}\mathbb{E}[Y] + \mathcal{O}(\alpha^{\frac{3}{4}}). \tag{13}$$

A few remarks are in order. Similar to the additive noise setting, Theorem 5 indicates that the steady-state bias of Q-learning, $\mathbb{E}[q^{(\alpha)}] - q^*$, is in general of order $O(\sqrt{\alpha})$ for small stepsize $\alpha$. Again, this distinctive $\sqrt{\alpha}$-bias result is due to the nonsmooth nature of the Q-learning dynamic; cf. function $f$ in equation (10). Our next theorem provides a more precise characterization on the bias.

The proof of Theorem 5 also uses our prelimit coupling technique, which can handle the multiplicative noise. On the contrary, the work [CMM22] only considers the additive noise setting and it is unclear how to generalize their analysis to the multiplicative noise case. Moreover, as a byproduct of our prelimit coupling, for the explicit Q-learning dynamic, we can obtain an $\mathcal{O}(\alpha^{\frac{1}{4}})$ convergence rate of $Y^{(\alpha)}$ to the limit $Y$. The proof of Theorem 5 is provided in Appendix G.

To discuss further properties of the limit $Y$, we need some definitions. We say that a state $s' \in \mathcal{S}$ is **rooted** if

$$P(s'|s,a) = 0, \quad \forall (s,a) \in \mathcal{S} \times \mathcal{A}.$$

Intuitively, a state $s'$ is rooted if it is not accessible from any other state in the MDP. Using the optimal Q-function $q^*$, we define $\mathcal{A}^*(s) := \arg\max_{a\in\mathcal{A}} q^*(s,a)$ as the optimal action set for each state $s \in \mathcal{S}$. Note that the action distribution $\pi^*(\cdot|s)$ of the optimal policy is supported on the set $\mathcal{A}^*(s)$ for each $s \in \mathcal{S}$. We say that a state $s \in \mathcal{S}$ is **tied** if $|\mathcal{A}^*(s)| > 1$, i.e., there is a tie among multiple optimal actions for $s$.

We classify all MDPs into two types:

- **Type A**: There exists at least one state that is tied and not rooted.
- **Type B** (i.e., not Type A): There is no tied state, or all tied states are rooted.

For each type of MDPs, we provide a more fine-grained characterization for the expectation of the limit $Y$ in the following theorem. Recall that $\mathbb{E}[Y]$ determines the order of the steady-state bias by equation (13).

**Theorem 6** (Bias Characterization)**.** *Under the same setting as Theorem 5, we have*

1. $\mathbb{E}[Y] \neq 0$ *if the underlying MDP is in Type A and* $\mathrm{Var}(\widetilde{\mathcal{H}}(q^*, \{D_0, P_0, r_0\}))$ *is positive definite.*
2. $\mathbb{E}[Y] = 0$ *if the underlying MDP is in Type B.*
3. *If the underlying MDP is in Type B and Assumption 4(n) holds for $n \geq 2$, then* $\mathbb{E}[q^{(\alpha)}] = q^* + \mathcal{O}(\alpha^n)$.

Note that for a Type-A MDP, the optimal policy is not unique due to the existence of multiple optimal actions for at least one state. In this case, Part (1) of the theorem implies $\mathbb{E}[Y] \neq 0$. Consequently, the asymptotic bias $\mathbb{E}[q^{(\alpha)}] - q^*$ of Q-learning is of $\sqrt{\alpha}$ order. As we will see in Section 4, the precise characterization of order-$\sqrt{\alpha}$ bias allows one to use the Richardson-Romberg extrapolation for bias reduction.

Parts (2) and (3) of the theorem imply that for Type-B MDPs (i.e., those with a unique optimal policy), the asymptotic bias can be controlled by the $n$-th order of the stepsize, as long as the noise has finite $2n$-th moment. For Q-learning, the random matrices $\{D_t, P_t\}_{t\geq 0}$ are bounded and thus all their moments are finite. If the rewards $\{r_t\}_{t\geq 0}$ also have finite arbitrary moments (e.g., they are Gaussian distributed or bounded), then the asymptotic bias is $\mathcal{O}(\alpha^n)$ for any $n \geq 1$, that is, the bias decays superpolynoimally with respect to the stepsize.

# 4 Polyak-Ruppert Averaging and Richardson-Romberg Extrapolation

In this section, we study the implications of our theoretical results for iterate averaging and extrapolation. In particular, we consider applying Polyak-Ruppert (PR) tail averaging [Rup88, PJ92, JKK+18] and Richardson-Romberg (RR) extrapolation [Hil87] to the iterates generated by contractive SA algorithms, and investigate the resulting estimation errors and biases in the presence of nonsmoothness.

To this end, we will first state two general results for PR averaging and RR extrapolation, respectively. We remark that these general results cover settings broader than those considered in this paper and may be of independent interest. We then apply these results to the contractive SA and Q-learning procedures studied in Section 2 and Section 3.

Let $\{\theta_t^{(\alpha)}\}_{t\geq 0}$ be a sequence of (raw) iterates in $\mathbb{R}^d$ generated by an SA procedure of the form

$$\theta_{t+1}^{(\alpha)} = \theta_t^{(\alpha)} + \alpha\big(\mathcal{H}(\theta_t^{(\alpha)}, w_t) - \theta_t^{(\alpha)}\big) \tag{14}$$

with a constant stepsize $\alpha > 0$. We assume that the noise sequence $\{w_t\}_{t\geq 0}$ is a uniformly ergodic Markov chain defined on a general state space $\mathcal{W}$ with transition kernel $p$ and stationary distribution $\mu_{\mathcal{W}}$, and let $\tau_\alpha$ denote its $\alpha$-mixing time, i.e., $\tau_\alpha := \min\{t \geq 0 : \max_{x\in X} \|p^t(x,\cdot) - \mu_{\mathcal{X}}\|_{\mathrm{TV}} \leq \alpha\}$, where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm. Note that a sequence of i.i.d. noise $\{w_t\}_{t\geq 0}$ is a uniformly ergodic Markov chain with $\tau_\alpha = 1$ for all $\alpha > 0$.

We introduce two conditions on the raw SA iterates $\{\theta_t^{(\alpha)}\}_{t\geq 0}$, which allow us to quantify the performance of PR averaging and RR extrapolation with respect to a target vector $\theta^*$.

**Condition 1** (Distributional convergence)**.** *There exist constants $C_0, C_1, \bar{\alpha} > 0$ satisfying $0 < 1 - \bar{\alpha}C_1 < 1$ such that for some random variable $\theta^{(\alpha)}$ it holds that*

$$W_2^2(\mathcal{L}(\theta_t^{(\alpha)}), \mathcal{L}(\theta^{(\alpha)})) \le C_0 \cdot (1 - \alpha C_1)^t, \quad \forall t \ge \tau_\alpha \text{ and } \forall \alpha \le \bar{\alpha}.$$

**Condition 2** (Asymptotic bias and variance)**.** *There exist constants $\beta > 0$ and $\delta \ge 0$ such that*

$$\mathbb{E}[\theta^{(\alpha)}] = \theta^* + \alpha^\beta B + o(\alpha^{\beta+\delta}), \tag{15}$$

*where $B \in \mathbb{R}^d$ is a vector independent of $t$ and $\alpha$. Moreover, $\mathbb{E}[\|\theta^{(\alpha)} - \theta^*\|_2^2] \in \mathcal{O}(\alpha\tau_\alpha)$.*

Note that as the stepsize $\alpha$ gets larger, we have a faster geometric convergence in Condition 1 but a greater bias in Condition 2. We later verify these conditions under our contractive SA and Q-learning settings.

## 4.1 Polyak-Ruppert Tail Averaging

Polyak-Ruppert (PR) averaging procedure [Rup88, PJ92, JKK+18] is a popular procedure for reducing the variance of the SA iterates and accelerating the convergence. Specifically, given a burn-in period $k_0 \ge 0$, we compute the tail-averaged iterates as:

$$\bar{\theta}_{k_0,k}^{(\alpha)} := \frac{1}{k - k_0} \sum_{t=k_0}^{k-1} \theta_t^{(\alpha)}, \quad \text{for } k \ge k_0 + 1.$$

The following proposition provides non-asymptotic bounds for the first two moments of the tailed-averaged iterate $\bar{\theta}_{k_0,k}^{(\alpha)}$.

**Proposition 3.** *Under Conditions 1 and 2, we have for all $k_0 \ge \frac{2}{\alpha C_1} \log\left(\frac{1}{\alpha\tau_\alpha}\right)$ and $k \ge k_0 + \tau_\alpha$:*

$$\mathbb{E}\big[\bar{\theta}_{k_0,k}^{(\alpha)}\big] - \theta^* = \alpha^\beta B + o(\alpha^{\beta+\delta}) + \mathcal{O}\Big(\frac{1}{\alpha(k-k_0)} \exp\big(-\frac{\alpha C_1 k_0}{2}\big)\Big), \tag{16}$$

$$\mathbb{E}\big[\left(\bar{\theta}_{k_0,k} - \theta^*\right)\left(\bar{\theta}_{k_0,k} - \theta^*\right)^\top\big] = \alpha^{2\beta} BB^\top + o(\alpha^{2\beta+\delta}) + \mathcal{O}\Big(\frac{1}{\alpha\left(k-k_0\right)^2} \exp\big(-\frac{\alpha C_1 k_0}{2}\big)\Big) + \mathcal{O}\Big(\frac{\tau_\alpha}{k-k_0}\Big). \tag{17}$$

The proof is provided in Section I, generalizing the arguments from [HCX23b] on Linear SA. As a typical application of the above result, let us set the burn-in parameter as $k_0 = k/2$ and consider the second moment bound in Equation (17). The first two terms on the right-hand side of (17) correspond to the squared asymptotic bias, which is the same as the bias of the raw iterates in Condition 2 and cannot be reduced by averaging. The third term captures the optimization error, which decays geometrically in $k$ due to the geometric distributional convergence in Conditions 1. The last right hand side term of (17) corresponds to the variance of averaged iterate $\bar{\theta}_{k/2,k}$, which decays at a rate $\mathcal{O}(1/k)$ due to averaging over $k/2$ raw iterates that are geometrically mixed.

## 4.2 Richardson-Romberg Extrapolation

With the fine-grained characterization of the asymptotic bias in Condition 2, one can use the RR extrapolation technique [Hil87] to reduce the bias to a higher order term of the stepsize $\alpha$. In particular, we consider first-order RR extrapolation, where we run two SA recursions (14) in parallel under two different stepsizes $\alpha$ and $2\alpha$, under the *same* sequence of noise $\{w_t\}_{t\ge0}$. The resulting tail-averaged iterates $\bar{\theta}_{k_0,k}^{(\alpha)}$ and $\bar{\theta}_{k_0,k}^{(2\alpha)}$ are defined as before. The RR extrapolated iterates are then computed as follows as a linear combination of the two averaged iterates:

$$\tilde{\theta}_{k_0,k}^{(\alpha)} = \frac{2^\beta}{2^\beta - 1} \bar{\theta}_{k_0,k}^{(\alpha)} - \frac{1}{2^\beta - 1} \bar{\theta}_{k_0,k}^{(2\alpha)}. \tag{18}$$

The coefficients of the above linear combination are chosen such that we cancel out the dominating terms $\alpha^\beta$ and $(2\alpha)^\beta$ in the biases.

**Proposition 4.** *Under Conditions 1 and 2, the RR extrapolated iterates defined in (18) satisfy the following bounds for all $k_0 \geq \frac{2}{\alpha C_1}\log\left(\frac{1}{\alpha\tau_\alpha}\right)$ and $k \geq k_0 + \tau_\alpha$:*

$$\mathbb{E}\left[\tilde{\theta}^{(\alpha)}_{k_0,k}\right] - \theta^* \in o(\alpha^{\beta+\delta}) + \mathcal{O}\Big(\frac{1}{\alpha(k-k_0)}\exp\big(-\frac{\alpha C_1 k_0}{2}\big)\Big), \tag{19}$$

$$\mathbb{E}\left[(\tilde{\theta}_{k-k_0} - \theta^*)(\tilde{\theta}_{k-k_0} - \theta^*)^\top\right] \in o(\alpha^{2\beta+2\delta}) + \mathcal{O}\Big(\frac{1}{\alpha\,(k-k_0)^2}\exp\big(-\frac{\alpha C_1 k_0}{2}\big)\Big) + \mathcal{O}\Big(\frac{2^{2\beta}}{(2^\beta-1)^2}\frac{\tau_\alpha}{k-k_0}\Big). \tag{20}$$

The proof is provided in Section J. Again focusing on the second moment bound (20) with $k_0 = k/2$, we see that the squared bias is reduced to $o(\alpha^{2\beta+2\delta})$, whereas we retain the geometric convergence of the optimization error (second right hand side term) and the $1/k$ rate of the variance (third right hand side term).

### 4.3 Applications to Contractive SA with Additive Noise and Q-Learning

First consider the contractive SA dynamic (7) with additive noise from Section 2. By Theorem 1, Condition 1 holds with $C_1 = 1-\sqrt{\gamma}$ and $\tau_\alpha = 1$, By Theorem 2, Condition 2 holds with $\tau_\alpha = 1, B = \mathbb{E}[Y], \beta = \frac{1}{2}$ and $\delta = 0$. Hence, Proposition 4 with $k_0 = k/2$ implies the following MSE bound:

$$\mathbb{E}\left\|\tilde{\theta}_{k/2} - \theta^*\right\|^2 \in o(\alpha) + \mathcal{O}\Big(\frac{1}{\alpha k^2}\exp\big(-\frac{\alpha(1-\sqrt{\gamma})k}{4}\big)\Big) + \mathcal{O}\Big(\frac{1}{k}\Big).$$

Similarly, for the Q-learning dynamic (10) in Section 3, Condition 1 holds with $C_1 = 1-\sqrt{\gamma_0}$ and $\tau_\alpha = 1$ by Theorem 4; Condition 2 holds with $\tau_\alpha = 1, B = \mathbb{E}[Y], \beta = \frac{1}{2}$ and $\delta = 1/4$ by Theorem 5. Consequently, we have the following MSE bound:

$$\mathbb{E}\left\|\tilde{\theta}_{k/2} - \theta^*\right\|^2 \in o(\alpha^{3/2}) + \mathcal{O}\Big(\frac{1}{\alpha k^2}\exp\big(-\frac{\alpha(1-\sqrt{\gamma_0})k}{4}\big)\Big) + \mathcal{O}\Big(\frac{1}{k}\Big).$$

In both cases, the asymptotic bias of the raw iterate is on the order of $\sqrt{\alpha}\mathbb{E}[Y]$, which is reduced to $o(\sqrt{\alpha})$ or $o(\alpha^{3/4})$ by RR extrapolation. We emphasize that the order of the bias here is different from the $O(\alpha)$ bias typically seen in smooth SGD/SA dynamics [DDB20, HCX23b]. Knowledge of the correct bias order, as provided by our theoretical results, is crucial for the RR extrapolation to be effective. We note that if $\mathbb{E}[Y] = 0$, the bias of the raw iterate is already $o(\sqrt{\alpha})$, in which case the above RR extrapolation scheme may not lead to further improvement but it does not hurt the performance either (up to constants). In Section 6, we provide numerical experiments demonstrating bias reduction by RR extrapolation.

## 5 Proof Outline

In this section, we outline the proofs of our main theoretical results. We focus on the additive noise setting and discuss how to generalize to the Q-learning setting with multiplicative noise. Without additional explanation, we default iterates are in $\mathbb{R}^d$.

Recall that $\mathcal{T}$ is contractive w.r.t. the norm $\|\cdot\|_c$. As $\phi(\cdot) = \frac{1}{2}\|\cdot\|_c^2$ is not necessarily differentiable, our analysis makes use of its *generalized Moreau envelope* [CMSS23, CMSS20], which can be thought of as a smooth surrogate of $\phi$. In particular, let $h(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, which is 1-smooth with respect to $\|\cdot\|_2$. Because all norms on $\mathbb{R}^d$ are equivalent [Fol99], there exist two positive constants $l_{cs}$ and $u_{cs}$ such that $l_{cs}\|\cdot\|_2 \leq \|\cdot\|_c \leq u_{cs}\|\cdot\|_2$. The generalized Moreau envelope $M_\eta : \mathbb{R}^d \to \mathbb{R}$ of $\phi$ with respect to $h$ is defined as

$$M_\eta(x) = \inf_{u\in\mathbb{R}^d}\Big\{\phi(u) + \frac{1}{\eta}h(x-u)\Big\}, \qquad \forall x \in \mathbb{R}^d. \tag{21}$$

The basic properties of $M_\eta$ are summarized below. The proof can be found in [CMSS23, Proposition 1] and [CMSS20, Lemma A.1].

**Proposition 5.** *$M_\eta$ has the following properties: (1) $M_\eta$ is convex and $\frac{1}{\eta}$-smooth with respect to $\|\cdot\|_2$; (2) there exists a norm $\|\cdot\|_m$ such that $M_\eta(x) = \frac{1}{2}\|x\|_m^2$; (3) it holds that $l_{cm}\|\cdot\|_m \leq \|\cdot\|_c \leq u_{cm}\|\cdot\|_m$, where $l_{cm} = (1+\eta l_{cs}^2)^{\frac{1}{2}}$ and $u_{cm} = (1+\eta u_{cs}^2)^{\frac{1}{2}}$; (4) $\langle\nabla M_\eta(x), y\rangle \leq \|x\|_m\|y\|_m$, $\forall x, y \in \mathbb{R}^d$.*

In this section, we omit the subscript in $\theta_t^{(\alpha)}$ when the dependence on the stepsize $\alpha$ is clear.

## 5.1 Proof Outline for Proposition 1 (Moment Bounds) and Theorem 1 (Distributional Convergence)

**Moment Bounds.** To bound the $(2n)$-th moment $\mathbb{E}\|\theta_t^{(\alpha)} - \theta^*\|_c^{2n}$, we use the generalized Moreau envelope $M_\eta$ as a Lyapunov function and generalize the arguments in [CMSS20, CMSS23] to higher moments by induction on $n$. In particular, using the contractive property of $\mathcal{T}$ and the properties of $M_\eta$, we can obtain

$$M_\eta(\theta_{t+1} - \theta^*) \le \underbrace{(1 - \alpha(1 - \sqrt{\gamma}))\, M_\eta(\theta_t - \theta^*)}_{T_1} + \underbrace{\alpha\langle \nabla M_\eta(\theta_t - \theta^*), w_t\rangle}_{T_2} + \underbrace{\alpha^2 \|w_t\|_c^2/\eta l_{cs}^2}_{T_3}. \tag{22}$$

Taking the $n$-th moment of both sides gives $\mathbb{E}[M_\eta^n(\theta_{t+1} - \theta^*)] \le \mathbb{E}[(T_1 + T_2 + T_3)^n]$. Expanding the right hand side and noting that $w_t$ is zero mean, we derive $\mathbb{E}[nT_1^{n-1}T_2] = 0$ and $\mathbb{E}[T_1^n] \le \big(1 - \alpha(1 - \sqrt{\gamma})\big)\, \mathbb{E}[M_\eta^n(\theta_t - \theta^*)]$. A careful calculation using the induction hypothesis shows that the cross terms satisfy $\mathbb{E}[\binom{n}{a}\binom{n-a}{b} T_1^a T_2^b T_3^c] \in \mathcal{O}(\alpha^{n+1})$. Combining these bounds gives

$$\mathbb{E}[M_\eta^n(\theta_t - \theta^*)] \le \mathbb{E}[M_\eta^n(\theta_{t_\alpha} - \theta^*)](1 - \alpha(1 - \sqrt{\gamma}))^{t - t_{\alpha,n}} + c_n \alpha^n, \quad \forall t \ge t_{\alpha,n}, \tag{23}$$

from which the desired moment bounds follow in light of part (c) of Proposition 5.

**Distributional Convergence.** Similarly to [DDB20, HCX23b, ZX24], the key step in proving Theorem 1 is establishing the convergence of $W_2^2\big(\mathcal{L}(\theta_t^{(\alpha)}), \mathcal{L}({\theta'}_t^{(\alpha)})\big)$ for two iterate sequences $\{\theta_t^{(\alpha)}\}_{t\ge 0}$ and $\{{\theta'}_t^{(\alpha)}\}_{t\ge 0}$ with different initialization. Coupling these two sequences by sharing the noise sequence $\{w_t\}_{t\ge 0}$, we further reduce the problem to bounding $\mathbb{E}\|\theta_t^{(\alpha)} - {\theta'}_t^{(\alpha)}\|_c^2$ and, in turn, to bounding $\mathbb{E}[M_\eta(\theta_t^{(\alpha)} - {\theta'}_t^{(\alpha)})]$. The latter can be done using an argument similar to equation (22).

**Proof for Q-learning:** Due to multiplicative noise, the error term $w_t$ depends on the iterate $q_t^{(\alpha)}$ itself. A more involved analysis using the structure of Q-learning allows us to control resulting additional error terms, thereby proving Proposition 2 and Theorem 4.

## 5.2 Proof Outline for Theorem 2 (Steady State Convergence)

The proof consists of three steps and employs coupling arguments applied to the prelimit rescaled random variables $Y_t^{(\alpha)} := (\theta_t^{(\alpha)} - \theta^*)/\sqrt{\alpha}$ with $\alpha > 0$ and $t < \infty$.

### 5.2.1 Step 1: Gaussian Noise and Rational Stepsize

In this step, we assume that the noise $w_t$ is Gaussian. We prove that $\{\mathcal{L}(Y^{(\alpha)})\}_{\alpha \in \mathbb{Q}^+}$ form a Cauchy sequence with respect to $W_2$, thus converging to a unique limit $\mathcal{L}(Y)$, i.e., $\lim_{\alpha\to 0, \alpha \in \mathbb{Q}^+} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) = 0$.

To this end, we first consider two stepsizes $\alpha$ and $\alpha/k$, where $k \in \mathbb{N}^+$ and study the rescaled iterates $Y_t^{(\alpha)}$ and $Y_t^{(\alpha/k)}$ generated by equation (7). As discussed in Section 1.2, we couple these two sequences such that one step of $Y_t^{(\alpha)}$ corresponds to $k$ steps of $Y_t^{(\alpha/k)}$. We take the generalized Moreau envelope of the difference sequence, $\{Y_t^{(\alpha)} - Y_{kt}^{(\alpha/k)}\}_{t\ge 0}$, with the goal of showing that

$$\mathbb{E}[M_\eta(Y_{t+1}^{(\alpha)} - Y_{kt+k}^{(\alpha/k)})] \le (1 - \alpha(1 - \sqrt{\gamma}))\, \mathbb{E}[M_\eta(Y_t^{(\alpha)} - Y_{kt}^{(\alpha/k)})] + \mathcal{O}(\alpha^{r_1 + 1}), \tag{24}$$

where $r_1$ is a constant. The proof of Equation (24) makes use of the $g \circ F$ decomposibility of the operator $\mathcal{T}$ and is the most critical sub-step in Step 1. Consequently, we have

$$\lim_{t\to\infty} \mathbb{E}[M_\eta(Y_t^{(\alpha)} - Y_{kt}^{(\alpha/k)})] \in \mathcal{O}(\alpha^{r_1}).$$

Combining with the distributional convergence result in Theorem 1, we obtain that

$$\begin{aligned} & W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\big) \\ \le & \lim_{t\to\infty} \Big[ W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y_t^{(\alpha)})\big) + W_2\big(\mathcal{L}(Y_t^{(\alpha)}), \mathcal{L}(Y_{kt}^{(\alpha/k)})\big) + W_2\big(\mathcal{L}(Y_{kt}^{(\alpha/k)}), \mathcal{L}(Y^{(\alpha/k)})\big) \Big] \end{aligned}$$

$$\leq \lim_{t\to\infty} \sqrt{\mathbb{E}[\|Y_t^{(\alpha)} - Y_{kt}^{(\alpha/k)}\|_c^2]} \leq \lim_{t\to\infty} \sqrt{2u_{cm}^2 \mathbb{E}[M_\eta(Y_t^{(\alpha)} - Y_{kt}^{(\alpha/k)})]} \in \mathcal{O}(\alpha^{\frac{r_1}{2}}).$$

Next we consider stepsizes $\alpha > 0$ and $\alpha/k$ with $k = p/q \in \mathbb{Q}^+$, where $p, q \in \mathbb{N}^+$ and $p > q$. By triangle equality for the $W_2$ metric, we have

$$\begin{aligned} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\big) \leq& W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/q)})\big) + W_2\big(\mathcal{L}(Y^{(\alpha/q)}), \mathcal{L}(Y^{(p\cdot\alpha/q)})\big) \\ \leq& \mathcal{O}(\alpha^{\frac{r_1}{2}}) + \mathcal{O}(\alpha^{\frac{r_1}{2}}) \in \mathcal{O}(\alpha^{\frac{r_1}{2}}). \end{aligned} \tag{25}$$

Therefore, for any rational sub-sequence $\{\alpha_j\}_{j=0}^\infty$ with $\alpha_j \to 0$, $\{\mathcal{L}(Y^{(\alpha_j)})\}_{j=0}^\infty$ is a Cauchy sequence with respect to $W_2$. Consequently, a limit $\mathcal{L}(Y)$ exists. Since two rational sub-sequences can be merged into one rational sub-sequence by staggered placement, the limit is unique.

#### 5.2.2 Step 2: General Stepsize

Still assuming Gaussian noise, we generalize the result in Step 1 to general stepsize. To this end, we prove that $Y^{(\alpha)}$ is continuous in $\alpha$ with respect to $W_2$. More specifically, we consider two real-valued stepsizes $\alpha$ and $\alpha'$, and couple the corresponding two sequences $Y_t^{(\alpha)}$ and $Y^{(\alpha')}$ by letting them share the same noise $\{w_t\}_{t\geq 0}$, as detailed in Section 1.2. We then obtain the following equation by applying the generalized Moreau envelope on the difference sequence $\{Y_t^{(\alpha)} - Y_t^{(\alpha')}\}_{t\geq 0}$:

$$\mathbb{E}\big[M_\eta(Y_{t+1}^{(\alpha)} - Y_{t+1}^{(\alpha')})\big] \leq (1 - \mathcal{O}(\alpha))\, \mathbb{E}\big[M_\eta(Y_t^{(\alpha)} - Y_t^{(\alpha')})\big] + \mathcal{O}\big(|\alpha - \alpha'|\big),$$

which implies that

$$\lim_{t\to\infty} \mathbb{E}[M_\eta(Y_t^{(\alpha)} - Y_t^{(\alpha')})] = \mathcal{O}\big(|\alpha - \alpha'|\big).$$

Following similar arguments as in Step 1, we have $\lim_{\alpha'\to\alpha} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha')})\big) = 0$, thereby concluding that $Y^{(\alpha)}$ is continuous in $\alpha$ with respect to $W_2$. Since the real numbers have the rational numbers as a dense subset, we obtain the desired convergence result $\lim_{\alpha\to 0} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) = 0$.

To obtain an explicit convergence rate to the above limit, we observe that

$$W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) \leq W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\big) + W_2\big(\mathcal{L}(Y^{(\alpha/k)}), \mathcal{L}(Y)\big), \qquad \forall k \in \mathbb{N}^+.$$

Sending $k \to \infty$ on both sides and applying the bound (25), we obtain the desired rate:

$$W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) \in \mathcal{O}(\alpha^{r_1/2}).$$

#### 5.2.3 Step 3: General Noise

Steps 1 and 2 above complete the proof of Theorem 2 for Gaussian noise. In this step, we consider general noise. To this end, we consider two sequences $Y_t'^{(\alpha)}$ and $Y_t^{(\alpha)}$, where $Y_t'^{(\alpha)}$ is driven by some general noise $w_t'$, and $Y_t^{(\alpha)}$ is driven by Gaussian noise $w_t$ whose first two moments match those of $w_t'$. The crucial idea in this step is to use a multivariate Berry-Esseen bound in Wasserstein distance [Bon20], which allows us to show that there exists a coupling between $w_t'$ and $w_t$ such that for $\kappa = \lfloor \alpha^{-1/2} \rfloor$,

$$\mathbb{E}\left\| \frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t - \frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t' \right\|_2^2 = W_2^2\left(\mathcal{L}\Big(\frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t\Big), \mathcal{L}\Big(\frac{1}{\sqrt{\kappa}} \sum_{t=1}^{\kappa} w_t'\Big)\right) \in \mathcal{O}\Big(\frac{1}{\kappa}\Big).$$

Under this noise coupling, we apply the generalized Moreau envelope on the difference sequence, $\{Y_{\kappa t}^{(\alpha)} - {Y'_{\kappa t}}^{(\alpha)}\}_{t\geq 0}$, to obtain that

$$\mathbb{E}\big[M_\eta(Y_{\kappa t+\kappa}^{(\alpha)} - {Y'_{\kappa t+\kappa}}^{(\alpha)})\big] \leq (1 - (1 - \sqrt{\gamma})\alpha\kappa)\, \mathbb{E}\big[M_\eta(Y_{\kappa t}^{(\alpha)} - {Y'_{\kappa t}}^{(\alpha)})\big] + \mathcal{O}(\alpha).$$

Here, the $\mathcal{O}(\alpha)$ term comes from the Berry-Esseen bound [Bon20]. It follows that for some constant $r_2$, we have

$$\lim_{t\to\infty} \mathbb{E}[M_\eta(Y_{\kappa t}^{(\alpha)} - {Y'_{\kappa t}}^{(\alpha)})] \in \mathcal{O}(\alpha^{\frac{1}{2}}).$$

Following the same line of arguments in Step 1, we conclude that $W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y'^{(\alpha)})\big) \in \mathcal{O}(\alpha^{\frac{1}{4}})$. Combining with the convergence rate result from Step 2 on $Y^{(\alpha)}$ with Gaussian noise, we obtain

$$\begin{aligned}W_2(\mathcal{L}\big(Y'^{(\alpha)}), \mathcal{L}(Y)\big) &\leq W_2\big(\mathcal{L}(Y'^{(\alpha)}), \mathcal{L}(Y^{(\alpha)})\big) + W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big)\\ &\leq \mathcal{O}(\alpha^{\frac{1}{4}}) + \mathcal{O}(\alpha^{\frac{r_1}{2}}) \in \mathcal{O}(\alpha^{\frac{\min(2r_1,1)}{4}}).\end{aligned}$$

This establishes that $Y'^{(\alpha)}$ with general noise converges in $W_2$ at a rate $\mathcal{O}(\alpha^{\frac{\min(2r_1,1)}{4}})$, which completes the proof of Theorem 2.

**Proof for Q-learning:** To prove Theorem 5, we need to couple the multiplicative noise for two sequences $Y_t^{(\alpha)}$ and $Y_t^{(\alpha')}$ in a similar manner as the additive noise case, with potentially mismatched stepsizes $(\alpha, \alpha')$ and time indices $(t, t')$. Importantly, in Step 3, in order to use the multivariate Berry-Esseen bound, we need to judiciously couple the general noisy sequence $\{(D'_t, P'_t, r'_t)\}$ with a carefully chosen Gaussian-distributed noisy sequence $\{(D_t, P_t, r_t)\}$ with matching joint covariance. Moreover, to obtain tight estimates for the squared distance of the form $\mathbb{E}\|Y_t^{(\alpha)} - Y_{t'}^{(\alpha')}\|_c^2$, we need to isolate the expected operator $\mathcal{H}$ from the noisy update (11). Doing so leads to more error terms that need to be carefully controlled.

## 5.3 Proof Outline for Theorem 3 (Bias Characterization)

Theorem 1 implies that the stochastic process $\{Y_t^{(\alpha)}\}_{t\geq 0}$ converges weakly in $W_2$ to a random variable $Y^{(\alpha)}$ corresponding to its stationary distribution. At stationarity we have the following equation in distribution:

$$Y^{(\alpha)} \overset{\mathrm{d}}{=} (1-\alpha)Y^{(\alpha)} + \sqrt{\alpha}\big(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*) + w\big). \tag{26}$$

Taking the expectation on both sides of the above equation yields

$$\mathbb{E}[Y^{(\alpha)}] = \frac{1}{\sqrt{\alpha}}\mathbb{E}[\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*)].$$

Recall that the operator $\mathcal{T}$ is $g \circ F$ decomposable in a local neighborhood $B^d(\theta^*, \epsilon)$ of $\theta^*$. We decompose the right-hand side of the above equation into two parts:

$$\begin{aligned}\mathbb{E}[Y^{(\alpha)}] =& \frac{1}{\sqrt{\alpha}}\mathbb{E}\big[(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))\big] && (T_1)\\ &+ \frac{1}{\sqrt{\alpha}}\mathbb{E}\big[(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon))\big]. && (T_2)\end{aligned}$$

For the term $T_1$, we make use of the contraction property of $\mathcal{T}$ and a concentration inequality to show that $\lim_{\alpha\to 0} T_1 = 0$. To analyze the term $T_2$, we consider two cases.

**Case 1:** If $g$ is smooth, then $\mathcal{T}$ is smooth on $B^d(\theta^*, \epsilon)$. By Taylor expansion of $\mathcal{T}$ and an argument similar to the proof of $\lim_{\alpha\to 0} T_1 = 0$, we have $\lim_{\alpha\to 0} T_2 = \nabla\mathcal{T}(\theta^*)\mathbb{E}[Y]$. Therefore, by letting $\alpha \to 0$, we obtain that

$$\mathbb{E}[Y] = \nabla\mathcal{T}(\theta^*)\mathbb{E}[Y].$$

By smoothness and contraction properties of $\mathcal{T}$, we can argue that $\mathbb{E}[Y] = 0$.

**Case 2:** If $g$ is nonsmooth, then by Taylor expansion of $F$ and continuity of $g$, we have

$$\mathbb{E}[Y] = \mathbb{E}[g(\nabla F(0)Y)].$$

We further consider two sub cases.

(a) If $\nabla F(0) = 0$, we have $\mathbb{E}[Y] = \mathbb{E}[g(0)] = 0$.

(b) If $\nabla F(0) \neq 0$, we define $h(Y) := g(\nabla F(0)Y)$. If the subdifferential of $h_1(\cdot)$ at 0 is not singleton, there exist $z_1, z_2 \in \mathbb{R}^d$ such that

$$h_1(Y) = h_1(Y) - h_1(0) \geq z_j^\top Y, \quad j = 1, 2.$$

Below we argue by contradiction that $\mathbb{E}[Y] \neq 0$.

Suppose that $\mathbb{E}[Y]=0$, in which case $\mathbb{E}[h(Y)]=0$. Therefore, we have $\mathbb{E}[h_1(Y)-z_j^\top Y]=0, \quad j=1,2$. Because $h_1(Y)-z_j^\top Y$ is always non-negative, we must have $h_1(Y)-z_j^T Y=0$ almost surely for $j=1,2$. Therefore, we have $z_1^\top Y=z_2^\top Y$ almost surely. Letting $\zeta=z_1-z_2$, we have $\zeta^\top Y=0$ almost surely, which implies $\mathbb{E}[(\zeta^T Y)^2]=0$.

By equation (26), we obtain

$$\begin{aligned}\mathbb{E}[(\zeta^T Y^{(\alpha)})^2] =&(1-\alpha)^2\mathbb{E}\big[(\zeta^T Y^{(\alpha)})^2\big]+2\sqrt{\alpha}(1-\alpha)\mathbb{E}\Big[\zeta^T Y^{(\alpha)}\cdot\zeta^T\big(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\mathcal{T}(\theta^*)\big)\Big]\\ &+\alpha\mathbb{E}\Big[\big(\zeta^T\big(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\mathcal{T}(\theta^*)+w\big)\big)^2\Big]\end{aligned}$$

Taking $\alpha\to 0$ to both sides of the above equation, we can finally obtain

$$\mathbb{E}[(\zeta^T Y)^2]\geq\mathbb{E}[(\zeta^T w)^2]=\zeta^T\operatorname{Var}(w)\zeta>0,$$

which contradicts with the equality $\mathbb{E}[(\zeta^T Y)^2]=0$ etablished above. We conclude that $\mathbb{E}[Y]\neq 0$.

**Proof for Q-learning:** In Theorem 6 we distinguish two types of MDP. When the MDP is Type A, the analysis is similar to Case 2(b) above. When MDP is Type B, the dynamic of Q-learning is locally linear around $\theta^*$. Therefore, the $T_2$ term above is almost proportional to $\mathbb{E}[Y^{(\alpha)}]$. For $T_1$, since the noise has finite $(2n)$-th moment, we can prove $T_1\in\mathcal{O}(\alpha^{n-\frac{1}{2}})$, which implies the desired bounds $\mathbb{E}[Y^{(\alpha)}]\in\mathcal{O}(\alpha^{n-\frac{1}{2}})$ and $\mathbb{E}[q^{(\alpha)}]=q^*+\mathcal{O}(\alpha^n)$.

# 6 Numerical Experiments

In this section, we provide numerical experiments for SA with additive noise and Q-learning.

For SA with additive noise, we consider the example in Section 1.1 with $b=0$. We run the update (7) initialized at $\theta_0^{(\alpha)}=1$, with stepsize $\alpha\in\{0.05,0.1,0.2,0.4\}$.[3] In Fig. 2(a), we plot the $\ell_1$ error $\|\theta-\theta^*\|_1$ for the tail-averaged (TA) iterates $\bar{\theta}_{0,k}^{(\alpha)}$, and the RR extrapolated iterates $\tilde{\theta}_{0,k}^{(\alpha)}$ with $\beta=\frac{1}{2}$. Theorems 2 and 3 show that the asymptotic bias of the TA iterates is $\Theta(\sqrt{\alpha})$, which can be reduced by RR extrapolation $o(\sqrt{\alpha})$. This bias reduction effect can be observed in Fig $2(a)$ by comparing the final errors for TA and RR iterates.

For Q-learning, we randomly generate an MDP with 3 states and 2 actions. The expected reward function $\bar{r}$ is sampled uniformly from $[0,1]^{|\mathcal{S}||\mathcal{A}|}$, and the rows of the transition kernel $P$ are sampled from Dirichlet($\mathbf{1}$), where $\mathbf{1}$ is the all-one vector. This random MDP is almost surely in Type B. We then generate Type A MDP by having the first two actions of the first state share the same transition and expected reward. The observed rewards are Gaussian: $r_t\sim\mathcal{N}(\bar{r},0.3I)$. We run Synchronous Q-learning initialized at $q_0^{(\alpha)}=\mathbf{1}$ with stepsize $\alpha\in\{0.02,0.04,0.08,0.16\}$. Theorem 5 and 6 show that for Type A MDP, the bias for TA is $\sqrt{\alpha}$ and can be order-wise reduced by RR extrapolation. This prediction is consistent with Fig $2(b)$. By Theorem 6 and the discussion in Section 4.3, for Type B MDP the bias is already small and of order $\mathcal{O}(\alpha^{\frac{3}{4}})$, for which RR extrapolation may not lead to obvious improvement. This is consistent with the result in Fig 2(c).

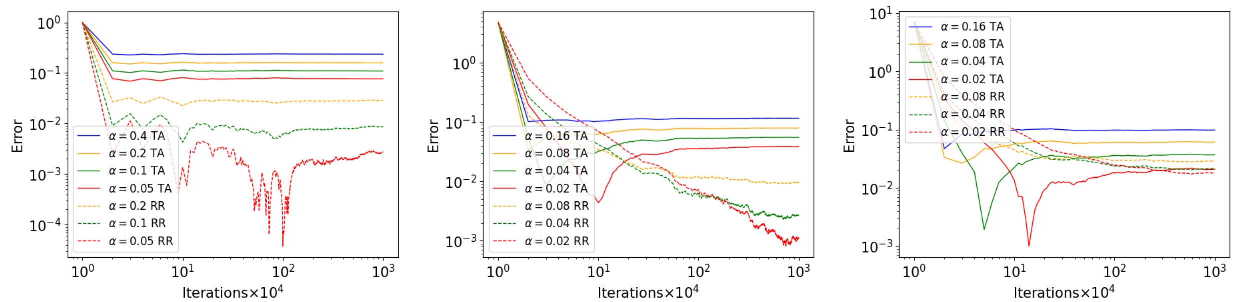
# 7 Additional Related Work

In this section, we discuss the existing results that are most relevant to our work.

## 7.1 Results on SA and SGD

The study of SA and SGD traces its origins to the seminal work by Robbins and Monro [RM51]. Classical works focus on diminishing stepsize regime [RM51, Blu54] and have established almost sure asymptotic convergence for SA and SGD algorithms. Subsequent works [Rup88, Pol90] propose the iterate averaging technique, now known as Polyak-Ruppert (PR) averaging, to accelerate convergence. The asymptotic convergence theory of SA and SGD is well-developed and extensively addressed in many exemplary textbooks [KY03, BMP12, WR22].

[3] The code can be found in https://colab.research.google.com/drive/1b2RVEhC5gMmtxgL7SOdekp-25UM2q2hV?usp=sharing.




(a) Example in Section 1.1 with $b = 0$.  (b) MDP in Type A.  (c) MDP in Type B.

Figure 2: The errors of tail-averaged (TA) and RR extrapolated iterates with different stepsizes $\alpha$. In the legends, $\alpha = x$ *RR* means RR extrapolation with two stepsizes $x$ and $2x$.

Some recent works [CZD+22, CBD22] study the non-asymptotic convergence with diminishing stepsize. The recent work [CMZ23] establishes the high probability bound on the estimation error of contractive SA with diminishing stepsize.

Recently, the study of constant stepsizes in SA and SGD has gained popularity. Many works in this line assume i.i.d. data. When using constant stepsize, one loses the almost sure convergence guarantee in the diminishing stepsize sequence regime, and at best can achieve distributional convergence, as shown in [DDB20, YBVE21, CMM22, HCX23b, ZX24]. Furthermore, a recurrent observation in the literature is the presence of asymptotic bias when using constant stepsize in SA, i.e., $\mathbb{E}[\theta^{(\alpha)}] \neq \theta^*$. When the SA update is locally smooth, the asymptotic bias has been demonstrated to be of $\Theta(\alpha)$ order in [DDB20, HCX23b, ZX24]. The work [YBVE21] considers nonsmooth SA but only provides an upper bound for the asymptotic bias, i.e. $|\mathbb{E}[\theta^{(\alpha)}] - \theta^*| \leq c\sqrt{\alpha}$. Many papers provide non-asymptotic MSE upper bounds. The work in [LS18, MLW+20] studies linear SA under i.i.d. data and provides an upper bound on the MSE. There are also works that analyze the MSE with Markovian data, such as [SY19, MPWB21, DMN+21, DMNS22]. The work in [CMSS23, CMSS20] introduce the generalized Moreau envelope (GME) to analyze the MSE of general contractive SA. In our work, we make use of the GME, but we extend this technique to analyze different and more general problems, specifically, generalizing to obtain upper bounds for any $2n$-th moment, proving weak convergence of SA iterates and proving steady-state convergence as stepsize $\alpha$ diminishes to 0.

## 7.2 Applications in Reinforcement Learning

Many widely employed iterative algorithms in reinforcement learning (RL) can be reformulated as SA problems [SB18, Ber19]. Among those, the two most well-known algorithms are the temporal-difference (TD) learning for policy evaluation [Sut88, DS94] and Q-learning for optimal policy learning[WD92]. The TD algorithms when incorporating linear function approximation can be cast into the framework of linear SA. Q-learning is a nonsmooth and nonlinear contractive SA, and has also been studied extensively in both classical works [Tsi94, Sze97, EDMB03] and recent works [CZD+22, CMSS23]. The work [ZX24] studies the stationary distribution of asynchronous Q-learning with Markovian data and characterizes the asymptotic bias under the assumption that MDP has no tied state.

## 7.3 Nonsmooth Function Class

Nonsmooth functions have been studied in many works, such as semi-smoothness in [Mif77], identifiable surfaces in [Wri93], $\mathcal{UV}$-structures in [LOS00, MS05], partly smoothness in [Lew02], $g \circ F$ decomposition in [Sha03, Sag13] and minimal identifiable sets in [DL14, DDJ23]. In our work, we adopt the definition in [Sha03] and extend it to a multidimensional function space to define the nonsmooth SA.

### 7.4 Results on Steady-State Convergence

The steady-state convergence is commonly studied in the realm of stochastic processes, with one well-known application being the steady-state convergence in queueing networks. As discussed, the classical method is through justifying the interchange of limits, as seen in [GZ06, Gur14, YY16, YY18]. An alternative approach is through the basic adjoint relationship (BAR) approach, which studies the generator of the Markov process, i.e., $\mathbb{E}[Gf(Y^{(\alpha)})] = 0$ as $\alpha \to 0$ [BDM17, BDM24, CMM22]. Another line of work related to steady-state convergence focuses on the unadjusted Langevin algorithm (ULA) [DM17, DM19]. These works take an approach similar to the justification of limit interchange in queueing networks, in which they first demonstrate the convergence of ULA to the corresponding stochastic differential equation (SDE), and then relate the convergence to the stationary distribution of the SDE.

## 8 Conclusion

In this work, we studied nonsmooth contractive SA with a constant stepsize. We developed prelimit coupling techniques for establishing steady-state convergence and characterizing the asymptotic bias, highlighting the impact of nonsmoothness on steady-state behavior. Our coupling techniques also bear potential for other nonsmooth dynamical systems such as piecewise smooth diffusion, stochastic differential equations and their discretization. Of immediate interest are to obtain more refined characterization of the steady-state distribution $Y^{(\alpha)}$ and its limit $Y$, such as higher moment results and other functionals of the distribution and obtain non-asymptotic results as a function of $\alpha$ and the level of nonsmoothness. Generalizing our results to general noise settings is another interesting future direction. For additive martingale difference noise, we believe the current analysis can be combined with an appropriate martingale Berry-Esseen Central Limit Theorem to establish similar distributional and steady-state convergence results. For more general multiplicative Markovian noise, however, establishing such results would require a better understanding of Markovian nonlinear SA and new coupling arguments.

## Acknowledgement

Q. Xie and Y. Zhang are supported in part by National Science Foundation (NSF) grants CNS-1955997 and EPCN-2339794. Y. Chen is supported in part by NSF grants CCF-1704828 and CCF-2233152. Y. Zhang is also supported in part by NSF Award DMS-2023239.

## References

[BDM17] Anton Braverman, J. G. Dai, and Masakiyo Miyazawa. Heavy traffic approximation for the stationary distribution of a generalized Jackson network: the BAR approach. *Stochastic Systems*, 7(1):143–196, May 2017.

[BDM24] Anton Braverman, J. G. Dai, and Masakiyo Miyazawa. The BAR approach for multiclass queueing networks with SBP service policies, 2024.

[Ber19] Dimitri P. Bertsekas. *Reinforcement learning and Optimal Control*. Athena Scientific, Belmont, Massachusetts, USA, 2019.

[Blu54] Julius R. Blum. Approximation Methods which Converge with Probability one. *The Annals of Mathematical Statistics*, 25(2):382 – 386, 1954.

[BMP12] Albert Benveniste, Michel Métivier, and Pierre Priouret. *Adaptive algorithms and stochastic approximations*, volume 22. Springer Science & Business Media, 2012.

[Bon20] Thomas Bonis. Stein's method for normal approximation in wasserstein distances with application to the multivariate central limit theorem. *Probability Theory and Related Fields*, 178(3-4):827–860, 2020.

[Bra98] Maury Bramson. State space collapse with application to heavy traffic limits for multiclass queueing networks. *QUESTA*, 30:89–140, 1998.

[CBD22] Siddharth Chandak, Vivek S. Borkar, and Parth Dodhia. Concentration of contractive stochastic approximation and reinforcement learning. *Stochastic Systems*, 12(4):411–430, 2022.

[CMM22] Zaiwei Chen, Shancong Mou, and Siva Theja Maguluri. Stationary behavior of constant stepsize sgd type algorithms: An asymptotic characterization. *Proc. ACM Meas. Anal. Comput. Syst.*, 6(1), feb 2022.

[CMSS20] Zaiwei Chen, Siva Theja Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. Finite-sample analysis of contractive stochastic approximation using smooth convex envelopes. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 8223–8234. Curran Associates, Inc., 2020.

[CMSS23] Zaiwei Chen, Siva T Maguluri, Sanjay Shakkottai, and Karthikeyan Shanmugam. A lyapunov theory for finite-sample guarantees of markovian stochastic approximation. *Operations Research*, 2023.

[CMZ23] Zaiwei Chen, Siva Theja Maguluri, and Martin Zubeldia. Concentration of contractive stochastic approximation: Additive and multiplicative noise. *arXiv preprint arXiv:2303.15740*, 2023.

[CZD+22] Zaiwei Chen, Sheng Zhang, Thinh T. Doan, John-Paul Clarke, and Siva Theja Maguluri. Finite-sample analysis of nonlinear stochastic approximation with applications in reinforcement learning. *Automatica*, 146:110623, 2022.

[DDB20] Aymeric Dieuleveut, Alain Durmus, and Francis Bach. Bridging the gap between constant step size stochastic gradient descent and Markov chains. *The Annals of Statistics*, 48(3):1348 – 1382, 2020.

[DDJ23] Damek Davis, Dmitriy Drusvyatskiy, and Liwei Jiang. Asymptotic normality and optimality in nonsmooth stochastic approximation, 2023.

[DJMS21] Alain Durmus, Pablo Jiménez, Eric Moulines, and Salem Said. On Riemannian stochastic approximation schemes with fixed step-size. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1018–1026. PMLR, 13–15 Apr 2021.

[DL14] Dmitriy Drusvyatskiy and Adrian S Lewis. Optimality, identifiability, and sensitivity. *Mathematical Programming*, 147(1-2):467–498, 2014.

[DM17] Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551 – 1587, 2017.

[DM19] Alain Durmus and Éric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854 – 2882, 2019.

[DMN+21] Alain Durmus, Eric Moulines, Alexey Naumov, Sergey Samsonov, Kevin Scaman, and Hoi-To Wai. Tight high probability bounds for linear stochastic approximation with fixed stepsize. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 30063–30074. Curran Associates, Inc., 2021.

[DMNS22] Alain Durmus, Eric Moulines, Alexey Naumov, and Sergey Samsonov. Finite-time high-probability bounds for Polyak-Ruppert averaged iterates of linear stochastic approximation, 2022.

[DS94] Peter Dayan and Terrence J. Sejnowski. TD($\lambda$) converges with probability 1. *Machine Learning*, 14(3):295–301, Mar 1994.

[Dur19] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, England, 5 edition, April 2019.

[EDMB03] Eyal Even-Dar, Yishay Mansour, and Peter Bartlett. Learning rates for Q-Learning. *Journal of Machine Learning Research*, 5(1):1–25, Dec 2003.

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, Nashville, TN, 2 edition, March 1999.

[Gur14] Itai Gurvich. Validity of heavy-traffic steady-state approximations in multiclass queueing networks: the case of queue-ratio disciplines. *Mathematics of Operations Research*, 39(1):121–162, 2014.

[GZ06] David Gamarnik and Assaf Zeevi. Validity of heavy traffic steady-state approximation in generalized Jackson networks. *Ann. Appl. Probab.*, 16(1):56–90, 2006.

[HCX23a] Dongyan Huo, Yudong Chen, and Qiaomin Xie. Effectiveness of constant stepsize in markovian lsa and statistical inference. *arXiv preprint arXiv:2312.10894*, 2023.

[HCX23b] Dongyan (Lucy) Huo, Yudong Chen, and Qiaomin Xie. Bias and extrapolation in markovian linear stochastic approximation with constant stepsizes. In *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 81–82, 2023.

[Hil87] F B Hildebrand. *Introduction to numerical analysis*. Dover Books on Mathematics. Dover Publications, Mineola, NY, 2 edition, June 1987.

[HJ12] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, England, 2 edition, October 2012.

[JKK+18] Prateek Jain, Sham M. Kakade, Rahul Kidambi, Praneeth Netrapalli, and Aaron Sidford. Parallelizing stochastic gradient descent for least squares regression: Mini-batching, averaging, and model misspecification. *Journal of Machine Learning Research*, 18(223):1–42, 2018.

[KY03] Harold J. Kushner and G. George Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Stochastic Modelling and Applied Probability. Springer, New York, NY, USA, 2nd edition, 2003.

[Lan20] Guanghui Lan. *First-order and stochastic optimization methods for machine learning*, volume 1. Springer, 2020.

[Lew02] Adrian S Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.

[LOS00] Claude Lemaréchal, François Oustry, and Claudia Sagastizábal. The $\mathcal{U}$-lagrangian of a convex function. *Transactions of the American mathematical Society*, 352(2):711–729, 2000.

[LS18] Chandrashekar Lakshminarayanan and Csaba Szepesvári. Linear stochastic approximation: How far does constant step-size and iterate averaging go? In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1347–1355. PMLR, 09–11 Apr 2018.

[MB11] Eric Moulines and Francis Bach. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

[Mif77] Robert Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization*, 15(6):959–972, 1977.

[MLW+20] Wenlong Mou, Chris Junchi Li, Martin J. Wainwright, Peter L. Bartlett, and Michael I. Jordan. On linear stochastic approximation: Fine-grained Polyak-Ruppert and non-asymptotic concentration. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2947–2997. PMLR, 09–12 Jul 2020.

[MPWB21] Wenlong Mou, Ashwin Pananjady, Martin J. Wainwright, and Peter L. Bartlett. Optimal and instance-dependent guarantees for markovian linear stochastic approximation, 2021.

[MS05] Robert Mifflin and Claudia Sagastizábal. A-algorithm for convex minimization. *Mathematical programming*, 104:583–608, 2005.

[PJ92] Boris T. Polyak and Anatoli B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838–855, 1992.

[Pol90] Boris T. Polyak. New stochastic approximation type procedures. *Automation and Remote Control*, 51(7):98–107, Jul 1990.

[RM51] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407, 1951.

[Rup88] David Ruppert. Efficient estimations from a slowly convergent robbins-monro process. Technical report, Cornell University Operations Research and Industrial Engineering, 1988.

[Sag13] Claudia Sagastizábal. Composite proximal bundle method. *Mathematical Programming*, 140(1):189–233, 2013.

[SB18] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018.

[Sha03] Alexander Shapiro. On a class of nonsmooth composite functions. *Mathematics of Operations Research*, 28(4):677–692, 2003.

[Sut88] Richard S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1):9–44, Aug 1988.

[SY19] Rayadurgam Srikant and Lei Ying. Finite-time error bounds for linear stochastic approximation and TD learning. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2803–2830. PMLR, 25–28 Jun 2019.

[Sze97] Csaba Szepesvári. The asymptotic convergence-rate of Q-Learning. In M. Jordan, M. Kearns, and S. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press, 1997.

[Tsi94] John N. Tsitsiklis. Asynchronous stochastic approximation and Q-Learning. *Machine Learning*, 16(3):185–202, Sep 1994.

[Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

[Wai19] Martin J Wainwright. Stochastic approximation with cone-contractive operators: Sharp $\ell_\infty$-bounds for $q$-learning. *arXiv preprint arXiv:1905.06265*, 2019.

[WD92] Christopher J. C. H. Watkins and Peter Dayan. Q-Learning. *Machine Learning*, 8(3):279–292, May 1992.

[WR22] Stephen J. Wright and Benjamin Recht. *Optimization for Data Analysis*. Cambridge University Press, 2022.

[Wri93] Stephen J Wright. Identifiable surfaces in constrained optimization. *SIAM Journal on Control and Optimization*, 31(4):1063–1079, 1993.

[YBVE21] Lu Yu, Krishnakumar Balasubramanian, Stanislav Volgushev, and Murat A Erdogdu. An analysis of constant step size SGD in the non-convex regime: Asymptotic normality and bias. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 4234–4248. Curran Associates, Inc., 2021.

[YY16] Heng-Qing Ye and David D. Yao. Diffusion limit of fair resource control—stationarity and interchange of limits. *Mathematics of Operations Research*, 41(4):1161–1207, 2016.

[YY18] Heng-Qing Ye and David D. Yao. Justifying diffusion approximations for multiclass queueing networks under a moment condition. *The Annals of Applied Probability*, 28(6):3652 – 3697, 2018.

[ZX24] Yixuan Zhang and Qiaomin Xie. Constant stepsize q-learning: Distributional convergence, bias and extrapolation, 2024.

# A Proof of Proposition 1

Proposition 1 follows from combining the following lemma and the property (3) in Proposition 5.

**Lemma 1.** *For each integer $n \geq 1$, under Assumption 1 and Assumption 2($\boldsymbol{n}$), there exists $\eta$, $\bar{\alpha}$ such that for any $\alpha \leq \bar{\alpha}$, there exist $t_{\alpha,n}$ such that*

$$\mathbb{E}[M_\eta^n(\theta_t - \theta^*)] \leq \mathbb{E}[M_\eta^n(\theta_{t_\alpha} - \theta^*)](1 - \alpha(1 - \sqrt{\gamma}))^{t - t_{\alpha,n}} + c_n \alpha^n$$

*holds for all $t \geq t_{\alpha,n}$, where $M_\eta(\cdot)$ is defined in (21) and $\{c_n\}_{n\geq 0}$ are universal constants that are independent with $\alpha$ and $t$. Moreover, $t_{\alpha,1} = 0$.*

## A.1 Proof of Lemma 1

We use induction on $n$ to prove Lemma 1

**Base Case:** $n = 1$.

By subtracting $\theta^*$ from both side of equation (7), we obtain

$$\theta_{t+1} - \theta^* = \theta_t - \theta^* + \alpha(\mathcal{T}(\theta_t) - \theta_t + w_t) = (1 - \alpha)(\theta_t - \theta^*) + \alpha(\mathcal{T}(\theta_t) - \mathcal{T}(\theta^*) + w_t), \tag{27}$$

where the second equality holds because $\mathcal{T}(\theta^*) = \theta^*$.

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of equation (27) and by property (1) in Proposition 5, we obtain

$$\begin{aligned}
M_\eta(\theta_{t+1} - \theta^*) \leq &(1 - \alpha)^2 M_\eta(\theta_t - \theta^*) \\
&+ (1 - \alpha)\alpha\langle \nabla M_\eta(\theta_t - \theta^*), \mathcal{T}(\theta_t) - \mathcal{T}(\theta^*) + w_t\rangle \qquad (28)\\
&+ \frac{\alpha^2}{2\eta}\|\mathcal{T}(\theta_t) - \mathcal{T}(\theta^*) + w_t\|_2^2. \qquad (29)
\end{aligned}$$

The term in (28) can be bounded as follows:

$$\begin{aligned}
(28) &= (1 - \alpha)\alpha\left(\langle \nabla M_\eta(\theta_t - \theta^*), \mathcal{T}(\theta_t) - \mathcal{T}(\theta^*)\rangle + \langle \nabla M_\eta(\theta_t - \theta^*), w_t\rangle\right) \\
&\overset{\text{(i)}}{\leq} (1 - \alpha)\alpha\left(\|\theta_t - \theta^*\|_m \|\mathcal{T}(\theta_t) - \mathcal{T}(\theta^*)\|_m + \langle \nabla M_\eta(\theta_t - \theta^*), w_t\rangle\right) \\
&\overset{\text{(ii)}}{\leq} \frac{(1 - \alpha)\alpha\gamma}{l_{cm}}\|\theta_t - \theta^*\|_m \|\theta_t - \theta^*\|_c + (1 - \alpha)\alpha\langle \nabla M_\eta(\theta_t - \theta^*), w_t\rangle \\
&\overset{\text{(iii)}}{\leq} \frac{2\alpha(1 - \alpha)\gamma u_{cm}}{l_{cm}} M_\eta(\theta_t - \theta^*) + (1 - \alpha)\alpha\langle \nabla M_\eta(\theta_t - \theta^*), w_t\rangle,
\end{aligned}$$

where (i) holds because of property (4) of Proposition 5, (ii) holds because of property (3) of Proposition 5 and $\gamma$-contraction of $\mathcal{T}(\cdot)$, and (iii) holds because of property (2) of Proposition 5.

The term in (29) can be bounded as follows:

$$\begin{aligned}(29) \leq & \frac{\alpha^2}{2\eta l_{cs}^2}\|\mathcal{T}(\theta_t)-\mathcal{T}(\theta^*)+w_t\|_c^2 \leq \frac{\alpha^2}{\eta l_{cs}^2}\left(\|\mathcal{T}(\theta_t)-\mathcal{T}(\theta^*)\|_c^2+\|w_t\|_c^2\right)\\ \leq & \frac{2\alpha^2\gamma^2u_{cm}^2}{\eta l_{cs}^2}M_\eta(\theta_t-\theta^*)+\frac{\alpha^2\|w_t\|_c^2}{\eta l_{cs}^2}.\end{aligned}$$

Combining the above bounds, we obtain

$$\begin{aligned}M_\eta(\theta_{t+1}-\theta^*) \leq & \Big(1-2\alpha(1-\frac{(1-\alpha)\gamma u_{cm}}{l_{cm}})+\alpha^2(1+\frac{2\gamma u_{cm}^2}{\eta l_{cs}^2})\Big)M_\eta(\theta_t-\theta^*)\\ & +(1-\alpha)\alpha\langle\nabla M_\eta(\theta_t-\theta^*),w_t\rangle+\frac{\alpha^2\|w_t\|_c^2}{\eta l_{cs}^2}.\end{aligned}$$

Recall that $\frac{u_{cm}}{l_{cm}}=\sqrt{\frac{1+\eta u_{cs}^2}{1+\eta l_{cs}^2}}$ by property (3) in Proposition 5. We can always choose a sufficient small $\eta>0$ such that $\frac{u_{cm}}{l_{cm}}\leq\frac{1}{\sqrt{\gamma}}$, which implies $-2\alpha(1-\frac{(1-\alpha)\gamma u_{cm}}{l_{cm}})\leq-2\alpha(1-(1-\alpha)\sqrt{\gamma})\leq-2\alpha(1-\sqrt{\gamma})$. Furthermore, there always exists $\bar{\alpha}>0$ such that $\bar{\alpha}<1$ and $\Big(1-2\alpha(1-\sqrt{\gamma})+\alpha^2(1+\frac{2\gamma^2u_{cm}^2}{\eta l_{cs}^2})\Big)\leq 1-\alpha(1-\sqrt{\gamma})<1$ when $\alpha\leq\bar{\alpha}$. Therefore, for $\forall\alpha\leq\bar{\alpha}$ and $t\geq 0$, we obtain

$$M_\eta(\theta_{t+1}-\theta^*)\leq(1-\alpha(1-\sqrt{\gamma}))\,M_\eta(\theta_t-\theta^*)+(1-\alpha)\alpha\langle\nabla M_\eta(\theta_t-\theta^*),w_t\rangle+\frac{\alpha^2\|w_t\|_c^2}{\eta l_{cs}^2}. \tag{30}$$

Taking expectation on both sides of equation (30), we obtain

$$\begin{aligned}\mathbb{E}[M_\eta(\theta_{t+1}-\theta^*)] &\leq (1-\alpha(1-\sqrt{\gamma}))\,\mathbb{E}[M_\eta(\theta_t-\theta^*)]+\frac{\alpha^2c_w}{\eta l_{cs}^2}\\ &\leq (1-\alpha(1-\sqrt{\gamma}))^{t+1}\,\mathbb{E}[M_\eta(\theta_0-\theta^*)]+\sum_{k=0}^{t}(1-\alpha(1-\sqrt{\gamma}))^k\,\frac{\alpha^2c_w}{\eta l_{cs}^2}\\ &\leq (1-\alpha(1-\sqrt{\gamma}))^{t+1}\,\mathbb{E}[M_\eta(\theta_0-\theta^*)]+\frac{\alpha c_w}{\eta l_{cs}^2(1-\sqrt{\gamma})},\end{aligned}$$

where $c_w=\mathbb{E}[\|w_t\|_c^2]$ and the first inequality holds because $w_t$ is zero mean and independent with $\theta_t-\theta^*$. Therefore, we complete the proof for the base case.

**Induction Step:** Given positive integer $k\geq 2$, assume Lemma 1 holds for all $n\leq k-1$. When $n=k$, take $k$-th moment to both side of equation (30) and we obtain

$$\begin{aligned}\mathbb{E}[M_\eta^k(\theta_{t+1}-\theta^*)] &\leq \mathbb{E}\Big[\Big(\underbrace{(1-\alpha(1-\sqrt{\gamma}))\,M_\eta(\theta_t-\theta^*)}_{T_1}+\underbrace{(1-\alpha)\alpha\langle\nabla M_\eta(\theta_t-\theta^*),w_t\rangle}_{T_2}+\underbrace{\frac{\alpha^2\|w_t\|_c^2}{\eta l_{cs}^2}}_{T_3}\Big)^k\Big]\\ &= \underbrace{\mathbb{E}\Big[\sum_{a+b=k}\tbinom{k}{a}\tbinom{k-a}{b}T_1^aT_2^b\Big]}_{S_1}+\underbrace{\mathbb{E}\Big[\sum_{a+b+c=k,c\geq 1}\tbinom{k}{a}\tbinom{k-a}{b}T_1^aT_2^bT_3^c\Big]}_{S_2}.\end{aligned} \tag{31}$$

We next analyze $S_1$ and $S_2$. For $S_1$ we have

$$\begin{aligned}
S_1 &= (1-\alpha(1-\sqrt{\gamma}))^k\,\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] + \mathbb{E}\Big[\sum_{a+b=k,b\ge 2}\tbinom{k}{a}\tbinom{k-a}{b}T_1^aT_2^b\Big] \\
&\le (1-\alpha(1-\sqrt{\gamma}))^k\,\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] + \mathbb{E}\Big[\sum_{a+b=k,b\ge 2,b \text{ is even}}\tbinom{k}{a}\tbinom{k-a}{b}\alpha^b M_\eta^a(\theta_t-\theta^*)\|\theta_t-\theta^*\|_m^b\|w_t\|_m^b\Big] \\
&\quad + \sum_{a+b=k,b\ge 3,b \text{ is odd}}\alpha^b\mathbb{E}\big[2^{\frac{b}{2}}\tbinom{k}{a}\tbinom{k-a}{b}\|w_t\|_m^b\big]\mathbb{E}\big[M_\eta^{a+\frac{b}{2}}(\theta_t-\theta^*)\big] \\
&\le (1-\alpha(1-\sqrt{\gamma}))^k\,\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] + \sum_{a+b=k,b\ge 2,b \text{ is even}}\alpha^b\underbrace{\mathbb{E}\big[2^{\frac{b}{2}}\tbinom{k}{a}\tbinom{k-a}{b}\|w_t\|_m^b\big]}_{\text{constant depends on } k}\underbrace{\mathbb{E}\big[M_\eta^{a+\frac{b}{2}}(\theta_t-\theta^*)\big]}_{\in\mathcal{O}(\alpha^{a+\frac{b}{2}}),\because a+\frac{b}{2}\le k-1} \\
&\quad + \sum_{a+b=k,b\ge 3,b \text{ is odd}}\alpha^b\underbrace{\mathbb{E}\big[2^{\frac{b}{2}}\tbinom{k}{a}\tbinom{k-a}{b}\|w_t\|_m^b\big]}_{\text{constant depends on } k}\sqrt{\underbrace{\mathbb{E}\big[M_\eta^{a+\frac{b+1}{2}}(\theta_t-\theta^*)\big]}_{\in\mathcal{O}(\alpha^{a+\frac{b+1}{2}}),\because a+\frac{b+1}{2}\le k-1}\underbrace{\mathbb{E}\big[M_\eta^{a+\frac{b-1}{2}}(\theta_t-\theta^*)\big]}_{\in\mathcal{O}(\alpha^{a+\frac{b-1}{2}})}} \\
&\overset{\text{(i)}}{\le} (1-\alpha(1-\sqrt{\gamma}))^k\,\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] + \mathcal{O}(\alpha^{k+1}) \le (1-\alpha(1-\sqrt{\gamma}))\,\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] + \mathcal{O}(\alpha^{k+1}),
\end{aligned}$$

where (i) holds by induction hypothesis and taking $t$ to be sufficiently large. For $S_2$ we have

$$\begin{aligned}
S_2 &\le \sum_{a+b+c=k,c\ge 1}\alpha^{b+2c}\mathbb{E}\Big[\tbinom{k}{a}\tbinom{k-a}{b}\frac{\|w_t\|_c^{2c}\|w_t\|_m^b}{\eta^c l_{cs}^{2c}}\Big]\mathbb{E}\big[M_\eta^{a+\frac{b}{2}}(\theta_t-\theta^*)\big] \\
&\le \sum_{a+b+c=k,c\ge 1,b \text{ is even}}\alpha^{b+2c}\underbrace{\mathbb{E}\Big[\tbinom{k}{a}\tbinom{k-a}{b}\frac{\|w_t\|_c^{2c}\|w_t\|_m^b}{\eta^c l_{cs}^{2c}}\Big]}_{\text{constant depends on } k}\underbrace{\mathbb{E}\big[M_\eta^{a+\frac{b}{2}}(\theta_t-\theta^*)\big]}_{\in\mathcal{O}(\alpha^{a+\frac{b}{2}}),\because a+\frac{b}{2}\le k-1} \\
&\quad + \sum_{a+b+c=k,c\ge 1,b \text{ is odd}}\alpha^{b+2c}\underbrace{\mathbb{E}\Big[\tbinom{k}{a}\tbinom{k-a}{b}\frac{\|w_t\|_c^{2c}\|w_t\|_m^b}{\eta^c l_{cs}^{2c}}\Big]}_{\text{constant depends on } k}\sqrt{\underbrace{\mathbb{E}\big[M_\eta^{a+\frac{b+1}{2}}(\theta_t-\theta^*)\big]}_{\in\mathcal{O}(\alpha^{a+\frac{b+1}{2}}),\because a+\frac{b+1}{2}\le k-1}\underbrace{\mathbb{E}\big[M_\eta^{a+\frac{b-1}{2}}(\theta_t-\theta^*)\big]}_{\in\mathcal{O}(\alpha^{a+\frac{b-1}{2}})}}.
\end{aligned}$$

By induction hypothesis and taking $t$ to be sufficiently large, we conclude that $S_2 \in \mathcal{O}(\alpha^{k+1})$.

Combining the bound of $S_1, S_2$ with equation (31), we obtain

$$\mathbb{E}[M_\eta^k(\theta_{t+1}-\theta^*)] \le (1-\alpha(1-\sqrt{\gamma}))\,\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] + \mathcal{O}(\alpha^{k+1}).$$

Therefore, for $\forall\alpha \le \bar{\alpha}$, there exist $t_{\alpha,k} > 0$ such that

$$\mathbb{E}[M_\eta^k(\theta_t-\theta^*)] \le \mathbb{E}[M_\eta^k(\theta_{t_{\alpha,k}}-\theta^*)](1-\alpha(1-\sqrt{\gamma}))^{t-t_{\alpha,k}} + c_k\alpha^k$$

holds for $\forall t \ge t_{\alpha,k}$, where $c_k$ are universal constants that are independent with $\alpha$ and $t$.

# B Proof of Theorem 1

We prove the three properties stated in Theorem 1 in the next three subsections, respectively.

## B.1 Unique Limit Distribution

We consider a pair of coupled , $\{\theta_t^{[1]}\}_{t\ge 0}$ and $\{\theta_t^{[2]}\}_{t\ge 0}$, defined as

$$\begin{aligned}
\theta_{t+1}^{[1]} &= \theta_t^{[1]} + \alpha\left(\mathcal{T}(\theta_t^{[1]}) - \theta_t^{[1]} + w_t\right), \\
\theta_{t+1}^{[2]} &= \theta_t^{[2]} + \alpha\left(\mathcal{T}(\theta_t^{[2]}) - \theta_t^{[2]} + w_t\right).
\end{aligned} \tag{32}$$

Here $\{\theta_t^{[1]}\}_{t\geq 0}$ and $\{\theta_t^{[2]}\}_{t\geq 0}$ are two iterates coupled by sharing $\{w_t\}_{t\geq 0}$. We assume that the initial iterates $\theta_0^{[1]}$ and $\theta_0^{[2]}$ may depend on each other.

Taking the difference of the two equations in (32), we obtain

$$\theta_{t+1}^{[1]} - \theta_{t+1}^{[2]} = (1-\alpha)(\theta_t^{[1]} - \theta_t^{[2]}) + \alpha\left(\mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\right).$$

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both side of above equation and by property (1) in Proposition 5, we obtain

$$\begin{aligned} M_\eta(\theta_{t+1}^{[1]} - \theta_{t+1}^{[2]}) \leq &(1-\alpha)^2 M_\eta(\theta_t^{[1]} - \theta_t^{[2]}) + \alpha(1-\alpha)\langle \nabla M_\eta(\theta_t^{[1]} - \theta_t^{[2]}), \mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\rangle \\ &+ \frac{\alpha^2}{2\eta}\|\mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\|_2^2. \end{aligned}$$

Taking expectation to both side of above equation, we obtain

$$\begin{aligned} \mathbb{E}[M_\eta(\theta_{t+1}^{[1]} - \theta_{t+1}^{[2]})] \leq &(1-\alpha)^2\mathbb{E}[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})] \\ &+ \underbrace{\alpha(1-\alpha)\mathbb{E}[\langle \nabla M_\eta(\theta_t^{[1]} - \theta_t^{[2]}), \mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\rangle]}_{T_1} + \underbrace{\frac{\alpha^2}{2\eta}\mathbb{E}\|\mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\|_2^2}_{T_2}. \end{aligned}$$

When $\alpha \leq 1$, we obtain

$$\begin{aligned} T_1 &\overset{\text{(i)}}{\leq} \alpha(1-\alpha)\mathbb{E}[\|\theta_t^{[1]} - \theta_t^{[2]}\|_m \|\mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\|_m] \\ &\overset{\text{(ii)}}{\leq} \frac{\alpha(1-\alpha)}{l_{cm}}\mathbb{E}[\|\theta_t^{[1]} - \theta_t^{[2]}\|_m \|\mathcal{T}(\theta_t^{[1]}) - \mathcal{T}(\theta_t^{[2]})\|_c] \\ &\leq \frac{\alpha(1-\alpha)\gamma}{l_{cm}}\mathbb{E}[\|\theta_t^{[1]} - \theta_t^{[2]}\|_m \|\theta_t^{[1]} - \theta_t^{[2]}\|_c] \\ &\overset{\text{(iii)}}{\leq} \frac{2\alpha(1-\alpha)\gamma u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})] \overset{\text{(iv)}}{\leq} 2\alpha\sqrt{\gamma}\mathbb{E}[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})], \end{aligned}$$

where (i) holds because of the property (4) of Proposition 5, (ii) and (iii) holds because of the property (2) and (3) of Proposition 5 and (iv) holds because $\frac{u_{cm}}{l_{cm}} = \sqrt{\frac{1+\eta u_{cs}^2}{1+\eta l_{cs}^2}}$ by property (3) in Proposition 5 and we can always choose a sufficient small $\eta > 0$ such that $\frac{u_{cm}}{l_{cm}} \leq \frac{1}{\sqrt{\gamma}}$.

By $\gamma$-contraction of $\mathcal{T}(\cdot)$, we obtain

$$T_2 \leq \frac{\alpha^2\gamma^2}{2\eta l_{cs}^2}\mathbb{E}\|\theta_t^{[1]} - \theta_t^{[2]}\|_c^2 \leq \frac{\alpha^2\gamma^2 u_{cm}^2}{\eta l_{cs}^2}\mathbb{E}[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})].$$

Combining the bound for $T_1$ and $T_2$, there exists $\bar{\alpha}' \leq \bar{\alpha}$ such that

$$\begin{aligned} \mathbb{E}[M_\eta(\theta_{t+1}^{[1]} - \theta_{t+1}^{[2]})] &\leq (1 - 2\alpha(1-\sqrt{\gamma}) + \mathcal{O}(\alpha^2))\mathbb{E}[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})] \\ &\leq (1 - \alpha(1-\sqrt{\gamma}))\mathbb{E}[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})], \end{aligned}$$

for $\forall \alpha \leq \bar{\alpha}'$. Therefore, we have

$$\begin{aligned} W_2^2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\theta_t^{[2]}\right)\right) &\leq \mathbb{E}\left[\left\|\theta_t^{[1]} - \theta_t^{[2]}\right\|_c^2\right] \\ &\leq 2u_{cm}^2\mathbb{E}\left[M_\eta(\theta_t^{[1]} - \theta_t^{[2]})\right] \leq 2u_{cm}^2\mathbb{E}\left[M_\eta(\theta_0^{[1]} - \theta_0^{[2]})\right](1-\alpha(1-\sqrt{\gamma}))^t, \end{aligned} \tag{33}$$

which implies $W_2^2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\theta_t^{[2]}\right)\right)$ decays geometrically. Note that equation (33) always holds for any joint distribution of initial iterates $(\theta_0^{[1]}, \theta_0^{[2]})$. Then, we use $\theta_{-1}^{[2]}$ to denote a random variable that satisfies $\theta_{-1}^{[2]} \overset{\text{d}}{=} \theta_0^{[1]}$ where $\overset{\text{d}}{=}$ denotes equality in distribution and $\theta_{-1}^{[2]}$ is independent of $\{x_t\}_{t\geq 0}$. Finally, we set $\theta_0^{[2]}$ as

$$\theta_0^{[2]} = (1-\alpha)\theta_{-1}^{[2]} + \alpha\left(\mathcal{T}(\theta_{-1}^{[2]}) + w_{-1}\right). \tag{34}$$

Given that $\theta_{-1}^{[2]} \overset{\mathrm{d}}{=} \theta_0^{[1]}$ and $\theta_{-1}^{[2]}$ is independent with $\{w_t\}_{t\geq -1}$, we can prove $\theta_t^{[2]} \overset{\mathrm{d}}{=} \theta_{t+1}^{[1]}$ for all $t \geq 0$ by comparing the dynamic of $(\theta_t^{[1]})_{t\geq 0}$ and $(\theta_t^{[2]})_{t\geq 0}$ as given in equations (32) and (34).

We thus have

$$\begin{aligned} W_2^2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\theta_{t+1}^{[1]}\right)\right) &= W_2^2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\theta_t^{[2]}\right)\right) \\ &\leq 2u_{cm}^2 \mathbb{E}\left[M_\eta(\theta_0^{[1]} - \theta_0^{[2]})\right](1-\alpha(1-\sqrt{\gamma}))^t, \end{aligned}$$

where the second inequality follows from equation (33). It follows that

$$\sum_{t=0}^{\infty} W_2^2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\theta_{t+1}^{[1]}\right)\right) \leq 2u_{cm}^2 \mathbb{E}\left[M_\eta(\theta_0^{[1]} - \theta_0^{[2]})\right] \sum_{t=0}^{\infty}(1-\alpha(1-\sqrt{\gamma}))^t < \infty.$$

Consequently, $\{\mathcal{L}(\theta_t^{[1]})\}_{t\geq 0}$ forms a Cauchy sequence with respect to the metric $W_2$. Since the space $\mathcal{P}_2(\mathbb{R}^d)$ endowed with $W_2$ is a Polish space, every Cauchy sequence converges [Vil09, Theorem 6.18]. Furthermore, convergence in Wasserstein 2-distance also implies weak convergence [Vil09, Theorem 6.9]. Therefore, we conclude that the sequence $\{\mathcal{L}(\theta_t^{[1]})\}_{t\geq 0}$ converges weakly to a limit distribution $\bar{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$.

Next, we show that $\bar{\mu}$ is independent of the initial iterate distribution of $\theta_0^{[1]}$. Suppose there exists another sequence $\{\widetilde{\theta}_t^{[1]}\}_{t\geq 0}$ with a different initial distribution that converges to a limit $\widetilde{\mu}$. By triangle inequality, we have

$$W_2(\bar{\mu}, \tilde{\mu}) \leq W_2\left(\bar{\mu}, \mathcal{L}\left(\theta_t^{[1]}\right)\right) + W_2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\widetilde{\theta}_t^{[1]}\right)\right) + W_2\left(\mathcal{L}\left(\widetilde{\theta}_t^{[1]}\right), \tilde{\mu}\right) \overset{t\to\infty}{\longrightarrow} 0.$$

Note that the last step holds since $W_2\left(\mathcal{L}\left(\theta_t^{[1]}\right), \mathcal{L}\left(\widetilde{\theta}_t^{[1]}\right)\right) \overset{t\to\infty}{\longrightarrow} 0$ by equation (58). We thus have $W_2(\bar{\mu}, \widetilde{\mu}) = 0$, which implies the uniqueness of the limit $\bar{\mu}$.

Finally, the following lemma bounds the second moment of the limit random vector $\theta^{(\alpha)}$.

**Lemma 2.** *Under Assumption 2(**1**), when $\alpha \leq \bar{\alpha}'$, we obtain*

$$\mathbb{E}[\|\theta^{(\alpha)} - \theta^*\|_2^2] \in \mathcal{O}(\alpha) \quad \text{and} \quad \mathbb{E}[\|\theta^{(\alpha)}\|_2^2] \in \mathcal{O}(1).$$

*Proof for Lemma 2.* We have shown that the sequence $\{\theta_t\}_{t\geq 0}$ converges weakly to $\theta^{(\alpha)}$ in $\mathcal{P}_2(\mathbb{R}^d)$. It is well known that weak convergence in $\mathcal{P}_2(\mathbb{R}^d)$ is equivalent to convergence in distribution and the convergence of the first two moments. As a result, we have

$$\mathbb{E}\left[\|\theta^{(\alpha)} - \theta^*\|_c^2\right] = \lim_{t\to\infty} \mathbb{E}\left[\|\theta_t - \theta^*\|_c^2\right]. \tag{35}$$

Taking $t \to \infty$ on both sides of equation (8) in Proposition 1 with $n = 1$ and combining with equation (35) yields

$$\mathbb{E}[\|\theta^{(\alpha)} - \theta^*\|_2^2] \leq \frac{1}{l_{cs}^2}\mathbb{E}[\|\theta^{(\alpha)} - \theta^*\|_c^2] \in \mathcal{O}(\alpha).$$

Therefore, we have

$$\mathbb{E}[\|\theta^{(\alpha)}\|_2^2] \leq 2\mathbb{E}(\|\theta^{(\alpha)} - \theta^*\|_2^2) + 2\|\theta^*\|_2^2 \in \mathcal{O}(1).$$

□

## B.2 Invariance

Moreover, we will show that the unique limit distribution $\bar{\mu}$ is also a stationary distribution for the Markov chain $\{\theta_t\}_{t\geq 0}$, as stated in the following lemma.

**Lemma 3.** *Let $\{\theta_t\}_{t\geq 0}$ and $\{\theta_t'\}_{t\geq 0}$ be two trajectories of iterates in equation (32), where $\mathcal{L}(\theta_0) = \bar{\mu}$ and $\mathcal{L}(\theta_0') \in \mathcal{P}_2(\mathbb{R}^{|\mathcal{S}||\mathcal{A}|})$ is arbitrary. we have*

$$W_2^2(\mathcal{L}(\theta_1), \mathcal{L}(\theta_1')) \leq \rho W_2^2(\mathcal{L}(\theta_0), \mathcal{L}(\theta_0')),$$

*where the quantity $\rho := \frac{u_{cm}^2}{l_{cm}^2}(1-\alpha(1-\sqrt{\gamma}))$ is independent of $\mathcal{L}(\theta_0')$. In particular, for any $t \geq 0$, if we set $\mathcal{L}(\theta_0') = \mathcal{L}(\theta_t)$, then*

$$W_2^2(\mathcal{L}(\theta_1), \mathcal{L}(\theta_{t+1})) \leq \rho W_2^2(\bar{\mu}, \mathcal{L}(\theta_t)).$$

*Proof of Lemma 3.* We couple the two processes $\{\theta_t\}_{t\geq 0}$ and $\{\theta'_t\}_{t\geq 0}$ such that

$$W_2^2\left(\mathcal{L}\left(\theta_0\right), \mathcal{L}(\theta'_0)\right) = \mathbb{E}\left[\|\theta_0 - \theta'_0\|_c^2\right].$$

Since $W_2$ is defined by infimum over all couplings, we have

$$\begin{aligned} W_2^2\left(\mathcal{L}\left(\theta_1\right), \mathcal{L}(\theta'_1)\right) &\leq \mathbb{E}\left[\|\theta_1 - \theta'_1\|_c^2\right] \\ &\leq 2u_{cm}^2 \mathbb{E}\left[M_\eta(\theta_1 - \theta'_1)\right] \\ &\leq 2u_{cm}^2 (1 - \alpha(1 - \sqrt{\gamma}))\mathbb{E}\left[M_\eta(\theta_0 - \theta'_0)\right] \\ &\leq \frac{u_{cm}^2}{l_{cm}^2}(1 - \alpha(1 - \sqrt{\gamma}))\mathbb{E}\left[\|\theta_0 - \theta'_0\|_c^2\right] = \rho W_2^2\left(\mathcal{L}\left(\theta_0\right), \mathcal{L}(\theta'_0)\right), \end{aligned}$$

where $\rho = \frac{u_{cm}^2}{l_{cm}^2}(1 - \alpha(1 - \sqrt{\gamma}))$. □

By triangle inequality, we obtain

$$\begin{aligned} W_2\left(\mathcal{L}\left(\theta_1\right), \bar{\mu}\right) &\leq W_2\left(\mathcal{L}\left(\theta_1\right), \mathcal{L}\left(\theta_{t+1}\right)\right) + W_2\left(\mathcal{L}\left(\theta_{t+1}\right), \bar{\mu}\right) \\ &\leq \sqrt{\rho} W^2\left(\bar{\mu}, \mathcal{L}(\theta_t)\right) + W_2\left(\mathcal{L}\left(\theta_{t+1}\right), \bar{\mu}\right) \stackrel{t\to\infty}{\longrightarrow} 0, \end{aligned} \tag{36}$$

where the second inequality holds by Lemma 3 and last step comes from the weak convergence result. Therefore, we have proved that $\{\theta_t\}_{t\geq 0}$ converges to a unique stationary distribution $\bar{\mu}$.

### B.3 Convergence rate

Consider the coupled processes defined as equation (32). Suppose that the initial iterate $\theta_0^{[2]}$ follows the stationary distribution $\bar{\mu}$, thus $\mathcal{L}(\theta_t^{[2]}) = \bar{\mu}$ for all $t \geq 0$. By equation (33), we have for all $t \geq 0$ :

$$\begin{aligned} W_2^2\left(\mathcal{L}(\theta_t^{[1]}), \bar{\mu}\right) &= W_2^2\left(\mathcal{L}(\theta_t^{[1]}), \mathcal{L}(\theta_t^{[2]})\right) \\ &\leq 2u_{cm}^2 \mathbb{E}\left[M_\eta(\theta_0^{[1]} - \theta_0^{[2]})\right](1 - \alpha(1 - \sqrt{\gamma}))^t \\ &\leq 2u_{cm}^2 \mathbb{E}\left[M_\eta(\theta_0^{[1]} - \theta^{(\alpha)})\right](1 - \alpha(1 - \sqrt{\gamma}))^t \end{aligned}$$

Lemma 2 states that the second moment of $\theta^{(\alpha)}$ is bounded by a constant. Combining this bound with above equation, we obtain the desired bound $W_2^2(\mathcal{L}(\theta_t), \mu) \leq c \cdot (1 - \alpha(1 - \sqrt{\gamma}))^t$, where $c$ is a universal constant that is independent with $\alpha$ and $t$.

## C Proof of Theorem 2

In this section, we prove Theorem 2, which establishes steady-state convergence under the additive noise setting. We follow the three-step strategy outlined in Section 1.2.

We start by using equation (7) to obtain the following dynamic for $Y_t$:

$$Y_{t+1} = (1 - \alpha)Y_t + \sqrt{\alpha}\left(\mathcal{T}(\sqrt{\alpha}Y_t + \theta^*) - \mathcal{T}(\theta^*) + w_t\right). \tag{37}$$

### C.1 Step 1: Gaussian Noise and Rational Stepsize

We consider a pair of coupled $\{Y_t\}_{t\geq 0}$ and $\{Y'_t\}_{t\geq 0}$, defined as

$$\begin{aligned} Y_{t+1} &= (1 - \alpha)Y_t + \sqrt{\alpha}\Big(\mathcal{T}(\sqrt{\alpha}Y_t + \theta^*) - \mathcal{T}(\theta^*) + \frac{w'_{kt} + \cdots + w'_{kt+k-1}}{\sqrt{k}}\Big), \\ Y'_{t+1} &= (1 - \frac{\alpha}{k})Y'_t + \sqrt{\frac{\alpha}{k}}\Big(\mathcal{T}\Big(\sqrt{\frac{\alpha}{k}}Y'_t + \theta^*\Big) - \mathcal{T}(\theta^*) + w'_t\Big), \end{aligned} \tag{38}$$

where $\{w'_t\}_{t\geq 0}$ are i.i.d noise with normal distribution, zero mean and the same variance as $\{w_t\}_{t\geq 0}$ and $k \geq 1$ is an integer. Because $\{w'_t\}_{t\geq 0}$ are i.i.d noise with normal distribution, $\frac{w'_{kt}+\cdots+w'_{kt+k-1}}{\sqrt{k}}$ has the same distribution as $w'_t$. Direct calculation gives

$$\begin{aligned} Y'_{kt+k} =& \left(1-\frac{\alpha}{k}\right)^k Y'_{kt} + \sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}\left(\left(1-\frac{\alpha}{k}\right)^j - 1\right)\left(\mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*\right) - \mathcal{T}(\theta^*) + w'_{kt+k-1-j}\right) \\ &+ \sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}\left(\mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*\right) - \mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*\right)\right) \\ &+ \sqrt{\alpha k}\left(\mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*\right) - \mathcal{T}(\theta^*)\right) + \sqrt{\alpha}\frac{w'_{kt}+\cdots+w'_{kt+k-1}}{\sqrt{k}}. \end{aligned} \tag{39}$$

Combining equations (38) and (39), we obtain

$$\begin{aligned} Y_{t+1} - Y'_{kt+k} =&(1-\alpha)(Y_t - Y'_{kt}) \\ &+ \left(1-\alpha-\left(1-\frac{\alpha}{k}\right)^k\right)Y'_{kt} + \sqrt{\alpha}\left(\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*) - \mathcal{T}(\sqrt{\alpha}Y'_{kt}+\theta^*)\right) \\ &+ \sqrt{\alpha}\left(\left(\mathcal{T}\left(\sqrt{\alpha}Y'_{kt}+\theta^*\right) - \mathcal{T}(\theta^*)\right) - \sqrt{k}\left(\mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*\right) - \mathcal{T}(\theta^*)\right)\right) \\ &+ \sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}\left(1-(1-\frac{\alpha}{k})^j\right)\left(\mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*\right) - \mathcal{T}(\theta^*) + w'_{kt+k-1-j}\right) \\ &+ \sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}\left(\mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*\right) - \mathcal{T}\left(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*\right)\right) \\ :=&(1-\alpha)(Y_t - Y'_{kt}) + A, \end{aligned}$$

where $A$ collects all but the first term on the RHS. Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of above equation and by property (1) in Proposition 5, we obtain

$$M_\eta(Y_{t+1} - Y'_{kt+k}) \leq (1-\alpha)^2 M_\eta(Y_t - Y'_{kt}) + (1-\alpha)\underbrace{\langle \nabla M_\eta(Y_t - Y'_{kt}), A\rangle}_{T_1} + \underbrace{\frac{1}{2\eta}\|A\|_2^2}_{T_2}. \tag{40}$$

The following lemmas, proved in Sections C.1.1 and C.1.2 to follow, control the $T_1$ and $T_2$ terms above.

**Lemma 4.** *Under the setting of Theorem 2, we have*

$$\mathbb{E}[T_1] \leq \frac{2\alpha\gamma u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(Y_t - Y'_{kt})] + o(\alpha).$$

**Lemma 5.** *Under the setting of Theorem 2, we have*

$$\mathbb{E}[T_2] \leq \frac{5\alpha^2 u_{cm}^2\gamma^2}{\eta l_{cs}^2}\mathbb{E}[M_\eta(Y_t - Y'_{kt})] + o(\alpha).$$

Plugging the above bounds for $T_1$ and $T_2$ into equation (40), we obtain

$$\mathbb{E}[M_\eta(Y_{t+1} - Y'_{kt+k})] \leq \left(1 - 2\alpha(1 - \frac{(1-\alpha)\gamma u_{cm}}{l_{cm}}) + \mathcal{O}(\alpha^2)\right)\mathbb{E}[M_\eta(Y_t - Y'_{kt})] + o(\alpha).$$

By the similar argument as in the proof of Lemma 1, we can always choose proper $\eta, \bar{\alpha}$ such that for $\forall\alpha \leq \bar{\alpha}$, there exist $t_\alpha$ such that for all $t \geq t_\alpha$, we obtain

$$\mathbb{E}[M_\eta(Y_{t+1} - Y'_{kt+k})] \leq (1 - \alpha(1-\sqrt{\gamma}))\,\mathbb{E}[M_\eta(Y_t - Y'_{kt})] + o(\alpha)$$

which implies

$$\lim_{t\to\infty} \mathbb{E}[M_\eta(Y_t - Y'_{kt})] \in o(1).$$

By triangle inequality, we have

$$\begin{aligned}
W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\big) &\le \lim_{t\to\infty} \Big\{ W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y_t)\big) + W_2\big(\mathcal{L}(Y_t), \mathcal{L}(Y'_{kt})\big) + W_2\big(\mathcal{L}(Y'_{kt}), \mathcal{L}(Y^{(\alpha/k)})\big) \Big\} \\
&\overset{\text{(i)}}{=} \lim_{t\to\infty} W_2\big(\mathcal{L}(Y_t), \mathcal{L}(Y'_{kt})\big) \\
&\overset{\text{(ii)}}{\le} \lim_{t\to\infty} \sqrt{\mathbb{E}[\|Y_t - Y'_{kt}\|_c^2]} \overset{\text{(iii)}}{\le} \lim_{t\to\infty} \sqrt{2u_{cm}^2 \mathbb{E}[M(Y_t - Y'_{kt})]} \in o(1),
\end{aligned}$$

where (i) follows from Theorem 1, (ii) holds by the definition of $W_2$ distance, and (iii) is true by Proposition 5.

Therefore, we have that for all $k \in \mathbb{N}^+$ and $\alpha > 0$,

$$W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\big) \in o(1). \tag{41}$$

When $k \in \mathbb{Q}^+, k > 1$ and $\alpha > 0$, let $k = \frac{p}{q}$. We have

$$\begin{aligned}
W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\big) &\le W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/p)})\big) + W_2\big(\mathcal{L}(Y^{(\alpha/p)}), \mathcal{L}(Y^{(\alpha/k)})\big) \\
&\overset{(i)}{\le} o(1) + o(1) \in o(1),
\end{aligned}$$

where (i) holds because $\frac{\alpha}{p} = \frac{\frac{\alpha}{k}}{q}$ and $\frac{\alpha}{k} \le \alpha$.

Then, for any rational sub-sequence $\{\alpha_j\}_{j=0}^\infty$, $\alpha_j \to 0$, $\{\mathcal{L}(Y^{(\alpha_j)})\}_{j=0}^\infty$ is a Cauchy sequence with respect to $W_2$, therefore has a limit. Assume we have two different rational sub-sequence $\{\alpha_j\}_{j=0}^\infty$ and $\{\beta_j\}_{j=0}^\infty$ such that the limits of $\{\mathcal{L}(Y^{(\alpha_j)})\}_{j=0}^\infty$ and $\{\mathcal{L}(Y^{(\beta_j)})\}_{j=0}^\infty$ are different with respect to $W_2$. Let $\mathcal{L}(\bar{Y})$ be the limit of $\{\mathcal{L}(Y^{(\alpha_j)})\}_{j=0}^\infty$ and $\mathcal{L}(\hat{Y})$ be the limit of $\{\mathcal{L}(Y^{(\beta_j)})\}_{j=0}^\infty$. Then, there exists $\epsilon > 0$, such that $W_2\big(\mathcal{L}(\bar{Y}), \mathcal{L}(\hat{Y})\big) > \epsilon$. Let $\gamma_{2j} = \alpha_j, \gamma_{2j+1} = \beta_j$. Then, $\{\gamma_j\}_{j=0}^\infty$ forms a rational sequence and $\gamma \to 0$. Then, we obtain

$$\begin{aligned}
&\lim_{j\to\infty} W_2\big(\mathcal{L}(Y^{(\gamma_{2j})}), \mathcal{L}(Y^{(\gamma_{2j+1})})\big) = \lim_{j\to\infty} W_2\big(\mathcal{L}(Y^{(\alpha_j)}), \mathcal{L}(Y^{(\beta_j)})\big) \\
&\quad = \lim_{j\to\infty} \Big\{ W_2\big(\mathcal{L}(\bar{Y}), \mathcal{L}(Y^{(\alpha_j)})\big) + W_2\big(\mathcal{L}(Y^{(\alpha_j)}), \mathcal{L}(Y^{(\beta_j)})\big) + W_2\big(\mathcal{L}(Y^{(\beta_j)}), \mathcal{L}(\hat{Y})\big) \Big\} \\
&\quad \ge W_2\big(\mathcal{L}(\bar{Y}), \mathcal{L}(\hat{Y})\big) > \epsilon,
\end{aligned}$$

which contradicts with the fact that $\{\mathcal{L}(Y^{(\gamma_j)})\}_{j=0}^\infty$ is a Cauchy sequence with respect to $W_2$. Therefore, for any rational sub-sequence $\{\alpha_j\}_{j=0}^\infty$, $\alpha_j \to 0$, $\{\mathcal{L}(Y^{(\alpha_j)})\}_{j=0}^\infty$ converge to a unique limit with respect to $W_2$. That is, there exists a unique random variable $Y$ such that

$$\lim_{\alpha\to 0, \alpha\in\mathbb{Q}^+} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) = 0.$$

This completes the first step of the proof of Theorem 2.

### C.1.1 Proof of Lemma 4 on $T_1$

By property (4) in Proposition 5 and $\{w'_{kt+k-1-j}\}_{j=0}^{k-1}$ being i.i.d. zero mean noise and independent with $Y_t$ and $Y'_{kt}$, we obtain

$$\begin{aligned}
\mathbb{E}[T_1] \le &\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|(1 - \alpha - (1 - \frac{\alpha}{k})^k) Y'_{kt}\|_m] && (T_{11}) \\
&+ \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sqrt{\alpha}\left(\mathcal{T}(\sqrt{\alpha} Y_t + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y'_{kt} + \theta^*)\right)\|_m] && (T_{12}) \\
&+ \mathbb{E}\Big[\|Y_t - Y'_{kt}\|_m \|\sqrt{\alpha}\Big(\big(\mathcal{T}(\sqrt{\alpha} Y'_{kt} + \theta^*) - \mathcal{T}(\theta^*)\big) - \sqrt{k}\big(\mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt} + \theta^*) - \mathcal{T}(\theta^*)\big)\Big)\|_m\Big] && (T_{13})
\end{aligned}$$

$$+\,\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j)\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*)-\mathcal{T}(\theta^*)\right)\|_m] \tag{$T_{14}$}$$

$$+\,\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*)-\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*)\right)\|_m]. \tag{$T_{15}$}$$

Below, we bound the terms $T_{11} \sim T_{15}$ separately.

**The $T_{11}$ Term:** We begin with

$$T_{11} = |1-\alpha-(1-\frac{\alpha}{k})^k|\mathbb{E}[\|Y_t - Y'_{kt}\|_m\|Y'_{kt}\|_m].$$

Note that $f(x) = (1-\frac{\alpha}{x})^x$ increases monotonically when $x \ge \alpha$. Therefore, when $\alpha \le 1$, we obtain

$$|1-\alpha-(1-\frac{\alpha}{k})^k| \le \lim_{k\to\infty}(1-\frac{\alpha}{k})^k - 1 + \alpha = \exp(-\alpha) - 1 + \alpha \in \mathcal{O}(\alpha^2). \tag{42}$$

By Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\mathbb{E}[\|Y_t - Y'_{kt}\|_m\|Y'_{kt}\|_m] &\le \mathbb{E}[\|Y_t\|_m\|Y'_{kt}\|_m] + \mathbb{E}[\|Y'_{kt}\|_m\|Y'_{kt}\|_m] \\ &\le \sqrt{\mathbb{E}[\|Y_t\|_m^2]\mathbb{E}[\|Y'_{kt}\|_m^2]} + \mathbb{E}[\|Y'_{kt}\|_m^2] \overset{\text{(i)}}{\in} \mathcal{O}(1),\end{aligned} \tag{43}$$

where (i) holds by the following Corollary 1(**1**) and choosing a sufficiently large $t$ (note that Corollary 1 is parameterized by an integer $n \ge 1$). Therefore, we conclude that $T_{11} \in \mathcal{O}(\alpha^2)$.

**Corollary 1** (***n***)**.** *For integer $n \ge 1$, under Assumption 2(**n**), there exists $\bar\alpha$ such that for any $\alpha \le \bar\alpha$, there exists $t_{\alpha,n} > 0$ and*

$$\mathbb{E}[\|Y_t^{(\alpha)}\|^{2n}] \le c_n \mathbb{E}[\|Y_{t_{\alpha,n}}^{(\alpha)}\|^{2n}](1-\alpha(1-\sqrt{\gamma}))^{t-t_{\alpha,n}} + c'_n, \quad \forall t \ge t_{\alpha,n},$$

*where $\|\cdot\|$ is an arbitrary norm and $\{c_n\}_{n\ge1}$ and $\{c'_n\}_{n\ge1}$ are universal constants that are independent with $\alpha$ and $t$. Moreover, $t_{\alpha,1} = 0$.*

*Proof of Corollary 1.* By the equivalence of all norms on $\mathbb{R}^d$, we can obtain the Corollary 1(**n**) by dividing $\alpha^n$ to both sides of equation (8) in Proposition 1(**n**). □

**The $T_{12}$ Term:** Turning to $T_{12}$, we have

$$T_{12} \le \frac{\sqrt{\alpha}}{l_{cm}}\mathbb{E}[\|Y_t - Y'_{kt}\|_m\|\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y'_{kt}+\theta^*)\|_c] \le \frac{2\alpha\gamma u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(Y_t - Y'_{kt})].$$

**The $T_{13}$ Term:** For $T_{13}$, by Cauchy–Schwarz inequality, we obtain

$$T_{13} \le \sqrt{\alpha}\sqrt{\mathbb{E}[\|Y_t - Y'_{kt}\|_m^2]\mathbb{E}[\|\left(\mathcal{T}(\sqrt{\alpha}Y'_{kt}+\theta^*)-\mathcal{T}(\theta^*)-\sqrt{k}\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*)-\mathcal{T}(\theta^*)\right)\right)\|_m^2]}.$$

Note that $\mathbb{E}[\|Y_t - Y'_{kt}\|_m^2] \le 2\mathbb{E}[\|Y_t\|_m^2] + 2\mathbb{E}[\|Y'_{kt}\|_m^2] \in \mathcal{O}(1)$. For the second expectation term on above RHS, we have

$$\mathbb{E}\Big[\Big\|\Big(\mathcal{T}(\sqrt{\alpha}Y'_{kt}+\theta^*)-\mathcal{T}(\theta^*)-\sqrt{k}\Big(\mathcal{T}\Big(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*\Big)-\mathcal{T}(\theta^*)\Big)\Big)\Big\|_m^2\Big] \tag{44}$$

$$\overset{\text{(i)}}{=}\mathbb{E}\Big[\Big\|g\big(F(\sqrt{\alpha}Y'_{kt})\big)-\sqrt{k}g\big(F(\sqrt{\frac{\alpha}{k}}Y'_{kt})\big)\Big\|_m^2\mathbb{1}(\alpha^{\frac14}Y'_{kt}\in B^d(0,\epsilon))\Big] \tag{45}$$

$$+\,\mathbb{E}\Big[\Big\|\mathcal{T}(\sqrt{\alpha}Y'_{kt}+\theta^*)-\mathcal{T}(\theta^*)-\sqrt{k}\Big(\mathcal{T}\Big(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*\Big)-\mathcal{T}(\theta^*)\Big)\Big\|_m^2\mathbb{1}(\alpha^{\frac14}Y'_{kt}\notin B^d(0,\epsilon))\Big]. \tag{46}$$

where (i) holds because of Assumption 3.

By Taylor expansion, when $\alpha \leq 1$, there always exist random variable $\lambda_1, \lambda_2 \in [0,1]^n$ such that

$$
\begin{aligned}
(45) &= \mathbb{E}[\|g(F(\sqrt{\alpha}Y'_{kt})) - g(\sqrt{k}F(\sqrt{\frac{\alpha}{k}}Y'_{kt}))\|_m^2 \mathbb{1}(\alpha^{\frac{1}{4}}Y'_{kt} \in B^d(0,\epsilon))] \\
&= \alpha\mathbb{E}[\|g(\nabla F(\lambda_1, \sqrt{\alpha}Y'_{kt})Y'_{kt}) - g(\nabla F(\lambda_2, \sqrt{\frac{\alpha}{k}}Y'_{kt})Y'_{kt})\|_m^2 \mathbb{1}(\alpha^{\frac{1}{4}}Y'_{kt} \in B^d(0,\epsilon))] \\
&= \alpha\mathbb{E}[\| \left(g(\nabla F(\lambda_1, \sqrt{\alpha}Y'_{kt})\frac{Y'_{kt}}{\|Y'_{kt}\|_2}) - g(\nabla F(\lambda_2, \sqrt{\frac{\alpha}{k}}Y'_{kt})\frac{Y'_{kt}}{\|Y'_{kt}\|_2})\right)\|_m^2 \|Y'_{kt}\|_2^2 \mathbb{1}(\alpha^{\frac{1}{4}}Y'_{kt} \in B^d(0,\epsilon))] \\
&\leq \alpha \sqrt{\underbrace{\mathbb{E}[\| \left(g(\nabla F(\lambda_1, \sqrt{\alpha}Y'_{kt})\frac{Y'_{kt}}{\|Y'_{kt}\|_2}) - g(\nabla F(\lambda_2, \sqrt{\frac{\alpha}{k}}Y'_{kt})\frac{Y'_{kt}}{\|Y'_{kt}\|_2})\right)\|_m^4 \mathbb{1}(\alpha^{\frac{1}{4}}Y'_{kt} \in B^d(0,\epsilon))]}_{T_{131}}} \underbrace{\sqrt{\mathbb{E}[\|Y'_{kt}\|_2^4]}}_{\in\mathcal{O}(1)},
\end{aligned}
$$

where we use $\nabla F(\lambda_i, \sqrt{\alpha}Y'_{kt})$ to denote the vector that $[\nabla F(\lambda_i, \sqrt{\alpha}Y'_{kt})]_j = \nabla F_j(\lambda_{ij}\sqrt{\alpha}Y'_{kt})$ for $i = 1,2$ and $j \in [n]$.

For $\forall\epsilon_0 > 0$, by continuity of $g(\cdot)$, $\exists\delta_0 > 0$, such that $\|g(\theta) - g(\nabla F(0))\|_2 \leq \epsilon_0$ when $\|\theta - \nabla F(0)\|_2 \leq \delta_0$. By the continuity of $\nabla F(\cdot)$ at 0, $\exists\delta_1 > 0$, such that $\|\nabla F(\theta) - \nabla F(0)\|_2 \leq \delta_0$ when $\|\theta\| \leq \delta_1$. Therefore, we obtain $\|g(\nabla F(\theta)) - g(\nabla F(0))\|_2 \leq \epsilon_0$ when $\|\theta\| \leq \delta_1$. Given $\alpha^{\frac{1}{4}}Y'_{kt} \in B^d(0,\epsilon)$, we can always let $\alpha$ small enough such that $\|\sqrt{\alpha}Y'_{kt}\|_2 \leq \delta_1, \|\sqrt{\frac{\alpha}{k}}Y'_{kt}\|_2 \leq \delta_1$. Therefore, the variables within the term $T_{131}$ are always bounded, which implies $\lim_{\alpha\to 0} T_{131} = 0$. Therefore, we have

$$
\mathbb{E}[\|g(F(\sqrt{\alpha}Y'_{kt})) - \sqrt{k}g(F(\sqrt{\frac{\alpha}{k}}Y'_{kt}))\|_m^2 \mathbb{1}(\alpha^{\frac{1}{4}}Y'_{kt} \in B^d(0,\epsilon))] \in o(\alpha).
$$

For the term in (46), by Cauchy–Schwarz inequality and Markov inequality, we obtain

$$
\begin{aligned}
(46) &\leq \frac{2\gamma\alpha u_{cm}^2}{l_{cm}^2}\mathbb{E}[\|Y'_{kt}\|_m^2 \mathbb{1}(\alpha^{\frac{1}{4}}Y'_{kt} \notin B^d(0,\epsilon))] \\
&\leq \frac{2\gamma\alpha u_{cm}^2}{l_{cm}^2}\sqrt{\mathbb{E}[\|Y'_{kt}\|_m^4]}\sqrt{\mathbb{P}(\|Y'_{kt}\|_2^4 \geq \frac{\epsilon^4}{\alpha})} \leq \frac{2\gamma\alpha u_{cm}^2}{l_{cm}^2}\sqrt{\mathbb{E}[\|Y'_{kt}\|_m^4]}\sqrt{\frac{\alpha\mathbb{E}[\|Y'_{kt}\|_2^4]}{\epsilon^4}} \in \mathcal{O}(\alpha^{\frac{3}{2}}),
\end{aligned}
$$

where the last step follows from $\mathbb{E}[\|Y'_{kt}\|_m^4] = O(1)$ and $\mathbb{E}[\|Y'_{kt}\|_2^4]/\epsilon^4 = O(1)$. Combining all the analysis together, we obtain that (44) $\in o(\alpha)$, which in turn implies $T_{13} \in o(\alpha)$.

**The $T_{14}$ Term:** For $T_{14}$, we have

$$
\begin{aligned}
T_{14} &= \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j)\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j} + \theta^*) - \mathcal{T}(\theta^*)\right)\|_m] \\
&\leq \sum_{j=0}^{k-1}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sqrt{\frac{\alpha}{k}}(1-(1-\frac{\alpha}{k})^j)\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j} + \theta^*) - \mathcal{T}(\theta^*)\right)\|_m] \\
&\leq \frac{\gamma}{l_{cm}}\sum_{j=0}^{k-1}\frac{\alpha}{k}(1-(1-\frac{\alpha}{k})^j)\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt+k-1-j}\|_c] \qquad (47) \\
&\leq \frac{\gamma}{l_{cm}}\sum_{j=0}^{k-1}\frac{\alpha}{k}(1-(1-\frac{\alpha}{k})^j)\left(\sqrt{\mathbb{E}[\|Y_t\|_m^2]\mathbb{E}[\|Y'_{kt+k-1-j}\|_c^2]} + \sqrt{\mathbb{E}[\|Y'_{kt}\|_m^2]\mathbb{E}[\|Y'_{kt+k-1-j}\|_c^2]}\right) \\
&\leq \frac{\gamma}{l_{cm}}\sum_{j=0}^{k-1}\frac{\alpha}{k}(1-(1-\frac{\alpha}{k})^j)\cdot\mathcal{O}(1) = \mathcal{O}\left(\alpha - 1 + (1-\frac{\alpha}{k})^k\right) \overset{\text{(i)}}{\in} \mathcal{O}(\alpha^2),
\end{aligned}
$$

where (i) holds by equation (42).

**The $T_{15}$ Term:** Finally, we turn to $T_{15}$:

$$
\begin{aligned}
T_{15} &= \mathbb{E}[\|Y_t - Y'_{kt}\|_m \| \sqrt{\frac{\alpha}{k}} \sum_{j=0}^{k-1} \left( \mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt} + \theta^*) - \mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt+k-1-j} + \theta^*) \right) \|_m] \\
&\leq \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt+k-1-j} - Y'_{kt}\|_c].
\end{aligned}
$$

By equation (38), we obtain

$$
\begin{aligned}
&\|Y'_{kt+k-1-j} - Y'_{kt}\|_c \\
=&\|((1-\frac{\alpha}{k})^{k-1-j} - 1)Y'_{kt} + \sqrt{\frac{\alpha}{k}} \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} \left( \mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt+k-1-j-l} + \theta^*) - \mathcal{T}(\theta^*) + w'_{kt+k-1-j-l} \right) \|_c.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
T_{15} \leq& \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|((1-\frac{\alpha}{k})^{k-1-j} - 1)Y'_{kt}\|_c] && (T_{151}) \\
&+ \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \| \sqrt{\frac{\alpha}{k}} \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} \left( \mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt+k-1-j-l} + \theta^*) - \mathcal{T}(\theta^*) \right) \|_c] && (T_{152}) \\
&+ \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \| \sqrt{\frac{\alpha}{k}} \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} w'_{kt+k-1-j-l} \|_c]. && (T_{153})
\end{aligned}
$$

We analyze three terms $T_{151}, T_{152}, T_{153}$ separately. Note that

$$
T_{151} \leq \frac{\alpha\gamma u_{cm}}{k l_{cm}} \underbrace{\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt}\|_m]}_{\in \mathcal{O}(1)} \sum_{j=0}^{k-1} (1-(1-\frac{\alpha}{k})^{k-1-j}) \leq \mathcal{O}(1) \cdot \frac{\alpha}{k} \sum_{j=0}^{k-1} ((1-\frac{\alpha}{k})^{k-1-j} - 1) \overset{(i)}{\in} \mathcal{O}(\alpha^2),
$$

where (i) holds by equation (42). For $T_{152}$, we have

$$
\begin{aligned}
T_{152} &\leq \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \sum_{l=1}^{k-1-j} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \| \sqrt{\frac{\alpha}{k}} (1-\frac{\alpha}{k})^{l-1} \left( \mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt+k-1-j-l} + \theta^*) - \mathcal{T}(\theta^*) \right) \|_c] \\
&\leq \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \sum_{l=1}^{k-1-j} \frac{\alpha}{k} (1-\frac{\alpha}{k})^{l-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt+k-1-j-l}\|_c] \\
&\leq \frac{\alpha\gamma}{k l_{cm}} \sum_{j=0}^{k-1} \sum_{l=1}^{k-1-j} \frac{\alpha}{k} (1-\frac{\alpha}{k})^{l-1} \underbrace{\left( \sqrt{\mathbb{E}[\|Y_t\|_m^2] \mathbb{E}[\|Y'_{kt+k-1-j-l}\|_c^2]} + \sqrt{\mathbb{E}[\|Y'_{kt}\|_m^2] \mathbb{E}[\|Y'_{kt+k-1-j-l}\|_c^2]} \right)}_{\in \mathcal{O}(1)} \\
&\leq \mathcal{O}(1) \cdot \frac{\alpha^2}{k^2} \sum_{j=0}^{k-1} \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} \leq \mathcal{O}(1) \cdot \frac{\alpha^2}{k^2} \cdot k^2 \in \mathcal{O}(\alpha^2).
\end{aligned}
$$

Lastly, we have

$$
\begin{aligned}
T_{153} &\overset{\text{(i)}}{=} \frac{\alpha\gamma}{kl_{cm}} \sum_{j=0}^{k-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m] \mathbb{E}[\| \sqrt{\frac{\alpha}{k}} \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} w'_{kt+k-1-j-l}\|_c] \\
&\overset{\text{(ii)}}{\leq} \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=0}^{k-1} \mathbb{E}[\| \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} w'_{kt+k-1-j-l}\|_2] \\
&\leq \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=0}^{k-1} \sqrt{\mathbb{E}[\| \sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{l-1} w'_{kt+k-1-j-l}\|_2^2]} \\
&\leq \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=0}^{k-1} \sqrt{\sum_{l=1}^{k-1-j} (1-\frac{\alpha}{k})^{2l-2}} \overset{\text{(iii)}}{\leq} \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=0}^{k-1} \sqrt{k-1-j} \in \mathcal{O}(\alpha^{\frac{3}{2}}),
\end{aligned}
$$

where (i) holds because $Y_t$ and $Y'_{kt}$ are independent with $w'_{kt+k-1-j-l}$ for $j = 0, \ldots, k-1$ and $l = 1, \ldots, k-1-j$, (ii) follows as $\mathbb{E}[\|Y_t - Y'_{kt}\|_m] \in \mathcal{O}(1)$, and (iii) holds because $\sum_{j=0}^{k-1} \sqrt{k-1-j} \in \mathcal{O}(k^{\frac{3}{2}})$.

Putting the bounds of $T_{151}, T_{152}$ and $T_{153}$ together, we obtain $T_{15} \in \mathcal{O}(\alpha^{\frac{3}{2}})$.

Finally, combining our bounds for $T_{11} \sim T_{15}$, we have

$$
\mathbb{E}[T_1] \leq \frac{2\alpha\gamma u_{cm}}{l_{cm}} \mathbb{E}[M_\eta(Y_t - Y'_{kt})] + o(\alpha),
$$

thereby completing the proof of Lemma 4.

### C.1.2 Proof of Lemma 5 on $T_2$

By Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
\mathbb{E}[T_2] \leq & \frac{5}{2\eta} \mathbb{E}[\|(1-\alpha-(1-\frac{\alpha}{k})^k) Y'_{kt}\|_2^2] && (T_{21}) \\
&+ \frac{5}{2\eta} \mathbb{E}[\| \sqrt{\alpha} \left(\mathcal{T}(\sqrt{\alpha} Y_t + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y'_{kt} + \theta^*)\right) \|_2^2] && (T_{22}) \\
&+ \frac{5}{2\eta} \mathbb{E}[\| \sqrt{\alpha} \left(\mathcal{T}(\sqrt{\alpha} Y'_{kt} + \theta^*) - \mathcal{T}(\theta^*) - \sqrt{k} \left(\mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt} + \theta^*) - \mathcal{T}(\theta^*)\right)\right) \|_2^2] && (T_{23}) \\
&+ \frac{5}{2\eta} \mathbb{E}[\| \sqrt{\frac{\alpha}{k}} \sum_{j=0}^{k-1} (1-(1-\frac{\alpha}{k})^j) \left(\mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt+k-1-j} + \theta^*) - \mathcal{T}(\theta^*) + w'_{kt+k-1-j}\right) \|_2^2] && (T_{24}) \\
&+ \frac{5}{2\eta} \mathbb{E}[\| \sqrt{\frac{\alpha}{k}} \sum_{j=0}^{k-1} \left(\mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt} + \theta^*) - \mathcal{T}(\sqrt{\frac{\alpha}{k}} Y'_{kt+k-1-j} + \theta^*)\right) \|_2^2]. && (T_{25})
\end{aligned}
$$

Below, we bound $T_{21} \sim T_{25}$ separately.

**The $T_{21}$ Term:** We begin with $T_{21}$:

$$
T_{21} \leq \frac{5}{2\eta} \left(1-\alpha-(1-\frac{\alpha}{k})^k\right)^2 \cdot \mathbb{E}[\|Y'_{kt}\|_2^2] \leq \frac{5}{2\eta} \left(1-\alpha-(1-\frac{\alpha}{k})^k\right)^2 \cdot \mathcal{O}(1) \overset{\text{(i)}}{\in} \mathcal{O}(\alpha^4),
$$

where (i) holds by equation (42).

**The $T_{22}$ Term:** For $T_{22}$, we have

$$
T_{22} \leq \frac{5\alpha}{2\eta l_{cs}^2} \mathbb{E}\left[\|\mathcal{T}(\sqrt{\alpha} Y_t + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y'_{kt} + \theta^*)\|_c^2\right] \leq \frac{5\alpha^2 u_{cm}^2 \gamma^2}{\eta l_{cs}^2} \mathbb{E}[M_\eta(Y_t - Y'_{kt})].
$$

**The $T_{23}$ Term:** Using the bound (44) $\in o(\alpha)$ and the equivalence of all norms in $\mathbb{R}^d$, we obtain that $T_{23} \in o(\alpha^2)$.

**The $T_{24}$ Term:** By Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
T_{24} \leq & \frac{5}{\eta}\mathbb{E}[\|\sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j)\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*)-\mathcal{T}(\theta^*)\right)\|_2^2] \\
& + \frac{5}{\eta}\mathbb{E}[\|\sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j)w'_{kt+k-1-j}\|_2^2] \\
\leq & \frac{5\gamma^2}{\eta l_{cs}^2}\frac{\alpha^2}{k^2}\mathbb{E}[\|\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j)Y'_{kt+k-1-j}\|_c^2] + \frac{5}{\eta}\frac{\alpha}{k}\sum_{j=0}^{k-1}\mathbb{E}[\|(1-(1-\frac{\alpha}{k})^j)w'_{kt+k-1-j}\|_2^2] \\
\leq & \frac{5\gamma^2}{\eta l_{cs}^2}\frac{\alpha^2}{k}\sum_{j=0}^{k-1}\mathbb{E}[\|(1-(1-\frac{\alpha}{k})^j)Y'_{kt+k-1-j}\|_c^2] + \frac{5}{\eta}\frac{\alpha}{k}\sum_{j=0}^{k-1}\mathbb{E}[\|(1-(1-\frac{\alpha}{k})^j)w'_{kt+k-1-j}\|_2^2] \\
\overset{(i)}{\leq} & \mathcal{O}(\alpha^2)+\mathcal{O}(1)\cdot\frac{\alpha}{k}\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j)^2 \leq \mathcal{O}(\alpha^2)+\mathcal{O}(1)\cdot\frac{\alpha}{k}\sum_{j=0}^{k-1}(1-(1-\frac{\alpha}{k})^j) \overset{(ii)}{\in} \mathcal{O}(\alpha^2),
\end{aligned}
$$

where (i) holds because $\sum_{j=0}^{k-1}\mathbb{E}[\|(1-(1-\frac{\alpha}{k})^j)Y'_{kt+k-1-j}\|_c^2] \in \mathcal{O}(k)$, and (ii) holds by equation (42).

**The $T_{25}$ Term:** For $T_{25}$, we have

$$
\begin{aligned}
T_{25} &\leq \frac{5}{2\eta}\mathbb{E}[\|\sqrt{\frac{\alpha}{k}}\sum_{j=0}^{k-1}\left(\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt}+\theta^*)-\mathcal{T}(\sqrt{\frac{\alpha}{k}}Y'_{kt+k-1-j}+\theta^*)\right)\|_2^2] \\
&\leq \mathcal{O}(1)\cdot\frac{\alpha^2}{k}\sum_{j=0}^{k-1}\mathbb{E}[\|Y'_{kt}-Y'_{kt+k-1-j}\|_c^2] \in \mathcal{O}(\alpha^2),
\end{aligned}
$$

where the last step holds because $\sum_{j=0}^{k-1}\mathbb{E}[\|Y'_{kt}-Y'_{kt+k-1-j}\|_c^2] \in \mathcal{O}(k)$.

Combining pieces, we conclude that

$$
T_2 \leq \frac{5\alpha^2 u_{cm}^2\gamma^2}{\eta l_{cs}^2}\mathbb{E}[M_\eta(Y_t - Y'_{kt})] + o(\alpha),
$$

which completes the proof of Lemma 5.

## C.2 Step 2: General Stepsize

In this subsection, we aim to prove that there exists an $\alpha_0$ such that $\mathcal{L}(Y^{(\alpha)})$ is continuous in $\alpha$ when $\alpha \in (0, \alpha_0)$ with respect to $W_2$. Let us consider two stepsizes $\alpha > 0$ and $\alpha' > 0$. For simplicity, we will let $\{Y_t\}_{t\geq 0}$ and $\{Y'_t\}_{t\geq 0}$ denote the sequence associated with stepsize $\alpha$ and $\alpha'$, respectively. We couple the two sequences $\{Y_t\}_{t\geq 0}$ and $\{Y'_t\}_{t\geq 0}$ by letting them share the same noise $\{w'_t\}_{t\geq 0}$ :

$$
\begin{aligned}
Y_{t+1} &= (1-\alpha)Y_t + \sqrt{\alpha}\big(\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*) - \mathcal{T}(\theta^*) + w'_t\big), \\
Y'_{t+1} &= (1-\alpha')Y'_t + \sqrt{\alpha'}\big(\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*) - \mathcal{T}(\theta^*) + w'_t\big).
\end{aligned}
$$

Then, we obtain

$$
\begin{aligned}
Y_{t+1} - Y'_{t+1} &= (1-\alpha)(Y_t - Y'_t) + \sqrt{\alpha}(\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*) - \mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*)) \\
&\quad + \sqrt{\alpha}(\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*) - \mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*)) \\
&\quad + (\sqrt{\alpha}-\sqrt{\alpha'})(\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*) - \mathcal{T}(\theta^*)) + (\sqrt{\alpha}-\sqrt{\alpha'})w'_t \\
&:= (1-\alpha)(Y_t - Y'_t) + A.
\end{aligned}
$$

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both side of above equation and by property (1) in Proposition 5, we obtain

$$\mathbb{E}[M_\eta(Y_{t+1}-Y'_{t+1})] \leq (1-\alpha)^2\mathbb{E}[M_\eta(Y_t-Y'_t)] + (1-\alpha)\underbrace{\mathbb{E}\langle\nabla M(Y_t-Y'_t),A\rangle}_{T_1} + \underbrace{(1/(2\eta))\mathbb{E}\|A\|_2^2}_{T_2}.$$

Below we separately bound the $T_1$ and $T_2$ terms.

**Bounding the $T_1$ Term:** By property (4) in Proposition 5 and $w'_t$ being i.i.d zero mean noise and independent with $Y_t$ and $Y'_t$, we obtain

$$\begin{aligned}
T_1 \leq& \sqrt{\alpha}\mathbb{E}[\|Y_t-Y'_t\|_m\|\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*)\|_m] && (T_{11})\\
&+\sqrt{\alpha}\mathbb{E}[\|Y_t-Y'_t\|_m\|\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*)\|_m] && (T_{12})\\
&+(\sqrt{\alpha}-\sqrt{\alpha'})\mathbb{E}[\|Y_t-Y'_t\|_m\|\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*)-\mathcal{T}(\theta^*)\|_m]. && (T_{13})
\end{aligned}$$

Let $\delta = |\alpha-\alpha'| \leq \min((\frac{1}{\sqrt{\gamma}}-1)\alpha, \frac{1}{2}\alpha)$. Below, we bound $T_{11} \sim T_{13}$ separately, beginning with $T_{11}$:

$$\begin{aligned}
T_{11} &\leq \frac{\sqrt{\alpha}}{l_{cm}}\mathbb{E}[\|Y_t-Y'_t\|_m\|\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*)\|_c]\\
&\overset{(i)}{\leq} \frac{2\sqrt{\alpha\alpha'}u_{cm}\gamma}{l_{cm}}\mathbb{E}[M_\eta(Y_t-Y'_t)] \leq 2\sqrt{\alpha\alpha'}\sqrt{\gamma}\mathbb{E}[M_\eta(Y_t-Y'_t)] \leq 2\alpha\gamma^{\frac{1}{4}}\mathbb{E}[M_\eta(Y_t-Y'_t)],
\end{aligned}$$

where (i) holds because we can always choose a proper $\eta$ such that $\frac{u_{cm}}{l_{cm}} \leq \frac{1}{\sqrt{\gamma}}$. We next have

$$\begin{aligned}
T_{12} &\leq \frac{\sqrt{\alpha}|\sqrt{\alpha}-\sqrt{\alpha'}|u_{cm}\gamma}{l_{cm}}\mathbb{E}[\|Y_t-Y'_t\|_m\|Y_t\|_m]\\
&\in \mathcal{O}\left(\sqrt{\alpha}|\sqrt{\alpha}-\sqrt{\alpha'}|\right) \in \mathcal{O}\left(\frac{\sqrt{\alpha}\delta}{\sqrt{\alpha}+\sqrt{\alpha'}}\right) \in \mathcal{O}\left(\frac{\sqrt{\alpha}\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\right),
\end{aligned}$$

where we use $\mathbb{E}[\|Y_t-Y'_t\|_m\|Y_t\|_m] \in \mathcal{O}(1)$. Similarly, we have

$$\begin{aligned}
T_{13} &\leq \frac{\sqrt{\alpha'}|\sqrt{\alpha}-\sqrt{\alpha'}|u_{cm}\gamma}{l_{cm}}\mathbb{E}[\|Y_t-Y'_t\|_m\|Y'_t\|_m]\\
&\in \mathcal{O}\left(\sqrt{\alpha'}|\sqrt{\alpha}-\sqrt{\alpha'}|\right) \in \mathcal{O}\left(\frac{\sqrt{\alpha'}\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\right) \in \mathcal{O}\left(\frac{\sqrt{\alpha}\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\right),
\end{aligned}$$

where the last inequality holds because $\delta \leq \min((\frac{1}{\sqrt{\gamma}}-1)\alpha, \frac{1}{2}\alpha)$.

**Bounding the $T_2$ Term:** We next have

$$\begin{aligned}
T_2 \leq& \frac{2\alpha}{\eta}\mathbb{E}\|(\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*))\|_2^2 && (T_{21})\\
&+\frac{2\alpha}{\eta}\mathbb{E}\|(\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*))\|_2^2 && (T_{22})\\
&+\frac{2(\sqrt{\alpha}-\sqrt{\alpha'})^2}{\eta}\mathbb{E}\|(\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*)-\mathcal{T}(\theta^*))\|_2^2 && (T_{23})\\
&+\frac{2(\sqrt{\alpha}-\sqrt{\alpha'})^2}{\eta}\mathbb{E}[\|w'_t\|_2^2]. && (T_{24})
\end{aligned}$$

Below, we bound $T_{21} \sim T_{24}$ separately. We begin with

$$T_{21} \leq \frac{2\alpha}{l_{cs}^2}\mathbb{E}\|(\mathcal{T}(\sqrt{\alpha'}Y_t+\theta^*)-\mathcal{T}(\sqrt{\alpha'}Y'_t+\theta^*))\|_c^2 \leq \frac{4\alpha\alpha'\gamma^2u_{cm}^2}{l_{cs}^2}\mathbb{E}[M_\eta(Y_t-Y'_t)] \leq \frac{6\alpha^2\gamma^2u_{cm}^2}{l_{cs}^2}\mathbb{E}[M_\eta(Y_t-Y'_t)],$$

where the last inequality holds because $\delta \le \min((\frac{1}{\sqrt{\gamma}}-1)\alpha, \frac{1}{2}\alpha)$. The next three terms satisfy

$$T_{22} \in \mathcal{O}\left(\alpha(\sqrt{\alpha}-\sqrt{\alpha'})^2\right) \in \mathcal{O}\left(\frac{\alpha\delta^2}{(\sqrt{\alpha}+\sqrt{\alpha'})^2}\right) \in \mathcal{O}\left(\frac{\alpha\delta^2}{\min(\alpha,\alpha')}\right).$$

$$T_{23} \in \mathcal{O}\left(\alpha'(\sqrt{\alpha}-\sqrt{\alpha'})^2\right) \in \mathcal{O}\left(\frac{\alpha'\delta^2}{\min(\alpha,\alpha')}\right) \in \mathcal{O}\left(\frac{\alpha\delta^2}{\min(\alpha,\alpha')}\right).$$

$$T_{24} \in \mathcal{O}\left(\frac{\delta^2}{\min(\alpha,\alpha')}\right).$$

Combining the above bounds for $T_1$ and $T_2$ and using the fact that there exist an $\alpha_0$ such that $0 < \left(1-2(1-\gamma^{\frac{1}{4}})\alpha_0+\mathcal{O}(\alpha_0^2)\right) < 1$, we see that for any $\alpha \le \alpha_0$, there exist $t_\alpha$ such that for any $t \ge t_\alpha$, we obtain

$$\mathbb{E}[M_\eta(Y_{t+1}-Y'_{t+1})] \le \left(1-2(1-\gamma^{\frac{1}{4}})\alpha+\mathcal{O}(\alpha^2)\right)\mathbb{E}[M_\eta(Y_t-Y'_t)] + \mathcal{O}\left(\frac{\sqrt{\alpha}\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\right).$$

Then, we obtain

$$\lim_{t\to\infty}\mathbb{E}[M_\eta(Y_t-Y'_t)] \in \mathcal{O}\left(\frac{\delta}{\min(\alpha,\alpha')}\right).$$

Hence,

$$\begin{aligned} W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y^{(\alpha')})\right) &\le \lim_{t\to\infty} W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y_t)\right) + W_2\left(\mathcal{L}(Y_t),\mathcal{L}(Y'_t)\right) + W_2\left(\mathcal{L}(Y_t),\mathcal{L}(Y^{(\alpha')})\right) \\ &\le \lim_{t\to\infty}\sqrt{\mathbb{E}[\|Y_t-Y'_t\|_c^2]} \le \lim_{t\to\infty}\sqrt{2u_{cm}^2\mathbb{E}[M_\eta(Y_t-Y'_t)]} \le \frac{c\sqrt{\delta}}{\min(\alpha,\alpha')^{\frac{1}{2}}}, \end{aligned}$$

where $c$ is a universal constant that is independent with $\alpha, \alpha'$.

Then, given $\alpha > 0$, for $\forall\epsilon > 0$, we can choose a sufficient small $\delta_\epsilon$ such that

$$\frac{c\sqrt{\delta_\epsilon}}{(\alpha-\delta_\epsilon)^{\frac{1}{2}}} \le \epsilon \quad \text{and} \quad 0 < \delta_\epsilon < \min\left(\left(\frac{1}{\sqrt{\gamma}}-1\right)\alpha, \frac{1}{2}\alpha\right).$$

Then, when $\alpha'$ is selected with $|\alpha-\alpha'| \le \delta_\epsilon$, we obtain

$$W_2\big(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y^{(\alpha')})\big) \le \epsilon.$$

Therefore, we complete the proof of continuity of $\mathcal{L}(Y^{(\alpha)})$ with respect to $W_2$.

Recall that

$$\lim_{\alpha\to 0,\alpha\in\mathbb{Q}^+} W_2\big(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\big) = 0.$$

Thus, for $\forall\epsilon > 0$, there exist $\delta > 0$, such that for all rational $\alpha \le \delta$, $W_2\big(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\big) \le \frac{\epsilon}{2}$.

Given arbitrary real number r such that $0 < r \le \frac{\delta}{2}$, there exist $q(r) \in \mathbb{Q}$ such that $|r-q(r)| \le \frac{\delta}{2}$ and $W_2\big(\mathcal{L}(Y^{(r)}),\mathcal{L}(Y^{(q(r))})\big) \le \frac{\epsilon}{2}$ by Section C.2. Then,

$$W_2\big(\mathcal{L}(Y^{(r)}),\mathcal{L}(Y)\big) \le W_2\big(\mathcal{L}(Y^{(r)}),\mathcal{L}(Y^{(q(r))})\big) + W_2\big(\mathcal{L}(Y^{(q(r))}),\mathcal{L}(Y)\big) \le \epsilon,$$

where the second inequality holds because $q(r) \le \frac{\delta}{2}+\frac{\delta}{2} = \delta$. We conclude that there exist a unique limit $\mathcal{L}(Y)$ such that

$$\lim_{\alpha\to 0} W_2\big(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\big) = 0.$$

This completes the second step of the proof of Theorem 2.

## C.3 Step 2.5: Convergence Rate under Gaussian Noise

By triangle inequality, we obtain the desired convergence rate:

$$
\begin{aligned}
W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\right) &\le W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y^{(\alpha/k)})\right)+W_2\left(\mathcal{L}(Y^{(\alpha/k)}),\mathcal{L}(Y)\right)\\
&\le o(1)+W_2\left(\mathcal{L}(Y^{(\alpha/k)}),\mathcal{L}(Y)\right)\\
&\le \lim_{k\to\infty} o(1)+W_2\left(\mathcal{L}(Y^{(\alpha/k)}),\mathcal{L}(Y)\right)\in o(1).
\end{aligned}
\tag{48}
$$

## C.4 Step 3: General Noise

By Section C.1, C.2 and C.3, we prove that under the noise with Gaussian distribution, there exist a unique random variable $Y$ such that $Y^{(\alpha)}$ converge to $Y$ with respect to $W_2$. In this subsection, we aim to prove that under general i.i.d zero mean noise with the same variance, the convergence result still holds and the limit is still $Y$.

Fix the stepsize $\alpha>0$. We consider two sequences $\{Y_t^{(\alpha)}\}_{t\ge0}$ and $\{Y_t'^{(\alpha)}\}_{t\ge0}$, where $\{Y_t^{(\alpha)}\}_{t\ge0}$ is associated with general noise $\{w_t\}_{t\ge0}$, and $\{Y_t'^{(\alpha)}\}_{t\ge0}$ is associated with Gaussian distributed noise $\{w_t'\}_{t\ge0}$. When the context is clear, we drop the supperscript $^{(\alpha)}$ for the ease of exposition. We will couple $\{Y_t\}_{t\ge0}$ and $\{Y_t'\}_{t\ge0}$ as follows:

$$
\begin{aligned}
Y_{t+1}&=(1-\alpha)Y_t+\sqrt{\alpha}\left(\mathcal{T}(\sqrt{\alpha}Y_t+\theta^*)-\mathcal{T}(\theta^*)+w_t\right),\\
Y_{t+1}'&=(1-\alpha)Y_t'+\sqrt{\alpha}\left(\mathcal{T}(\sqrt{\alpha}Y_t'+\theta^*)-\mathcal{T}(\theta^*)+w_t'\right),
\end{aligned}
\tag{49}
$$

where $w_t,w_t'$ have zero mean and the same variance. Here $w_t$ and $w_t'$ are not necessarily independent of each other, and we assume that $w_t$ has finite fourth moment. The specific coupling between $\{w_t'\}_{t\ge0}$ and $\{w_t\}_{t\ge0}$ will be specified later.

Let $\kappa=\lfloor\alpha^{-\frac12}\rfloor$. Direct calculation gives

$$
\begin{aligned}
Y_{\kappa t+\kappa}=&(1-\alpha)^\kappa Y_{\kappa t}+\sqrt{\alpha}\kappa(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}+\theta^*)-\mathcal{T}(\theta^*))\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}+\theta^*)\right)\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}+\theta^*)-\mathcal{T}(\theta^*)\right)+\sqrt{\alpha}\sum_{j=1}^{\kappa}(1-\alpha)^{j-1}w_{\kappa t+\kappa-j}.
\end{aligned}
$$

and

$$
\begin{aligned}
Y_{\kappa t+\kappa}'=&(1-\alpha)^\kappa Y_{\kappa t}'+\sqrt{\alpha}\kappa(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}'+\theta^*)-\mathcal{T}(\theta^*))\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}'+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}'+\theta^*)\right)\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}'+\theta^*)-\mathcal{T}(\theta^*)\right)+\sqrt{\alpha}\sum_{j=1}^{\kappa}(1-\alpha)^{j-1}w_{\kappa t+\kappa-j}'.
\end{aligned}
$$

Taking the difference of the last two equations, we get

$$
\begin{aligned}
Y_{\kappa t+\kappa}-Y_{\kappa t+\kappa}'=&(1-\alpha)^\kappa(Y_{\kappa t}-Y_{\kappa t}')+\sqrt{\alpha}\kappa(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}'+\theta^*))\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}'+\theta^*)+\mathcal{T}(\sqrt{\alpha}Y_{\kappa t}'+\theta^*)\right)\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}+\theta^*)-\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}'+\theta^*)\right)\\
&+\sqrt{\alpha}\sum_{j=1}^{\kappa}(w_{\kappa t+\kappa-j}-w_{\kappa t+\kappa-j}')+\sqrt{\alpha}\sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)(w_{\kappa t+\kappa-j}-w_{\kappa t+\kappa-j}')\\
:=&(1-\alpha)^\kappa(Y_{\kappa t}-Y_{\kappa t}')+A,
\end{aligned}
$$

where we collect in $A$ all but the first term on the RHS. Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both side of above equation and by property (1) in Proposition 5, we obtain

$$\mathbb{E}[M(Y_{\kappa t+\kappa} - Y'_{\kappa t+\kappa})] \leq (1-\alpha)^{2\kappa}\mathbb{E}[M(Y_{\kappa t} - Y'_{\kappa t})] + (1-\alpha)^{\kappa}\underbrace{\mathbb{E}\langle \nabla M(Y_{\kappa t} - Y'_{\kappa t}), A\rangle}_{T_1} + \frac{1}{2\eta}\underbrace{\mathbb{E}\|A\|_2^2}_{T_2}. \tag{50}$$

The following lemmas, proved in Sections C.4.1 and C.4.2 to follow, control the $T_1$ and $T_2$ terms above.

**Lemma 6.** *Under the setting of Theorem 2, we have*

$$\mathbb{E}[T_1] \leq 2\alpha\kappa\sqrt{\gamma}\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] + \mathcal{O}(\alpha^{\frac{3}{2}}\kappa^{\frac{3}{2}}).$$

**Lemma 7.** *Under the setting of Theorem 2 and some proper couplings between $\{w_t\}_{t\geq 0}$ and $\{w'_t\}_{t\geq 0}$, we have*

$$\mathbb{E}[T_2] \leq \frac{10\alpha^2\kappa^2\gamma^2 u_{cm}^2}{l_{cs}^2}\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] + \mathcal{O}(\alpha).$$

Plugging the above bounds for $T_1$ and $T_2$ into equation (50), there exist an $\alpha_0$ such that for any $\alpha \leq \alpha_0$, there exist $t_\alpha$ such that for any $t \geq t_\alpha$, we obtain

$$\begin{aligned}\mathbb{E}[M_\eta(Y_{\kappa t+\kappa} - Y'_{\kappa t+\kappa})] &\leq \Big((1-\alpha)^\kappa + 2\alpha\kappa\sqrt{\gamma} + \frac{10\alpha^2\kappa^2\gamma^2 u_{cm}^2}{l_{cs}^2}\Big)\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] + \mathcal{O}(\alpha)\\ &\leq (1-(1-\sqrt{\gamma})\alpha\kappa)\,\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] + \mathcal{O}(\alpha).\end{aligned}$$

Therefore, we obtain

$$\lim_{t\to\infty}\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] \in \mathcal{O}(\alpha^{\frac{1}{2}}).$$

By triangle inequality, we have

$$\begin{aligned}W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y'^{(\alpha)})\big) &\leq \lim_{t\to\infty} W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y_{\kappa t})\big) + W_2\big(\mathcal{L}(Y_{\kappa t}), \mathcal{L}(Y'_{\kappa t})\big) + W_2\big(\mathcal{L}(Y'_{\kappa t}), \mathcal{L}(Y'^{(\alpha)})\big)\\ &\leq \lim_{t\to\infty}\sqrt{\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_c^2]} \leq \lim_{t\to\infty}\sqrt{2u_{cm}^2\mathbb{E}[M(Y_{\kappa t} - Y'_{\kappa t})]} \in \mathcal{O}(\alpha^{\frac{1}{4}}).\end{aligned}$$

Therefore, by equation (48), we obtain

$$W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\big) \leq W_2\big(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}((Y')^{(\alpha)})\big) + W_2\big(\mathcal{L}(Y'^{(\alpha)}), \mathcal{L}(Y)\big) \in o(1),$$

which implies

$$\lim_{\alpha\to 0} W_2(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)) = 0.$$

This completes the last step of the proof of Theorem 2. We have proved Theorem 2.

### C.4.1 Proof of Lemma 6 on $T_1$

By property (4) in Proposition 5 and $w_{\kappa t+\kappa-j}$ and $w'_{\kappa t+\kappa-j}$ being zero mean noise and independent with $Y_{\kappa t}$ and $Y'_{\kappa t}$, we obtain

$$\begin{aligned}T_1 \leq &\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_\kappa\|\sqrt{\alpha}\kappa(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t} + \theta^*) - \mathcal{T}(\sqrt{\alpha}Y'_{\kappa t} + \theta^*))\|_m] &(T_{11})\\ &+ \mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_\kappa\|\sqrt{\alpha}\sum_{j=1}^{\kappa}\big(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j} + \theta^*) - \mathcal{T}(\sqrt{\alpha}Y_{\kappa t} + \theta^*)\big)\|_m] &(T_{12})\\ &+ \mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|\sqrt{\alpha}\sum_{j=1}^{\kappa}\big(\mathcal{T}(\sqrt{\alpha}Y'_{\kappa t+\kappa-j} + \theta^*) - \mathcal{T}(\sqrt{\alpha}Y'_{\kappa t} + \theta^*)\big)\|_m] &(T_{13})\\ &+ \mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|\sqrt{\alpha}\sum_{j=1}^{\kappa}((1-\alpha)^{j-1} - 1)\big(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j} + \theta^*) - \mathcal{T}(\sqrt{\alpha}Y'_{\kappa t+\kappa-j} + \theta^*)\big)\|_m]. &(T_{14})\end{aligned}$$

Below, we bound terms $T_{11} \sim T_{14}$ separately.

$$T_{11} \leq \frac{2\alpha\kappa\gamma u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] \leq 2\alpha\kappa\sqrt{\gamma}\mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})],$$

where the last inequality holds because we can always choose a proper $\eta$ such that $\frac{u_{cm}}{l_{cm}} \leq \frac{1}{\sqrt{\gamma}}$.

$$T_{12} \leq \frac{\alpha\gamma}{l_{cm}}\sum_{j=1}^{\kappa}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|Y_{\kappa t+\kappa-j} - Y_{\kappa t}\|_c].$$

By equation (49), we obtain

$$\begin{aligned}
&\|Y_{\kappa t+\kappa-j} - Y_{\kappa t}\|_c \\
=&\left\|((1-\alpha)^{\kappa-j}-1)Y_{\kappa t} + \sqrt{\alpha}\sum_{l=1}^{\kappa-j}(1-\alpha)^{l-1}\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j-l}+\theta^*) - \mathcal{T}(\theta^*) + w_{\kappa t+\kappa-j-l}\right)\right\|_c.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
T_{12} \leq &\frac{\alpha\gamma}{l_{cm}}\sum_{j=1}^{\kappa}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|((1-\alpha)^{\kappa-j}-1)Y_{\kappa t}\|_c] && (T_{121})\\
&+\frac{\alpha\gamma}{l_{cm}}\sum_{j=1}^{\kappa}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|\sqrt{\alpha}\sum_{l=1}^{\kappa-j}(1-\alpha)^{l-1}\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j-l}+\theta^*) - \mathcal{T}(\theta^*)\right)\|_c] && (T_{122})\\
&+\frac{\alpha\gamma}{l_{cm}}\sum_{j=1}^{\kappa}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|\sqrt{\alpha}\sum_{l=1}^{\kappa-j}(1-\alpha)^{l-1}w_{\kappa t+\kappa-j-l}\|_c]. && (T_{123})
\end{aligned}$$

Observe that

$$\begin{aligned}
T_{121} &\leq \frac{\alpha\gamma u_{cm}}{l_{cm}}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|Y_{\kappa t}\|_m]\sum_{j=1}^{\kappa}(1-(1-\alpha)^{\kappa-j})\\
&\leq \mathcal{O}(1)\cdot\alpha\sum_{j=1}^{\kappa}(1-(1-\alpha)^{\kappa-j}) \leq \mathcal{O}(1)\cdot(\alpha\kappa - 1 + (1-\alpha)^{\kappa})\\
&\leq \mathcal{O}(1)\cdot(\alpha\kappa - 1 + (1-\frac{\alpha\kappa}{\kappa})^{\kappa}) \overset{\text{(i)}}{\in} \mathcal{O}(\alpha^2\kappa^2),
\end{aligned}$$

where (i) holds by equation (43). We also have

$$\begin{aligned}
T_{122} &\leq \frac{\alpha\gamma}{l_{cm}}\sum_{j=1}^{\kappa}\sum_{l=1}^{\kappa-j}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|\sqrt{\alpha}(1-\alpha)^{l-1}\left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j-l}+\theta^*) - \mathcal{T}(\theta^*)\right)\|_c]\\
&\leq \frac{\alpha\gamma^2}{l_{cm}}\sum_{j=1}^{\kappa}\sum_{l=1}^{\kappa-j}\alpha(1-\alpha)^{l-1}\mathbb{E}[\|Y_{\kappa t} - Y'_{\kappa t}\|_m\|Y_{\kappa t+\kappa-j-l}\|_c]\\
&\leq \frac{\alpha^2\gamma^2}{l_{cm}}\sum_{j=1}^{\kappa}\sum_{l=1}^{\kappa-j}(1-\alpha)^{l-1}(\sqrt{\mathbb{E}[\|Y_{\kappa t}\|_m^2]\mathbb{E}[\|Y_{\kappa t+\kappa-j-l}\|_c^2]} + \sqrt{\mathbb{E}[\|Y'_{\kappa t}\|_\kappa^2]\mathbb{E}[\|Y_{\kappa t+\kappa-j-l}\|_c^2]})\\
&\overset{\text{(i)}}{\leq} \mathcal{O}(1)\cdot\alpha^2\sum_{j=1}^{\kappa}\sum_{l=1}^{\kappa-j}(1-\alpha)^{l-1} \leq \mathcal{O}(1)\cdot\alpha^2\kappa^2 \in \mathcal{O}(\alpha^2\kappa^2),
\end{aligned}$$

where (i) follows as $\mathbb{E}[\|Y_t'\|_m^2] \in \mathcal{O}(1)$ and $\mathbb{E}[\|Y_t'\|_m^2] \in \mathcal{O}(1)$ for all $t \geq 0$. Lastly, for term $T_{123}$, we have

$$\begin{aligned}
T_{123} &\overset{\text{(i)}}{=} \frac{\alpha\gamma}{l_{cm}} \sum_{j=1}^{\kappa} \mathbb{E}[\|Y_{\kappa t} - Y_{\kappa t}'\|_m] \mathbb{E}[\|\sqrt{\alpha} \sum_{l=1}^{\kappa-j} (1-\alpha)^{l-1} w_{\kappa t+\kappa-j-l}\|_c] \\
&\overset{\text{(ii)}}{\leq} \mathcal{O}(1) \cdot \alpha^{\frac{3}{2}} \sum_{j=1}^{\kappa} \mathbb{E}[\| \sum_{l=1}^{\kappa-j} (1-\alpha)^{l-1} w_{\kappa t+\kappa-j-l}\|_2] \\
&\leq \mathcal{O}(1) \cdot \alpha^{\frac{3}{2}} \sum_{j=1}^{\kappa} \sqrt{\mathbb{E}[\| \sum_{l=1}^{\kappa-j} (1-\alpha)^{l-1} w_{\kappa t+\kappa-j-l}\|_2^2]} \\
&\leq \mathcal{O}(1) \cdot \alpha^{\frac{3}{2}} \sum_{j=1}^{\kappa} \sqrt{\sum_{l=1}^{\kappa-j} (1-\alpha)^{2l-2}} \leq \mathcal{O}(1) \cdot \alpha^{\frac{3}{2}} \sum_{j=1}^{\kappa} \sqrt{\kappa - j} \overset{\text{(iii)}}{=} \mathcal{O}(\alpha^{\frac{3}{2}} \kappa^{\frac{3}{2}}),
\end{aligned}$$

where (i) holds because $Y_{\kappa t}$ and $Y_{\kappa t}'$ are independent with $w_{\kappa t+\kappa-j-l}$, (ii) follows as $\mathbb{E}[\|Y_{\kappa t} - Y_{\kappa t}'\|_m] \in \mathcal{O}(1)$, and (iii) holds because $\sum_{j=1}^{\kappa} \sqrt{\kappa - j} \in \mathcal{O}(\kappa^{\frac{3}{2}})$.

Combining the bound of $T_{121}, T_{122}$ and $T_{123}$ together, we obtain $T_{12} \in \mathcal{O}(\alpha^{\frac{3}{2}} \kappa^{\frac{3}{2}})$. Similarly, we have $T_{13} \in \mathcal{O}(\alpha^{\frac{3}{2}} \kappa^{\frac{3}{2}})$. For $T_{14}$, we have

$$\begin{aligned}
T_{14} &\leq \frac{\alpha}{l_{cm}} \sum_{j=1}^{\kappa} ((1-\alpha)^{j-1} - 1) \mathbb{E}[\|Y_{\kappa t} - Y_{\kappa t}'\|_m \|Y_{\kappa t+\kappa-j} - Y_{\kappa t+\kappa-j}'\|_c] \\
&\overset{\text{(i)}}{\leq} \mathcal{O}(1) \cdot \alpha \sum_{j=1}^{\kappa} (1 - (1-\alpha)^{j-1}) \leq \mathcal{O}(1) \cdot (1 - \alpha\kappa - (1-\alpha)^{\kappa}) \in \mathcal{O}(\alpha^2 \kappa^2),
\end{aligned}$$

where in (i) we use $\mathbb{E}[\|Y_{\kappa t} - Y_{\kappa t}'\|_m \|Y_{\kappa t+\kappa-j} - Y_{\kappa t+\kappa-j}'\|_c] \in \mathcal{O}(1)$.

Combining the bound for $T_{11} \sim T_{14}$ together, we obtain

$$T_1 \leq 2\alpha\kappa\sqrt{\gamma}\mathbb{E}[M_\eta(Y_{\kappa t} - Y_{\kappa t}')] + \mathcal{O}(\alpha^{\frac{3}{2}} \kappa^{\frac{3}{2}}),$$

thereby completing the proof of Lemma 6.

#### C.4.2 Proof of Lemma 7 on $T_2$

$$\begin{aligned}
T_2 \leq\, & 5\alpha\kappa^2 \mathbb{E}[\|\mathcal{T}(\sqrt{\alpha} Y_{\kappa t} + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y_{\kappa t}' + \theta^*)\|_2^2] && (T_{21}) \\
&+ 5\alpha \mathbb{E}[\| \sum_{j=1}^{\kappa} \left( \mathcal{T}(\sqrt{\alpha} Y_{\kappa t+\kappa-j} + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y_{\kappa t} + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y_{\kappa t+\kappa-j}' + \theta^*) + \mathcal{T}(\sqrt{\alpha} Y_{\kappa t}' + \theta^*) \right) \|_2^2] \\
& && (T_{22}) \\
&+ 5\alpha \mathbb{E}[\| \sum_{j=1}^{\kappa} ((1-\alpha)^{j-1} - 1) \left( \mathcal{T}(\sqrt{\alpha} Y_{\kappa t+\kappa-j} + \theta^*) - \mathcal{T}(\sqrt{\alpha} Y_{\kappa t+\kappa-j}' + \theta^*) \right) \|_2^2] && (T_{23}) \\
&+ 5\alpha \mathbb{E}[\| \sum_{j=1}^{\kappa} (w_{\kappa t+\kappa-j} - w_{\kappa t+\kappa-j}') \|_2^2] && (T_{24}) \\
&+ 5\alpha \mathbb{E}[\| \sum_{j=1}^{\kappa} ((1-\alpha)^{j-1} - 1)(w_{\kappa t+\kappa-j} - w_{\kappa t+\kappa-j}') \|_2^2]. && (T_{25})
\end{aligned}$$

Below, we bound $T_{21} \sim T_{25}$ separately. For $T_{21}$:

$$T_{21} \leq \frac{5\alpha^2\kappa^2\gamma^2}{l_{cm}^2} \mathbb{E}[\|Y_{\kappa t} - Y_{\kappa t}'\|_c^2] \leq \frac{10\alpha^2\kappa^2\gamma^2 u_{cm}^2}{l_{cs}^2} \mathbb{E}[M_\eta(Y_{\kappa t} - Y_{\kappa t}')].$$

Next, for $T_{22}$, we have

$$\begin{aligned}
T_{22} &\leq 5\alpha\kappa \sum_{j=1}^{\kappa} \mathbb{E}[\| \left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}+\theta^*) - \mathcal{T}(\sqrt{\alpha}Y_{\kappa t}+\theta^*) - \mathcal{T}(\sqrt{\alpha}Y'_{\kappa t+\kappa-j}+\theta^*) + \mathcal{T}(\sqrt{\alpha}Y'_{\kappa t}+\theta^*)\right) \|_2^2] \\
&\leq \frac{10\alpha^2\kappa\gamma^2}{l_{cs}^2} \sum_{j=1}^{\kappa} (\mathbb{E}[\|Y_{\kappa t+\kappa-j} - Y_{\kappa t}\|_c^2] + \mathbb{E}[\|Y'_{\kappa t+\kappa-j} - Y'_{\kappa t}\|_c^2]) \in \mathcal{O}(\alpha^2\kappa^2).
\end{aligned}$$

Continuing, we have

$$\begin{aligned}
T_{23} &\leq 5\alpha\kappa \sum_{j=1}^{\kappa} \mathbb{E}[\|((1-\alpha)^{j-1}-1) \left(\mathcal{T}(\sqrt{\alpha}Y_{\kappa t+\kappa-j}+\theta^*) - \mathcal{T}(\sqrt{\alpha}Y'_{\kappa t+\kappa-j}+\theta^*)\right) \|_2^2] \\
&\leq \frac{10\alpha^2\gamma^2\kappa}{l_{cs}^2} \sum_{j=1}^{\kappa} \mathbb{E}[\|((1-\alpha)^{j-1}-1) \left(Y_{\kappa t+\kappa-j} - Y'_{\kappa t+\kappa-j}\right) \|_c^2] \\
&\leq \mathcal{O}(\alpha^2\kappa) \cdot \sum_{j=1}^{\kappa} ((1-\alpha)^{j-1}-1) \in \mathcal{O}(\alpha^3\kappa^3).
\end{aligned}$$

Lastly, we have

$$T_{24} = 5\alpha\kappa\mathbb{E}[\|\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w_{\kappa t+\kappa-j} - \frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w'_{\kappa t+\kappa-j})\|_2^2].$$

We restate Theorem 1 in [Bon20] in the following lemma.

**Lemma 8** (Theorem 1 in [Bon20])**.** *Let $X_1, \ldots, X_n$ be $n$ i.i.d random variables taking values in $\mathbb{R}^d$ with zero mean and identity variance matrix. Let $\nu$ be the d-dimensional standard Gaussian measure and $X'_1, \ldots, X'_n$ be $n$ i.i.d random variables distributed as $\nu$. Assume that $\mathbb{E}[\|X_1\|_2^4] < \infty$. Let $S_n = \frac{X_1+\cdots+X_n}{\sqrt{n}}$ and $S'_n = \frac{X'_1+\cdots+X'_n}{\sqrt{n}}$ Then, we have*

$$W_{2,2}(\mathcal{L}(S_n), \mathcal{L}(S'_n)) = W_{2,2}(\mathcal{L}(S_n), \nu) \in \mathcal{O}(\frac{1}{\sqrt{n}}),$$

*where $W_{2,2}$ denotes the Wasserstein distances of order 2 with $\ell_2$-norm.*

We can always choose a coupling between $w_t$ and $w'_t$ such that

$$\mathbb{E}[\|\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w_{\kappa t+\kappa-j} - \frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w'_{\kappa t+\kappa-j})\|_2^2 = W_{2,2}^2(\mathcal{L}(\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w_{\kappa t+\kappa-j}), \mathcal{L}(\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w'_{\kappa t+\kappa-j})).$$

Let $C = \mathbb{E}[w_1 w_1^T]$. Because $C$ is positive semidefinite, by [HJ12, Theorem 7.2.6], there always exists a symmetric matrix $C^{\frac{1}{2}}$ such that $C = C^{\frac{1}{2}}C^{\frac{1}{2}}$. Then, by Lemma 8, we obtain

$$\begin{aligned}
T_{24} &= 5\alpha\kappa W_{2,2}^2(\mathcal{L}(\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w_{\kappa t+\kappa-j}), \mathcal{L}(\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w'_{\kappa t+\kappa-j})) \\
&= 5\alpha\kappa \cdot \inf \left( \mathbb{E}[\|\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w_{\kappa t+\kappa-j} - \frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} w'_{\kappa t+\kappa-j})\|_2^2 \right) \\
&= 5\alpha\kappa \cdot \inf \left( \mathbb{E}[\|\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} C^{\frac{1}{2}}X_j - \frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} C^{\frac{1}{2}}X'_j)\|_2^2 \right) \\
&\leq 5\alpha\kappa\|C^{\frac{1}{2}}\|_2^2 \cdot \inf \left( \mathbb{E}[\|\frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} X_j - \frac{1}{\sqrt{\kappa}}\sum_{j=1}^{\kappa} X'_j)\|_2^2 \right) \\
&= 5\alpha\kappa\|C^{\frac{1}{2}}\|_2^2 W_{2,2}^2(\mathcal{L}(S_\kappa), \mathcal{L}(S'_\kappa)) \in \mathcal{O}(\alpha).
\end{aligned}$$

where all the infimums are took by considering all the joint distributions with the same marginal distribution.

$$
\begin{aligned}
T_{25} &= 5\alpha \sum_{j=1}^{\kappa} \mathbb{E}[\|((1-\alpha)^{j-1}-1)(w_{\kappa t+\kappa-j} - w'_{\kappa t+\kappa-j})\|_2^2] \\
&\le \mathcal{O}(\alpha) \cdot \sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)^2 \le \mathcal{O}(\alpha) \cdot \sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1) \in \mathcal{O}(\alpha^2\kappa^2).
\end{aligned}
$$

Recall that $\kappa = \lfloor \alpha^{-\frac{1}{2}} \rfloor$, we obtain

$$
T_2 \le \frac{10\alpha^2\kappa^2\gamma^2 u_{cm}^2}{l_{cs}^2} \mathbb{E}[M_\eta(Y_{\kappa t} - Y'_{\kappa t})] + \mathcal{O}(\alpha),
$$

thereby completing the proof of Lemma 7

# D Proof of Theorem 3

By equation (37) and Theorem 1, we obtain the following equation in distribution:

$$
Y^{(\alpha)} \overset{\mathrm{d}}{=} (1-\alpha)Y^{(\alpha)} + \sqrt{\alpha}\left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*) - \mathcal{T}(\theta^*) + w\right).
$$

After taking expectation to both sides of the above equation, we obtain

$$
\begin{aligned}
\mathbb{E}[Y^{(\alpha)}] =& \frac{1}{\sqrt{\alpha}}\mathbb{E}[\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*) - \mathcal{T}(\theta^*)] \\
=& \underbrace{\frac{1}{\sqrt{\alpha}}\mathbb{E}[(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*) - \mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))]}_{T_1} \\
&+ \underbrace{\frac{1}{\sqrt{\alpha}}\mathbb{E}[(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*) - \mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon))]}_{T_2}.
\end{aligned}
\tag{51}
$$

By Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
\|T_1\|_c &\le \frac{1}{\sqrt{\alpha}}\mathbb{E}[\|\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*) - \mathcal{T}(\theta^*)\|_c \mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))] \\
&\le \frac{1}{\sqrt{\alpha}}\sqrt{\mathbb{E}[\|\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*) - \mathcal{T}(\theta^*)\|_c^2]}\sqrt{\mathbb{P}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))} \\
&\le \gamma\sqrt{\mathbb{E}[\|Y^{(\alpha)}\|_c^2]}\sqrt{\mathbb{P}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))} \\
&\le \gamma\sqrt{\mathbb{E}[\|Y^{(\alpha)}\|_c^2]}\sqrt{\mathbb{P}(\|Y^{(\alpha)}\|_2^2 \ge \frac{\epsilon^2}{\sqrt{\alpha}})} \\
&\le \gamma\sqrt{\mathbb{E}[\|Y^{(\alpha)}\|_c^2]}\sqrt{\frac{\sqrt{\alpha}\mathbb{E}\|Y^{(\alpha)}\|_2^2}{\epsilon^2}} \overset{\text{(i)}}{\in} \mathcal{O}(\alpha^{\frac{1}{4}}),
\end{aligned}
$$

where (i) holds because the equivalence of all norms in $\mathbb{R}^d$, Fatou's lemma [Dur19, Exercise 3.2.4] and Corollary 1(1). Therefore, we obtain $\lim_{\alpha\to 0} T_1 = 0$.

Below, we discuss two cases.

**Case 1:** If $g(\cdot)$ is smooth, because $F(\cdot)$ is also smooth, we conclude that $\mathcal{T}(\cdot)$ is smooth in $B^d(\theta^*,\epsilon)$ by chain rule. Therefore, we obtain

$$
\begin{aligned}
T_2 =& \frac{1}{\sqrt{\alpha}}\mathbb{E}[(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]\\
=& \mathbb{E}[\nabla\mathcal{T}(\lambda_\alpha\sqrt{\alpha}Y^{(\alpha)}+\theta^*)Y^{(\alpha)}\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]\\
=& \underbrace{\mathbb{E}[(\nabla\mathcal{T}(\lambda_\alpha\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\nabla\mathcal{T}(\theta^*))Y^{(\alpha)}\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]}_{T_{21}}\\
&-\underbrace{\nabla\mathcal{T}(\theta^*)\mathbb{E}[Y^{(\alpha)}\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\notin B^d(0,\epsilon))]}_{T_{22}}+\nabla\mathcal{T}(\theta^*)\mathbb{E}[Y^{(\alpha)}].
\end{aligned}
$$

By Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
\|T_{21}\|_c &\leq \mathbb{E}[\|(\nabla\mathcal{T}(\lambda_\alpha\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\nabla\mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))\|_c\|Y^{(\alpha)}\|_c]\\
&\leq \underbrace{\sqrt{\mathbb{E}[\|(\nabla\mathcal{T}(\lambda_\alpha\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\nabla\mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))\|_c^2]}}_{\in o(1)}\underbrace{\sqrt{\mathbb{E}[\|Y^{(\alpha)}\|_c^2]}}_{\in\mathcal{O}(1)}\\
&\overset{\text{(i)}}{\in} o(1),
\end{aligned}
$$

where (i) holds because $\mathcal{T}(\cdot)$ is smooth in $B^d(\theta^*,\epsilon)$.

$$
\|T_{22}\|_c \leq \|\nabla\mathcal{T}(\theta^*)\|_c\underbrace{\sqrt{\mathbb{E}[\|Y^{(\alpha)}\|_c^2]}}_{\in\mathcal{O}(1)}\underbrace{\sqrt{\mathbb{P}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\notin B^d(0,\epsilon))}}_{\in\mathcal{O}(\alpha^{\frac{1}{4}})}\in\mathcal{O}(\alpha^{\frac{1}{4}}).
$$

Therefore, we obtain

$$
\lim_{\alpha\to 0}T_{21}=\lim_{\alpha\to 0}T_{22}=0.
$$

Taking $\alpha\to 0$ to both sides of equation (51), we obtain

$$
\mathbb{E}[Y]=\nabla\mathcal{T}(\theta^*)\mathbb{E}[Y].
$$

If $\mathbb{E}[Y]\neq 0$, let $y=\mathbb{E}[Y], y\neq 0$. Let $\bar{y}_{\epsilon_1}=\frac{\epsilon_1\epsilon y}{\|y\|_2}$, where $\epsilon_1<1$. Then, we have $\bar{y}_{\epsilon_1}\in B^d(0,\epsilon)$. Therefore, we obtain

$$
\begin{aligned}
\|\mathcal{T}(\bar{y}_{\epsilon_1}+\theta^*)-\mathcal{T}(\theta^*)\|_c &= \|\nabla\mathcal{T}(\lambda\bar{y}_{\epsilon_1}+\theta^*)\bar{y}_{\epsilon_1}\|_c\\
&\geq \|\nabla\mathcal{T}(\theta^*)\bar{y}_{\epsilon_1}\|_c-\|(\nabla\mathcal{T}(\lambda\bar{y}_{\epsilon_1}+\theta^*)-\nabla\mathcal{T}(\theta^*))\bar{y}_{\epsilon_1}\|_c\\
&\geq \|\bar{y}_{\epsilon_1}\|_c-\|(\nabla\mathcal{T}(\lambda\bar{y}_{\epsilon_1}+\theta^*)-\nabla\mathcal{T}(\theta^*))\|_c\|\bar{y}_{\epsilon_1}\|_c.
\end{aligned}
$$

By the smoothness of $\mathcal{T}(\cdot)$ in $B^d(\theta^*,\epsilon)$, we can always have an efficiently small $\epsilon_1$ such that

$$
\|\mathcal{T}(\bar{y}_{\epsilon_1}+\theta^*)-\mathcal{T}(\theta^*)\|_c>\gamma\|\bar{y}_{\epsilon_1}\|_c,
$$

which contradicts with the fact that $\mathcal{T}(\cdot)$ is a contraction.

Therefore, we know $y=0$ and $\mathbb{E}[Y]=0$.

**Case 2:** If $g(\cdot)$ is not smooth, by equation (51), we obtain

$$
\begin{aligned}
E[Y] &= \lim_{\alpha\to 0}\frac{1}{\sqrt{\alpha}}\mathbb{E}[(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)}+\theta^*)-\mathcal{T}(\theta^*))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]\\
&= \lim_{\alpha\to 0}\mathbb{E}[g(\frac{F(\sqrt{\alpha}Y^{(\alpha)})}{\sqrt{\alpha}})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]\\
&= \lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(\lambda\sqrt{\alpha}Y^{(\alpha)})Y^{(\alpha)})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]\\
&= \lim_{\alpha\to 0}\mathbb{E}[(g(\nabla F(\lambda\sqrt{\alpha}Y^{(\alpha)})Y^{(\alpha)})-g(\nabla F(0)Y^{(\alpha)}))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))]\\
&\quad+\lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)Y^{(\alpha)})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)}\in B^d(0,\epsilon))].
\end{aligned}
$$

For the first term, we have

$$\begin{aligned}
&\lim_{\alpha\to 0} \|\mathbb{E}[(g(\nabla F(\lambda\sqrt{\alpha}Y^{(\alpha)})Y^{(\alpha)}) - g(\nabla F(0)Y^{(\alpha)}))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon))]\|_c \\
=&\lim_{\alpha\to 0} \|\mathbb{E}[(g(\nabla F(\lambda\sqrt{\alpha}Y^{(\alpha)})\frac{Y^{(\alpha)}}{\|Y^{(\alpha)}\|_2}) - g(\nabla F(0)\frac{Y^{(\alpha)}}{\|Y^{(\alpha)}\|_2}))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon))\|Y^{(\alpha)}\|_2]\|_c \\
\le&\underbrace{\lim_{\alpha\to 0} \sqrt{\mathbb{E}\|(g(\nabla F(\lambda\sqrt{\alpha}Y^{(\alpha)})\frac{Y^{(\alpha)}}{\|Y^{(\alpha)}\|_2}) - g(\nabla F(0)\frac{Y^{(\alpha)}}{\|Y^{(\alpha)}\|_2}))\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon))\|_c^2}}_{\in o(1)}\underbrace{\sqrt{\mathbb{E}\|Y^{(\alpha)}\|_2^2}}_{\in\mathcal{O}(1)} \\
=&0.
\end{aligned}$$

Therefore, we have

$$\mathbb{E}[Y] = \lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)Y^{(\alpha)})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon))].$$

$$\begin{aligned}
&\lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)Y^{(\alpha)})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))] \\
=&\lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)\frac{Y^{(\alpha)}}{\|Y^{(\alpha)}\|_2})\|Y^{(\alpha)}\|_2\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))] \\
\overset{\text{(i)}}{\le}&\max_{\theta:\|\theta\|_2=1} g(\nabla F(0)\theta)\lim_{\alpha\to 0}\mathbb{E}[\|Y^{(\alpha)}\|_2\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))] \\
\le&\max_{\theta:\|\theta\|_2=1} g(\nabla F(0)\theta)\lim_{\alpha\to 0}\underbrace{\sqrt{\mathbb{E}[\|Y^{(\alpha)}\|_2^2]}}_{\in\mathcal{O}(1)}\underbrace{\sqrt{\mathbb{P}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))}}_{\in\mathcal{O}(\alpha^{\frac{1}{4}})} = 0.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\mathbb{E}[Y] =& \lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)Y^{(\alpha)})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \in B^d(0,\epsilon)) \\
&+ \lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)Y^{(\alpha)})\mathbb{1}(\alpha^{\frac{1}{4}}Y^{(\alpha)} \notin B^d(0,\epsilon))] \\
=& \lim_{\alpha\to 0}\mathbb{E}[g(\nabla F(0)Y^{(\alpha)})].
\end{aligned}$$

By [Dur19, Exercise 3.2.5], we obtain

$$\mathbb{E}[Y] = \mathbb{E}[g(\nabla F(0)Y)].$$

If $\nabla F(0) = 0$, we obtain $\mathbb{E}[Y] = 0$.

Now suppose that $\nabla F(0) \neq 0$. Let $h(Y) := g(\nabla F(0)Y)$. If there exists $i \in [d]$ such that the subdifferential or supdifferential of $h_i(\cdot)$ at 0 is not singleton. Without loss of generality, the subdifferential of $h_1(\cdot)$ at 0 is not singleton. Then, there exists $z_1, z_2 \in \mathbb{R}^d$ such that

$$h_1(Y) = h_i(Y) - h_i(0) \ge z_j^T Y, \quad j = 1, 2.$$

If $\mathbb{E}[Y] = 0$, then $\mathbb{E}[h(Y)] = 0$. Therefore, we have

$$\mathbb{E}[h_1(Y) - z_j^T Y] = 0, \quad j = 1, 2.$$

Because $h_1(Y) - z_j^T Y = 0$ are always nonnegative for $j = 1, 2$. We have $h_1(Y) - z_j^T Y = 0$ almost surely for $j = 1, 2$. Therefore, we have $z_1^T Y = z_2^T Y$ almost surely. Let $\zeta = z_1 - z_2$ and we obtain $\zeta^T Y = 0$ almost surely, which implies

$$\mathbb{E}[(\zeta^T Y)^2] = 0. \tag{52}$$

By equation (37) and Theorem 1, we obtain the following equation in distribution.

$$\zeta^T Y^{(\alpha)} = (1-\alpha)\zeta^T Y^{(\alpha)} + \sqrt{\alpha}\zeta^T\left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*) + w\right).$$

Taking second moment to both sides of the above equation, we obtain

$$\begin{aligned}\mathbb{E}[(\zeta^T Y^{(\alpha)})^2] =&(1-\alpha)^2\mathbb{E}[(\zeta^T Y^{(\alpha)})^2] + 2\sqrt{\alpha}(1-\alpha)\mathbb{E}[\zeta^T Y^{(\alpha)} \cdot \zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*)\right)] \\ &+ \alpha\mathbb{E}[(\zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*) + w\right))^2].\end{aligned}$$

By simultaneously subtracting $(1-\alpha)^2\mathbb{E}[(\zeta^T Y^{(\alpha)})^2]$ and dividing $\alpha$ to both sides of the above equation, we obtain

$$\begin{aligned}\underbrace{(2-\alpha)\mathbb{E}[(\zeta^T Y^{(\alpha)})^2]}_{T_1} =&\frac{2(1-\alpha)}{\sqrt{\alpha}}\mathbb{E}[\zeta^T Y^{(\alpha)} \cdot \zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*)\right)] \\ &+ \mathbb{E}[(\zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*) + w\right))^2] \\ =&\underbrace{\frac{2(1-\alpha)}{\sqrt{\alpha}}\mathbb{E}[\zeta^T Y^{(\alpha)} \cdot \zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*)\right)]}_{T_2} \\ &+ \underbrace{\mathbb{E}[(\zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*)\right))^2]}_{T_3} + \underbrace{\mathbb{E}[(\zeta^T w)^2]}_{T_4}.\end{aligned}$$

By equation (52) and Theorem 2, we obtain $\lim_{\alpha\to 0} T_1 = 0$. By Cauchy–Schwarz inequality, we can bound $T_2$ and $T_3$ as follows

$$\begin{aligned}\lim_{\alpha\to 0}|T_2| \leq& \lim_{\alpha\to 0}\frac{2|1-\alpha|}{\sqrt{\alpha}}\sqrt{\mathbb{E}[(\zeta^T Y^{(\alpha)})^2]}\sqrt{\mathbb{E}[\left(\zeta^T \left(\mathcal{T}(\sqrt{\alpha}Y^{(\alpha)} + \theta^*) - \mathcal{T}(\theta^*)\right)\right)^2]} \\ \leq& \lim_{\alpha\to 0}\underbrace{\sqrt{\mathbb{E}[(\zeta^T Y^{(\alpha)})^2]}}_{\in o(1)}\underbrace{\sqrt{\mathbb{E}[\|\zeta\|_c^2\|Y^{(\alpha)}\|_c^2]}}_{\in\mathcal{O}(1)} = 0.\end{aligned}$$

$$\lim_{\alpha\to 0} T_3 \leq \alpha\underbrace{\mathbb{E}[\|\zeta\|_c^2\|Y^{(\alpha)}\|_c^2]}_{\in\mathcal{O}(1)} = 0.$$

Because $\mathrm{Var}(w)$ is positive definite, we obtain

$$T_4 = \zeta^T\,\mathrm{Var}(w)\zeta > 0.$$

However, we have $T_4 = 0$ by letting $\alpha \to 0$, which contradicts with the fact that $T_4 > 0$. Therefore, we have $\mathbb{E}[Y] \neq 0$.

# E Proof of Proposition 2

We first present the following lemma, whose proof is given at the end of this subsection

**Lemma 9.** *Consider iterates $\{q_t\}_{t\geq 0}$ generated by equation (10). For integer $n \geq 1$, under assumption 4(**n**), there exists $\eta$, $\alpha_n$ such that for any $\alpha \leq \alpha_n$, there exist $t_{\alpha,n}$ such that*

$$\mathbb{E}[M_\eta^n(q_t - q^*)] \leq \mathbb{E}[M_\eta^n(q_{t_{\alpha,n}} - q^*)](1-\alpha(1-\sqrt{\gamma_0}))^{t-t_{\alpha,n}} + c_n\alpha^n$$

*holds for all $t \geq t_{\alpha,n}$, where $M_\eta(x)$ is constructed by equation (21) and $\{c_n\}_{n\geq 1}$ are constants that are independent with $\alpha$ and $t$. Moreover, $t_{\alpha,1} = 0$.*

Then, by the property (3) in Proposition 5, we complete the proof of Proposition 2

## E.1 Proof of Lemma 9

We use the induction to give the proof of Lemma 9

**Base Case:** $n = 1$.

By subtracting $q^*$ to both side of equation (10), we obtain

$$\begin{aligned} q_{t+1} - q^* &= (1-\alpha)(q_t - q^*) + \alpha(\gamma D_t P_t f(q_t) + (I - D_t)q_t + D_t r_t - q^*) \\ &\overset{\text{(i)}}{=} (1-\alpha)(q_t - q^*) + \alpha\Big(\mathcal{T}(q_t) - \mathcal{T}(q^*) + \gamma(D_t P_t - DP)f(q_t) + (D - D_t)q_t + (D_t r_t - D\bar{r})\Big) \\ &\overset{\text{(ii)}}{=} (1-\alpha)(q_t - q^*) + \alpha\Big(\mathcal{T}(q_t) - \mathcal{T}(q^*) + A_t f(q_t) + B_t q_t + C_t\Big), \end{aligned} \tag{53}$$

where (i) holds by $\gamma DPf(q^*) + Dr = Dq^*$ and denoting

$$\mathcal{T}(q) := \gamma DPf(q) + (I - D)q, \tag{54}$$

and (ii) holds by denoting $A_t = \gamma D_t P_t - \gamma DP$, $B_t = D - D_t$ and $C_t = D_t r_t - D\bar{r}$, thereby $\{(A_t, B_t, C_t)\}_{t\geq 0}$ are i.i.d. zero mean random variables and $(A_t, B_t, C_t)$ is independent with $q_t$. By [CMZ23, Proposition 3.3], we obtain that $\mathcal{T}(\cdot)$ is a $\gamma_0$-contraction with respect to $\|\cdot\|_c$, where $\gamma_0 = 1 - (1-\gamma)\min_{i\in\mathcal{S}\times\mathcal{A}} D_{ii}$.

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of equation (53) and by property (1) in Proposition 5, we obtain

$$\begin{aligned} M_\eta(q_{t+1} - q^*) \leq &(1-\alpha)^2 M_\eta(q_t - q^*) + \underbrace{(1-\alpha)\alpha\langle\nabla M_\eta(q_t - q^*), \mathcal{T}(q_t) - \mathcal{T}(q^*) + A_t f(q_t) + B_t q_t + C_t\rangle}_{T_1} \\ &+ \underbrace{\frac{\alpha^2}{2\eta}\|\mathcal{T}(q_t) - \mathcal{T}(q^*) + A_t f(q_t) + B_t q_t + C_t\|_2^2}_{T_2}. \end{aligned}$$

For $T_1$ we have

$$\begin{aligned} T_1 &= (1-\alpha)\alpha\left(\langle\nabla M_\eta(q_t - q^*), \mathcal{T}(q_t) - \mathcal{T}(q^*)\rangle + \langle\nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle\right) \\ &\overset{\text{(i)}}{\leq} (1-\alpha)\alpha\left(\|q_t - q^*\|_m\|\mathcal{T}(q_t) - \mathcal{T}(q^*)\|_m + \langle\nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle\right) \\ &\overset{\text{(ii)}}{\leq} \frac{(1-\alpha)\alpha\gamma_0}{l_{cm}}\|q_t - q^*\|_m\|q_t - q^*\|_c + (1-\alpha)\alpha\langle\nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle \\ &\overset{\text{(iii)}}{\leq} \frac{2\alpha(1-\alpha)\gamma_0 u_{cm}}{l_{cm}} M_\eta(q_t - q^*) + (1-\alpha)\alpha\langle\nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle, \end{aligned}$$

where (i) follows from property (4) of Proposition 5, (ii) follows from property (3) of Proposition 5 and $\gamma_0$-contraction of $\mathcal{T}(\cdot)$, and (iii) follows from property (2) of Proposition 5. For $T_2$ we have

$$\begin{aligned} T_2 &\leq \frac{\alpha^2}{2\eta l_{cs}^2}\|\mathcal{T}(q_t) - \mathcal{T}(q^*) + A_t f(q_t) + B_t q_t + C_t\|_c^2 \\ &\leq \frac{\alpha^2}{\eta l_{cs}^2}\left(\|\mathcal{T}(q_t) - \mathcal{T}(q^*)\|_c^2 + \|A_t f(q_t) + B_t q_t + C_t\|_c^2\right) \\ &\leq \frac{2\alpha^2\gamma_0^2 u_{cm}^2}{\eta l_{cs}^2} M_\eta(q_t - q^*) + \frac{\alpha^2\|A_t f(q_t) + B_t q_t + C_t\|_c^2}{\eta l_{cs}^2}. \end{aligned}$$

Combining the bound for $T_1$, $T_2$, we obtain

$$\begin{aligned} M_\eta(q_{t+1} - q^*) \leq &\left(1 - 2\alpha(1 - \frac{(1-\alpha)\gamma_0 u_{cm}}{l_{cm}}) + \alpha^2(1 + \frac{2\gamma_0^2 u_{cm}^2}{\eta l_{cs}^2})\right) M_\eta(q_t - q^*) \\ &+ (1-\alpha)\alpha\langle\nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle + \mathcal{O}(\alpha^2)\|A_t f(q_t) + B_t q_t + C_t\|_c^2. \end{aligned}$$

Recall that $\frac{u_{cm}}{l_{cm}} = \sqrt{\frac{1+\eta u_{cs}^2}{1+\eta l_{cs}^2}}$ by property (3) in Proposition 5. We can always choose a sufficient small $\eta > 0$ such that $\frac{u_{cm}}{l_{cm}} \leq \frac{1}{\sqrt{\gamma_0}}$, which implies $-2\alpha(1 - \frac{(1-\alpha)\gamma_0 u_{cm}}{l_{cm}}) \leq -2\alpha(1-(1-\alpha)\sqrt{\gamma_0}) \leq -2\alpha(1-\sqrt{\gamma_0})$. Furthermore, there always exists $\alpha_0 > 0$ such that $\left(1 - 2\alpha(1-\sqrt{\gamma_0}) + \alpha^2(1+\frac{2\gamma_0^2 u_{cm}^2}{\eta l_{cs}^2})\right) \leq 1 - \frac{3}{2}\alpha(1-\sqrt{\gamma_0}) < 1$ when $\alpha \leq \alpha_0$. Therefore, for $\forall \alpha \leq \alpha_0$ and $t \geq 0$, we obtain

$$\begin{aligned} M_\eta(q_{t+1} - q^*) \leq & \left(1 - \frac{3}{2}\alpha(1-\sqrt{\gamma_0})\right) M_\eta(q_t - q^*) + (1-\alpha)\alpha\langle \nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle \\ & + \mathcal{O}(\alpha^2)\|A_t f(q_t) + B_t q_t + C_t\|_c^2. \end{aligned} \tag{55}$$

Taking expectation to equation (55), there exist $\alpha_1 \leq \alpha_0$ such that for $\forall \alpha \leq \alpha_1$, we obtain

$$\begin{aligned} & \mathbb{E}[M_\eta(q_{t+1} - q^*)] \\ \leq & \left(1 - \frac{3}{2}\alpha(1-\sqrt{\gamma_0})\right) \mathbb{E}[M_\eta(q_t - q^*)] \\ & + \mathcal{O}(\alpha^2)\mathbb{E}[\|A_t(f(q_t) - f(q^*)) + B_t(q_t - q^*) + A_t f(q^*) + B_t q^* + C_t\|_c^2] \\ \overset{\text{(i)}}{\leq} & \left(1 - \frac{3}{2}\alpha(1-\sqrt{\gamma_0}) + \mathcal{O}(\alpha^2)\right) \mathbb{E}[M_\eta(q_t - q^*)] + \mathcal{O}(\alpha^2) \\ \overset{\text{(ii)}}{\leq} & (1 - \alpha(1-\sqrt{\gamma_0}))\, \mathbb{E}[M_\eta(q_t - q^*)] + \mathcal{O}(\alpha^2) \\ \leq & (1 - \alpha(1-\sqrt{\gamma_0}))^{t+1}\, \mathbb{E}[M_\eta(q_0 - q^*)] + \sum_{k=0}^{t} (1 - \alpha(1-\sqrt{\gamma_0}))^k\, \mathcal{O}(\alpha^2) \\ \leq & (1 - \alpha(1-\sqrt{\gamma_0}))^{t+1}\, \mathbb{E}[M_\eta(q_0 - q^*)] + \mathcal{O}(\alpha), \end{aligned}$$

where (i) holds because the second moment of $(A_t, B_t, C_t)$ is finite and there exist $\alpha_1$ such that (ii) holds for $\forall \alpha \leq \alpha_1$.

**Induction Step:** Given positive integer $k \geq 2$, assume Proposition 2 holds for all $n \leq k-1$. When $n = k$, we let

$$\begin{aligned} T_1 &= \left(1 - \frac{3}{2}\alpha(1-\sqrt{\gamma_0})\right) M_\eta(q_t - q^*) \\ T_2 &= (1-\alpha)\alpha\langle \nabla M_\eta(q_t - q^*), A_t f(q_t) + B_t q_t + C_t\rangle \\ T_3 &= \mathcal{O}(\alpha^2)\|A_t f(q_t) + B_t q_t + C_t\|_c^2. \end{aligned}$$

Take $k$-th moment to both sides of equation (55) and we obtain

$$\mathbb{E}[M_\eta^k(q_{t+1} - q^*)] \leq \mathbb{E}\left[(T_1 + T_2 + T_3)^k\right] = \underbrace{\mathbb{E}\left[\sum_{a+b=k} \tbinom{k}{a}\tbinom{k-a}{b} T_1^a T_2^b\right]}_{S_1} + \underbrace{\mathbb{E}\left[\sum_{a+b+c=k, c\geq 1} \tbinom{k}{a}\tbinom{k-a}{b} T_1^a T_2^b T_3^c\right]}_{S_2}. \tag{56}$$

For $S_1$ we have

$$
\begin{aligned}
S_1 \le& \Big(1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0})\Big)^k \mathbb{E}[M_\eta^k(q_t-q^*)] \\
&+ \mathbb{E}\Big[\sum_{a+b=k, b\ge 2} \tbinom{k}{a}\tbinom{k-a}{b}\alpha^b M_\eta^a(q_t-q^*)\|q_t-q^*\|_m^b\|A_tf(q_t)+B_tq_t+C_t\|_m^b\Big] \\
\le& \Big(1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0})\Big)^k \mathbb{E}[M_\eta^k(q_t-q^*)] \\
&+ \mathbb{E}\Big[\sum_{a+b=k, b\ge 2} \mathcal{O}(\alpha^b)M_\eta^{a+\frac{b}{2}}(q_t-q^*)+\mathcal{O}(\alpha^b)M_\eta^{a+b}(q_t-q^*)\Big] \\
\le& \Big((1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0}))^k+\sum_{b=2}^k\mathcal{O}(\alpha^b)\Big)\mathbb{E}[M_\eta^k(q_t-q^*)]+\mathbb{E}\Big[\sum_{a+b=k,b\ge 2}\mathcal{O}(\alpha^b)M_\eta^{a+\frac{b}{2}}(q_t-q^*)\Big] \\
\le& \Big((1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0}))^k+\sum_{b=2}^k\mathcal{O}(\alpha^b)\Big)\mathbb{E}[M_\eta^k(q_t-q^*)] \\
&+ \sum_{a+b=k,b\ge 2, b\text{ is even}}\mathbb{E}\Big[\mathcal{O}(\alpha^b)M_\eta^{a+\frac{b}{2}}(q_t-q^*)\Big]+\sum_{a+b=k,b\ge 3, b\text{ is odd}}\mathbb{E}\Big[\mathcal{O}(\alpha^b)M_\eta^{a+\frac{b}{2}}(q_t-q^*)\Big] \\
\le& \Big((1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0}))^k+\mathcal{O}(\alpha^2)\Big)\mathbb{E}[M_\eta^k(q_t-q^*)] \\
&+ \sum_{a+b=k,b\ge 2, b\text{ is even}}\mathcal{O}(\alpha^b)\underbrace{\mathbb{E}\Big[M_\eta^{a+\frac{b}{2}}(q_t-q^*)\Big]}_{\in\mathcal{O}(\alpha^{a+\frac{b}{2}}),\because a+\frac{b}{2}\le k-1} \\
&+ \sum_{a+b=k,b\ge 3, b\text{ is odd}}\mathcal{O}(\alpha^b)\sqrt{\underbrace{\mathbb{E}\Big[M_\eta^{a+\frac{b+1}{2}}(q_t-q^*)\Big]}_{\in\mathcal{O}(\alpha^{a+\frac{b+1}{2}}),\because a+\frac{b+1}{2}\le k-1}\underbrace{\mathbb{E}\Big[M_\eta^{a+\frac{b-1}{2}}(q_t-q^*)\Big]}_{\in\mathcal{O}(\alpha^{a+\frac{b-1}{2}})}} \\
\overset{\text{(i)}}{\le}& \Big((1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0}))^k+\mathcal{O}(\alpha^2)\Big)\mathbb{E}[M_\eta^k(q_t-q^*)]+\mathcal{O}(\alpha^{k+1}),
\end{aligned}
$$

where (i) holds by induction and taking t to be sufficiently large and $\alpha\le\min(\alpha_1,\dots,\alpha_{k-1})$.

For $S_2$ we have

$$
\begin{aligned}
S_2 \le& \mathbb{E}\Big[\sum_{a+b+c=k,c\ge 1}\mathcal{O}(\alpha^{b+2c})\cdot M_\eta^{a+\frac{b}{2}}(q_t-q^*)\|A_tf(q_t)+B_tq_t+C_t\|_c^{b+2c}\Big] \\
\le& \sum_{a+b+c=k,c\ge 1}\mathcal{O}(\alpha^{b+2c})\mathbb{E}[M_\eta^k(q_t-q^*)]+\sum_{a+b+c=k,c\ge 1,b\text{ is even}}\mathcal{O}(\alpha^{b+2c})\underbrace{\mathbb{E}\Big[M_\eta^{a+\frac{b}{2}}(q_t-q^*)\Big]}_{\in\mathcal{O}(\alpha^{a+\frac{b}{2}}),\because a+\frac{b}{2}\le k-1} \\
&+ \sum_{a+b+c=k,c\ge 1,b\text{ is odd}}\mathcal{O}(\alpha^{b+2c})\sqrt{\underbrace{\mathbb{E}\Big[M_\eta^{a+\frac{b+1}{2}}(q_t-q^*)\Big]}_{\in\mathcal{O}(\alpha^{a+\frac{b+1}{2}}),\because a+\frac{b+1}{2}\le k-1}\underbrace{\mathbb{E}\Big[M_\eta^{a+\frac{b-1}{2}}(q_t-q^*)\Big]}_{\in\mathcal{O}(\alpha^{a+\frac{b-1}{2}})}} \\
\overset{\text{(i)}}{\le}& \mathcal{O}(\alpha^2)\mathbb{E}[M_\eta^k(q_t-q^*)]+\mathcal{O}(\alpha^{k+1}),
\end{aligned}
$$

where (i) holds by induction and taking t to be sufficiently large and $\alpha\le\min(\alpha_1,\dots,\alpha_{k-1})$.

Combining the bound of $S_1, S_2$ with equation (56), we obtain

$$
\mathbb{E}[M_\eta^k(q_{t+1}-q^*)]\le\left(1-\frac{3}{2}\alpha(1-\sqrt{\gamma_0})\right)\mathbb{E}[M_\eta^k(q_t-q^*)]+\mathcal{O}(\alpha^2)\mathbb{E}[M_\eta^k(q_t-q^*)]+\mathcal{O}(\alpha^{k+1}).
$$

Therefore, there exist $\alpha_k\le\min(\alpha_1,\dots,\alpha_{k-1})$ and for $\forall\alpha\le\alpha_k$, there exist $t_{\alpha,k}$ such that

$$
\mathbb{E}[M_\eta^k(q_t-q^*)]\le\mathbb{E}[M_\eta^k(q_{t_{\alpha,k}}-\theta^*)](1-\alpha(1-\sqrt{\gamma_0}))^{t-t_{\alpha,k}}+c_k\alpha^k
$$

holds for $\forall t\ge t_{\alpha,k}$, where $c_k$ is a constant that is independent with $\alpha$ and $t$.

# F Proof of Theorem 4

## F.1 Unique Limit Distribution

We consider a pair of coupled , $\{q_t^{[1]}\}_{t\geq 0}$ and $\{q_t^{[2]}\}_{t\geq 0}$, defined as

$$\begin{aligned} q_{t+1}^{[1]} &= (1-\alpha)q_t^{[1]} + \alpha\Big(\gamma D_t P_t f(q_t^{[1]}) + (I-D_t)q_t^{[1]} + D_t r_t\Big), \\ q_{t+1}^{[2]} &= (1-\alpha)q_t^{[2]} + \alpha\Big(\gamma D_t P_t f(q_t^{[2]}) + (I-D_t)q_t^{[2]} + D_t r_t\Big). \end{aligned} \tag{57}$$

Here $\{q_t^{[1]}\}_{t\geq 0}$ and $\{q_t^{[2]}\}_{t\geq 0}$ are two iterates coupled by sharing $\{(D_t, P_t, r_t)\}_{t\geq 0}$. We assume that the initial iterates $q_0^{[1]}$ and $q_0^{[2]}$ may depend on each other.

Taking difference to equation (57), we obtain

$$q_{t+1}^{[1]} - q_{t+1}^{[2]} = (1-\alpha)(q_t^{[1]} - q_t^{[2]}) + \alpha\Big(\gamma D_t P_t (f(q_t^{[1]}) - f(q_t^{[2]})) + (I-D_t)(q_t^{[1]} - q_t^{[1]})\Big).$$

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of above equation and by property (1) in Proposition 5, we obtain

$$\begin{aligned} M_\eta(q_{t+1}^{[1]} - q_{t+1}^{[2]}) \leq{} & (1-\alpha)^2 M_\eta(q_t^{[1]} - q_t^{[2]}) \\ & + \alpha(1-\alpha)\langle \nabla M_\eta(q_t^{[1]} - q_t^{[2]}), \gamma D_t P_t (f(q_t^{[1]}) - f(q_t^{[2]})) + (I-D_t)(q_t^{[1]} - q_t^{[1]})\rangle \\ & + \frac{\alpha^2}{2\eta}\|\gamma D_t P_t (f(q_t^{[1]}) - f(q_t^{[2]})) + (I-D_t)(q_t^{[1]} - q_t^{[1]})\|_2^2. \end{aligned}$$

Taking expectations to both sides of above equation, we obtain

$$\begin{aligned} \mathbb{E}[M(q_{t+1}^{[1]} - q_{t+1}^{[2]})] \leq{} & (1-\alpha)^2\mathbb{E}[M(q_t^{[1]} - q_t^{[2]})] + \underbrace{\alpha(1-\alpha)\mathbb{E}[\langle \nabla M(q_t^{[1]} - q_t^{[2]}), \mathcal{T}(q_t^{[1]}) - \mathcal{T}(q_t^{[2]})\rangle]}_{T_1} \\ & + \underbrace{\frac{\alpha^2}{2\eta}\mathbb{E}\|\gamma D_t P_t (f(q_t^{[1]}) - f(q_t^{[2]})) + (I-D_t)(q_t^{[1]} - q_t^{[1]})\|_2^2}_{T_2}. \end{aligned}$$

For $T_1$ we have

$$\begin{aligned} T_1 &\overset{\text{(i)}}{\leq} \alpha(1-\alpha)\mathbb{E}[\|q_t^{[1]} - q_t^{[2]}\|_m \|\mathcal{T}(q_t^{[1]}) - \mathcal{T}(q_t^{[2]})\|_m] \\ &\overset{\text{(ii)}}{\leq} \frac{\alpha(1-\alpha)}{l_{cm}}\mathbb{E}[\|q_t^{[1]} - q_t^{[2]}\|_m \|\mathcal{T}(q_t^{[1]}) - \mathcal{T}(q_t^{[2]})\|_c] \\ &\leq \frac{\alpha(1-\alpha)\gamma_0}{l_{cm}}\mathbb{E}[\|q_t^{[1]} - q_t^{[2]}\|_m \|q_t^{[1]} - q_t^{[2]}\|_c] \\ &\overset{\text{(iii)}}{\leq} \frac{2\alpha(1-\alpha)\gamma_0 u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(q_t^{[1]} - q_t^{[2]})] \overset{\text{(iv)}}{\leq} 2\alpha\sqrt{\gamma_0}\mathbb{E}[M_\eta(q_t^{[1]} - q_t^{[2]})], \end{aligned}$$

where (i) holds because of the property (4) of Proposition 5, (ii) and (iii) holds because of the property (2) and (3) of Proposition 5 and (iv) holds because $\frac{u_{cm}}{l_{cm}} = \sqrt{\frac{1+\eta u_{cs}^2}{1+\eta l_{cs}^2}}$ by property (3) in Proposition 5 and we can always choose a sufficient small $\eta > 0$ such that $\frac{u_{cm}}{l_{cm}} \leq \frac{1}{\sqrt{\gamma_0}}$.

It is easy to verify that $f(\cdot)$ is a non-expansion with respect to $\|\|_c$, by Cauchy-Schwarz inequality and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$T_2 \leq \frac{\alpha^2}{\eta}(2\mathbb{E}\|\gamma D_t P_t (f(q_t^{[1]}) - f(q_t^{[2]}))\|_2^2 + 2\mathbb{E}\|(I-D_t)(q_t^{[1]} - q_t^{[1]})\|_2^2) \in \mathcal{O}(\alpha^2)\mathbb{E}[M_\eta(q_t^{[1]} - q_t^{[2]})].$$

Combining the bound for $T_1$ and $T_2$, there exists $\bar{\alpha}' \leq \alpha_1$ such that

$$\begin{aligned}\mathbb{E}[M(q_{t+1}^{[1]} - q_{t+1}^{[2]})] &\leq (1 - 2\alpha(1-\sqrt{\gamma}) + \mathcal{O}(\alpha^2))\mathbb{E}[M(q_t^{[1]} - q_t^{[2]})] \\ &\leq (1 - \alpha(1-\sqrt{\gamma}))\mathbb{E}[M(q_t^{[1]} - q_t^{[2]})],\end{aligned}$$

for $\forall \alpha \leq \bar{\alpha}'$. Therefore, we have

$$\begin{aligned}W_2^2\left(\mathcal{L}\left(q_t^{[1]}\right), \mathcal{L}\left(q_t^{[2]}\right)\right) &\leq \mathbb{E}\left[\left\|q_t^{[1]} - q_t^{[2]}\right\|_c^2\right] \\ &\leq 2u_{cm}^2\mathbb{E}\left[M(q_t^{[1]} - q_t^{[2]})\right] \leq 2u_{cm}^2\mathbb{E}\left[M(q_0^{[1]} - q_0^{[2]})\right](1 - \alpha(1-\sqrt{\gamma_0}))^t.\end{aligned} \tag{58}$$

Therefore, $W_2^2\left(\mathcal{L}\left(q_t^{[1]}\right), \mathcal{L}\left(q_t^{[2]}\right)\right)$ decays geometrically. Similarly to the argument in Section B.1, we see that the sequence $\{\mathcal{L}(q_t^{[1]})\}_{t\geq 0}$ converges weakly to a unique limit distribution $\bar{\mu} \in \mathcal{P}_2(\mathbb{R}^{|\mathcal{S}||\mathcal{A}|})$ that is independent of the initial iterate distribution of $q_0^{[1]}$.

Finally, we establish the following lemma to bound the second moment of the limit random vector $q^{(\alpha)}$.

**Lemma 10.** *Under Assumption 4, when $\alpha \leq \bar{\alpha}_0'$, we obtain*

$$\mathbb{E}[\|q^{(\alpha)} - q^*\|_2^2] \in \mathcal{O}(\alpha) \quad \text{and} \quad \mathbb{E}[\|q^{(\alpha)}\|_2^2] \in \mathcal{O}(1).$$

*Proof for Lemma 10.* We have shown that the sequence $\{q_t\}_{t\geq 0}$ converges weakly to $q^{(\alpha)}$ in $\mathcal{P}_2(\mathbb{R}^{|\mathcal{S}||\mathcal{A}|})$. It is well known that weak convergence in $\mathcal{P}_2(\mathbb{R}^{|\mathcal{S}||\mathcal{A}|})$ is equivalent to convergence in distribution and the convergence of the first two moments. As a result, we have

$$\mathbb{E}\left[\|q^{(\alpha)} - q^*\|_c^2\right] = \lim_{t\to\infty}\mathbb{E}\left[\|q_t - q^*\|_c^2\right]. \tag{59}$$

Taking $t \to \infty$ on both sides of equation (12) in Proposition 2 with $n = 1$ and combining with equation (59) yields

$$\mathbb{E}[\|q^{(\alpha)} - q^*\|_2^2] \leq \frac{1}{l_{cs}^2}\mathbb{E}[\|q^{(\alpha)} - q^*\|_c^2] \in \mathcal{O}(\alpha).$$

Since $2\|q^*\|_2^2 \in \mathcal{O}(1)$, it follows that

$$\mathbb{E}[\|q^{(\alpha)}\|_2^2] \leq 2\mathbb{E}(\|q^{(\alpha)} - q^*\|_2^2) + 2\|q^*\|_2^2 \in \mathcal{O}(1).$$

□

## F.2 Invariance

Moreover, we will show that the unique limit distribution $\bar{\mu}$ is also a stationary distribution for the Markov chain $\{q_t\}_{t\geq 0}$, as stated in the following lemma.

**Lemma 11.** *Let $\{q_t\}_{t\geq 0}$ and $\{q_t'\}_{t\geq 0}$ be two trajectories of iterates in equation (57), where $\mathcal{L}(q_0) = \bar{\mu}$ and $\mathcal{L}(q_0') \in \mathcal{P}_2(\mathbb{R}^{|\mathcal{S}||\mathcal{A}|})$ is arbitrary. we have*

$$W_2^2(\mathcal{L}(q_1), \mathcal{L}(q_1')) \leq \rho W_2^2(\mathcal{L}(q_0), \mathcal{L}(q_0')),$$

*where the quantity $\rho := \frac{u_{cm}^2}{l_{cm}^2}(1 - \alpha(1-\sqrt{\gamma_0}))$ is independent of $\mathcal{L}(q_0')$. In particular, for any $t \geq 0$, if we set $\mathcal{L}(q_0') = \mathcal{L}(q_t)$, then*

$$W_2^2(\mathcal{L}(q_1), \mathcal{L}(q_{t+1})) \leq \rho W_2^2(\bar{\mu}, \mathcal{L}(q_t)).$$

*Proof of Lemma 11.* We prove this lemma by coupling the two processes $\{q_t\}_{t\geq 0}$ and $\{q_t'\}_{t\geq 0}$ such that

$$W_2^2(\mathcal{L}(q_0), \mathcal{L}(q_0')) = \mathbb{E}\left[\|q_0 - q_0'\|_c^2\right].$$

Since $W_2$ is defined by infimum over all couplings, we have

$$\begin{aligned}W_2^2\left(\mathcal{L}\left(q_1\right),\mathcal{L}(q_1')\right) &\leq \mathbb{E}\left[\|q_1-q_1'\|_c^2\right]\\ &\leq 2u_{cm}^2\mathbb{E}\left[M_\eta(q_1-q_1')\right]\\ &\leq 2u_{cm}^2(1-\alpha(1-\sqrt{\gamma_0}))\mathbb{E}\left[M_\eta(q_0-q_0')\right]\\ &\leq \frac{u_{cm}^2}{l_{cm}^2}(1-\alpha(1-\sqrt{\gamma_0}))\mathbb{E}\left[\|q_0-q_0'\|_c^2\right]=\rho W_2^2\left(\mathcal{L}\left(q_0\right),\mathcal{L}(q_0')\right),\end{aligned}$$

where $\rho=\frac{u_{cm}^2}{l_{cm}^2}(1-\alpha(1-\sqrt{\gamma_0}))$. □

By triangle inequality, we obtain

$$\begin{aligned}W_2\left(\mathcal{L}\left(q_1\right),\bar{\mu}\right) &\leq W_2\left(\mathcal{L}\left(q_1\right),\mathcal{L}\left(q_{t+1}\right)\right)+W_2\left(\mathcal{L}\left(q_{t+1}\right),\bar{\mu}\right)\\ &\leq \sqrt{\rho}W^2\left(\bar{\mu},\mathcal{L}(q_t)\right)+W_2\left(\mathcal{L}\left(q_{t+1}\right),\bar{\mu}\right)\overset{t\to\infty}{\longrightarrow}0,\end{aligned}\tag{60}$$

where the second inequality holds by Lemma 11 and last step comes from the weak convergence result. Therefore, we have proved that $\{q_t\}_{t\geq 0}$ converge to a unique stationary distribution $\bar{\mu}$.

### F.3 Convergence rate

Consider the coupled processes defined as equation (57). Suppose that the initial iterate $q_0^{[2]}$ follows the stationary distribution $\bar{\mu}$, thus $\mathcal{L}(q_t^{[2]})=\bar{\mu}$ for all $t\geq 0$. By equation (58), we have for all $t\geq 0$ :

$$\begin{aligned}W_2^2\left(\mathcal{L}(q_t^{[1]}),\bar{\mu}\right) &= W_2^2\left(\mathcal{L}(q_t^{[1]}),\mathcal{L}(q_t^{[2]})\right)\\ &\leq 2u_{cm}^2\mathbb{E}\left[M_\eta(q_0^{[1]}-q_0^{[2]})\right](1-\alpha(1-\sqrt{\gamma_0}))^t\\ &\leq 2u_{cm}^2\mathbb{E}\left[M_\eta(q_0^{[1]}-q^{(\alpha)})\right](1-\alpha(1-\sqrt{\gamma_0}))^t.\end{aligned}$$

Lemma 10 states that the second moment of $q^{(\alpha)}$ is bounded by a constant. Combining this bound with above equation, we obtain

$$W_2^2(\mathcal{L}(q_t),\mu)\leq c\cdot(1-\alpha(1-\sqrt{\gamma_0}))^t,$$

where $c$ is a universal constant that is independent with $\alpha$ and $t$.

# G Proof of Theorem 5

We can obtain the following dynamic for $Y_t$ by equation (53)

$$Y_{t+1}=(1-\alpha)Y_t+\alpha\Big(\mathcal{T}(Y_t+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(\frac{q^*}{\sqrt{\alpha}})\Big)+\alpha A_t f(Y_t+\frac{q^*}{\sqrt{\alpha}})+\alpha B_t Y_t+\sqrt{\alpha}B_t q^*+\sqrt{\alpha}C_t,\tag{61}$$

where $\{(A_t,B_t,C_t)\}_{t\geq 0}$ are zero mean variables.

Define $g(x):\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}\to\mathbb{R}^{|\mathcal{S}|}$ such that

$$g_s(x):=\max_{a\in\mathcal{A}^*(s)}x(s,a),$$

and $h(x,y):=f(x+y)-f(y)-g(x)$. Therefore, $g(x)$ is a non-expansion mapping with respect to $\|\cdot\|_c$ by [CMZ23, Proposition 3.3]. We define $\mathcal{T}_0(q):=\gamma DPg(q)+(I-D)q$ and it is easy to verify that $\mathcal{T}_0(\cdot)$ is a $\gamma_0$-contraction with respect to $\|\|_c$. By definition of $g(x)$ and $h(x,y)$, we can reformulate equation (61) as

$$\begin{aligned}Y_{t+1}=&(1-\alpha)Y_t+\alpha\mathcal{T}_0(Y_t)+\alpha\gamma DPh(Y_t,\frac{q^*}{\sqrt{\alpha}})+\\ &+\alpha A_t h(Y_t,\frac{q^*}{\sqrt{\alpha}})+\sqrt{\alpha}A_t f(q^*)+\alpha A_t g(Y_t)+\alpha B_t Y_t+\sqrt{\alpha}B_t q^*+\sqrt{\alpha}C_t\\ \overset{\text{(i)}}{=}&(1-\alpha)Y_t+\alpha\mathcal{T}_0(Y_t)+\alpha\gamma DPh(Y_t,\frac{q^*}{\sqrt{\alpha}})+\alpha A_t h(Y_t,\frac{q^*}{\sqrt{\alpha}})\\ &+\alpha A_t g(Y_t)+\alpha B_t Y_t+\sqrt{\alpha}E_t,\end{aligned}\tag{62}$$

where (i) holds because we denote $E_t := A_t f(q^*) + B_t q^* + C_t$.

Furthermore, we have the following lemma to bound the second moment of $h(Y_t, \frac{q^*}{\sqrt{\alpha}})$.

**Lemma 12.** *Consider iterates $\{Y_t\}_{t\geq 0}$ generated by equation (62) with stepsize $\alpha$, under the same setting as Proposition 2 with $n = 2$, we obtain*

$$\mathbb{E}\left[\|h(Y_t, \frac{q^*}{\sqrt{\alpha}})\|_c^2\right] \in \mathcal{O}(\alpha).$$

*Proof of Lemma 12.* By definition, for $\forall s \in \mathcal{S}$, we obtain

$$\begin{aligned}
h_s(Y_t, \frac{q^*}{\sqrt{\alpha}}) &= f_s(Y_t + \frac{q^*}{\sqrt{\alpha}}) - f_s(\frac{q^*}{\sqrt{\alpha}}) - g_s(Y_t) \\
&= \max_{a\in\mathcal{A}} \left(Y_t(s,a) + \frac{q^*(s,a)}{\sqrt{\alpha}}\right) - \max_{a\in\mathcal{A}} \left(\frac{q^*(s,a)}{\sqrt{\alpha}}\right) - \max_{a\in\mathcal{A}^*(s)} Y_t(s,a) \\
&= \max_{a\in\mathcal{A}} \left(Y_t(s,a) + \frac{q^*(s,a)}{\sqrt{\alpha}}\right) - \max_{a\in\mathcal{A}^*(s)} \left(Y_t(s,a) + \frac{q^*(s,a)}{\sqrt{\alpha}}\right),
\end{aligned}$$

where the last inequality holds because $\mathcal{A}^*(s) = \arg\max_{a\in\mathcal{A}} q^*(s,a)$.

We can easily observe that $h_s(Y_t, \frac{q^*}{\sqrt{\alpha}}) \geq 0$. We define

$$\Delta(s) := \begin{cases} \infty & \text{if } \mathcal{A} = \mathcal{A}^*(s), \\ \max_{a\in\mathcal{A}} q^*(s,a) - \max_{a\in\mathcal{A}\setminus\mathcal{A}^*(s)} q^*(s,a) & \text{if } \mathcal{A} \neq \mathcal{A}^*(s). \end{cases}$$

Then, we can observe that when $\|Y_t\|_c \leq \frac{\Delta(s)}{2\sqrt{\alpha}}$, $h_s(Y_t, \frac{q^*}{\sqrt{\alpha}}) = 0$.

Therefore, we can conclude that

$$\begin{aligned}
0 \leq h_s(Y_t, \frac{q^*}{\sqrt{\alpha}}) &\leq \left(f_s(Y_t + \frac{q^*}{\sqrt{\alpha}}) - f_s(\frac{q^*}{\sqrt{\alpha}}) - g_s(Y_t)\right) \mathbb{1}_{\{\|Y_t\|_c \geq \frac{\Delta(s)}{2\sqrt{\alpha}}\}} \\
&\leq (f_s(Y_t) - g_s(Y_t))\, \mathbb{1}_{\{\|Y_t\|_c \geq \frac{\Delta(s)}{2\sqrt{\alpha}}\}} \\
&\leq (f_s(Y_t) - g_s(Y_t))\, \mathbb{1}_{\{\|Y_t\|_c \geq \frac{\Delta}{2\sqrt{\alpha}}\}},
\end{aligned} \tag{63}$$

where $\Delta = \min_{s\in\mathcal{S}} \Delta(s)$.

By Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
\mathbb{E}\left[\|h(Y_t, \frac{q^*}{\sqrt{\alpha}})\|_c^2\right] &\leq \sqrt{\mathbb{E}\|f(Y_t) - g(Y_t)\|_c^4} \cdot \sqrt{\mathbb{P}(\|Y_t\|_c \geq \frac{\Delta}{2\sqrt{\alpha}})} \\
&\leq \sqrt{8\mathbb{E}\|f(Y_t)\|_c^4 + 8\mathbb{E}\|g(Y_t)\|_c^4} \cdot \sqrt{\mathbb{P}(\|Y_t\|_c \geq \frac{\Delta}{2\sqrt{\alpha}})} \\
&\overset{\text{(i)}}{\leq} \sqrt{16\mathbb{E}\|Y_t\|_c^4} \cdot \sqrt{\mathbb{P}(\|Y_t\|_c \geq \frac{\Delta}{2\sqrt{\alpha}})} \\
&\overset{\text{(ii)}}{\leq} \mathcal{O}\left(\sqrt{\mathbb{P}(\|Y_t\|_c \geq \frac{\Delta}{2\sqrt{\alpha}})}\right) \leq \mathcal{O}\left(\sqrt{\frac{\mathbb{E}(\|Y_t\|_c^4)16\alpha^2}{\Delta^4}}\right) \overset{\text{(iii)}}{\in} \mathcal{O}(\alpha),
\end{aligned}$$

where (i) holds because the non-expansion of $f(\cdot)$ and $g(\cdot)$ with respect to $\|\cdot\|_c$ and (ii) and (iii) hold because of the following Corollary 2 with $n = 2$. $\square$

**Corollary 2.** *For integer $n \geq 1$, under Assumption 4(**n**), there exists $\alpha_n > 0$ such that for any $\alpha \leq \alpha_n$, there exist $t_{\alpha,n} > 0$ such that*

$$\mathbb{E}[\|Y_t^{(\alpha)}\|^{2n}] \leq c_n \mathbb{E}[\|Y_{t_{\alpha,n}}^{(\alpha)}\|^{2n}](1 - \alpha(1 - \sqrt{\gamma}))^{t - t_{\alpha,n}} + c'_n, \quad t \geq t_{\alpha,n},$$

*where $\|\cdot\|$ is an arbitrary norm and $\{c_n\}_{n\geq 1}$ and $\{c'_n\}_{n\geq 1}$ are universal constants that are independent with $\alpha$ and $t$. Moreover, $t_{\alpha,1} = 0$.*

*Proof of Corollary 2.* By the equivalence of all norms on $\mathbb{R}^d$, we can obtain the Corollary 2(**n**) by dividing $\alpha^n$ to both sides of equation (12) in Proposition 2. $\square$

## G.1 Step 1: Gaussian Noise and Rational Stepsize

We consider a pair of coupled $\{Y_t\}_{t\geq 0}$ and $\{Y_t'\}_{t\geq 0}$, defined as

$$
\begin{aligned}
Y_{t+1} =&(1-\alpha)Y_t + \alpha\mathcal{T}_0(Y_t) + \alpha\gamma DPh(Y_t, \frac{q^*}{\sqrt{\alpha}}) + \alpha\frac{A'_{kt}+\cdots+A'_{kt+k-1}}{\sqrt{k}}h(Y_t, \frac{q^*}{\sqrt{\alpha}})\\
&+\alpha\frac{A'_{kt}+\cdots+A'_{kt+k-1}}{\sqrt{k}}g(Y_t) + \alpha\frac{B'_{kt}+\cdots+B'_{kt+k-1}}{\sqrt{k}}Y_t + \sqrt{\alpha}\frac{E'_{kt}+\cdots+E'_{kt+k-1}}{\sqrt{k}}, \qquad (64)\\
Y'_{t+1} =&(1-\frac{\alpha}{k})Y'_t + \frac{\alpha}{k}\mathcal{T}_0(Y'_t) + \frac{\alpha}{k}\gamma DPh(Y'_t, \frac{q^*}{\sqrt{\frac{\alpha}{k}}}) + \frac{\alpha}{k}A'_t h(Y'_t, \frac{q^*}{\sqrt{\frac{\alpha}{k}}}) + \frac{\alpha}{k}A'_t g(Y'_t) + \frac{\alpha}{k}B'_t Y'_t + \sqrt{\frac{\alpha}{k}}E'_t,
\end{aligned}
$$

where $\{(A'_t, B'_t, E'_t)\}_{t\geq 0}$ are i.i.d. noise with normal distribution, zero mean and the same variance as $\{(A_t, B_t, E_t)\}_{t\geq 0}$ and $k\geq 1$ is an integer. Therefore, $(\frac{A'_{kt}+\cdots+A'_{kt+k-1}}{\sqrt{k}}, \frac{B'_{kt}+\cdots+B'_{kt+k-1}}{\sqrt{k}}, \frac{E'_{kt}+\cdots+E'_{kt+k-1}}{\sqrt{k}})$ has the same distribution as $(A'_t, B'_t, E'_t)$.

Therefore, we have

$$
\begin{aligned}
Y'_{kt+k} =&(1-\frac{\alpha}{k})^k Y'_{kt} + \frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}\Big(\mathcal{T}_0(Y'_{kt+k-j}) + (\gamma DP + A'_{kt+k-j})h(Y'_{kt+k-j}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})\Big)\\
&+\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}\Big(B'_{kt+k-j}Y'_{kt+k-j} + A'_{kt+k-j}g(Y'_{kt+k-j})\Big) + \sqrt{\frac{\alpha}{k}}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}E'_{kt+k-j}\\
\overset{\text{(i)}}{=}&(1-\alpha)Y'_{kt} + \mathcal{O}(\alpha^2)Y'_{kt} + \alpha\mathcal{T}_0(Y'_{kt}) + \frac{\alpha}{k}\sum_{j=1}^{k}\mathcal{T}_0(Y'_{kt+k-j}) - \mathcal{T}_0(Y'_{kt})\\
&+\frac{\alpha}{k}\sum_{j=1}^{k}((1-\frac{\alpha}{k})^{j-1}-1)\mathcal{T}_0(Y'_{kt+k-j}) \qquad (65)\\
&+\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}(\gamma DP + A'_{kt+k-j})h(Y'_{kt+k-j}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}}) + \frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}A'_{kt+k-j}g(Y'_{kt+k-j})\\
&+\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}B'_{kt+k-j}Y'_{kt+k-j} + \sqrt{\frac{\alpha}{k}}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}E'_{kt+k-j},
\end{aligned}
$$

where (i) holds by equation (42).

Combining equation (64) and (65), we obtain

$$
\begin{aligned}
Y_{t+1}-Y'_{kt+k} =&(1-\alpha)(Y_t-Y'_{kt})+\mathcal{O}(\alpha^2)Y'_{kt}+\alpha(\mathcal{T}_0(Y_t)-\mathcal{T}_0(Y'_{kt}))-\frac{\alpha}{k}\sum_{j=1}^{k}\mathcal{T}_0(Y'_{kt+k-j})-\mathcal{T}_0(Y'_{kt})\\
&-\frac{\alpha}{k}\sum_{j=1}^{k}((1-\frac{\alpha}{k})^{j-1}-1)\mathcal{T}_0(Y'_{kt+k-j})\\
&+\alpha\gamma DPh(Y_t,\frac{q^*}{\sqrt{\alpha}})-\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}\gamma DPh(Y'_{kt+k-j},\frac{q^*}{\sqrt{\frac{\alpha}{k}}})\\
&+\alpha\frac{A'_{kt}+\cdots+A'_{kt+k-1}}{\sqrt{k}}h(Y_t,\frac{q^*}{\sqrt{\alpha}})-\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}A'_{kt+k-j}h(Y'_{kt+k-j},\frac{q^*}{\sqrt{\frac{\alpha}{k}}})\\
&+\alpha\frac{A'_{kt}+\cdots+A'_{kt+k-1}}{\sqrt{k}}g(Y_t)-\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}A'_{kt+k-j}g(Y'_{kt+k-j})\\
&+\alpha\frac{B'_{kt}+\cdots+B'_{kt+k-1}}{\sqrt{k}}Y_t-\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}B'_{kt+k-j}Y'_{kt+k-j}\\
&+\sqrt{\frac{\alpha}{k}}\sum_{j=1}^{k}(1-(1-\frac{\alpha}{k})^{j-1})E'_{kt+k-j}\\
:=&(1-\alpha)(Y_t-Y'_{kt})+R,
\end{aligned}
$$

where $R$ collects all but the first terms on the right hand side.

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of the above equation and by property (1) in Proposition 5, we obtain

$$
M_\eta(Y_{t+1}-Y'_{kt+k})\le(1-\alpha)^2M_\eta(Y_t-Y'_{kt})+(1-\alpha)\underbrace{\langle\nabla M_\eta(Y_t-Y'_{kt}),R\rangle}_{T_1}+\frac{1}{2\eta}\underbrace{\|R\|_2^2}_{T_2}. \tag{66}
$$

The following lemmas, proved in Sections G.1.1 and G.1.2 to follow, control the $T_1$ and $T_2$ terms above.

**Lemma 13.** *Under the setting of Theorem 5, we have*

$$
\mathbb{E}[T_1]\le\frac{2\alpha\gamma_0u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(Y_t-Y'_{kt})]+\mathcal{O}(\alpha^{\frac{3}{2}}).
$$

**Lemma 14.** *Under the setting of Theorem 5, we have*

$$
\mathbb{E}[T_2]\le\mathcal{O}(\alpha^2)\cdot\mathbb{E}[M_\eta(Y_t-Y'_{kt})]+\mathcal{O}(\alpha^2).
$$

Plugging the above bounds for $T_1$ and $T_2$ into equation (66), we obtain

$$
\mathbb{E}[M_\eta(Y_{t+1}-Y'_{kt+k})]\le(1-2\alpha(1-\frac{\gamma_0u_{cm}}{l_{cm}})+\mathcal{O}(\alpha^2))\mathbb{E}[M_\eta(Y_t-Y'_{kt})]+\mathcal{O}\left(\alpha^{\frac{3}{2}}\right)
$$

By the similar argument as in the proof of Lemma 1, we can always choose proper $\eta,\bar{\alpha}$ such that for $\forall\alpha\le\bar{\alpha}$, there exist $t_\alpha$ such that for all $t\ge t_\alpha$, we obtain

$$
\mathbb{E}[M_\eta(Y_{t+1}-Y'_{kt+k})]\le(1-\alpha(1-\sqrt{\gamma_0}))\,\mathbb{E}[M_\eta(Y_t-Y'_{kt})]+\mathcal{O}\left(\alpha^{\frac{3}{2}}\right),
$$

which implies

$$
\lim_{t\to\infty}\mathbb{E}[M_\eta(Y_t-Y'_{kt})]=\mathcal{O}\left(\alpha^{\frac{1}{2}}\right).
$$

By triangle inequality, we have

$$
\begin{aligned}
W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\right) &\le \lim_{t\to\infty} \underbrace{W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y_t)\right)}_{\to 0} + W_2\left(\mathcal{L}(Y_t), \mathcal{L}(Y'_{kt})\right) + \underbrace{W_2\left(\mathcal{L}(Y'_{kt}), \mathcal{L}(Y^{(\alpha/k)})\right)}_{\to 0} \\
&\le \lim_{t\to\infty} \sqrt{\mathbb{E}[\|Y_t - Y'_{kt}\|_c^2]} \le \lim_{t\to\infty} \sqrt{2u_{cm}^2 \mathbb{E}[M(Y_t - Y'_{kt})]} \in \mathcal{O}(\alpha^{\frac{1}{4}}).
\end{aligned}
$$

Then, we can say for all $k \in \mathbb{N}^+$ and $\alpha > 0$,

$$
W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\right) = \mathcal{O}\left(\alpha^{\frac{1}{4}}\right).
$$

When $k \in \mathbb{Q}^+, k > 1$ and $\alpha > 0$, let $k = \frac{p}{q}$. Therefore, we obtain

$$
\begin{aligned}
W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/k)})\right) &\le W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y^{(\alpha/p)})\right) + W_2\left(\mathcal{L}(Y^{(\alpha/p)}), \mathcal{L}(Y^{(\alpha/k)})\right) \\
&\overset{(i)}{\le} \mathcal{O}\left(\alpha^{\frac{1}{4}}\right) + \mathcal{O}\left(\frac{\alpha^{\frac{1}{4}}}{k^{\frac{1}{4}}}\right) \in \mathcal{O}\left(\alpha^{\frac{1}{4}}\right),
\end{aligned}
$$

where (i) holds because $\frac{\alpha}{p} = \frac{\frac{\alpha}{k}}{q}$ and $\frac{\alpha}{k} \le \alpha$.

Then, by the same argument at the end of Section C.1, there exists a unique random variable $Y$ such that

$$
\lim_{\alpha\to 0, \alpha\in\mathbb{Q}^+} W_2\left(\mathcal{L}(Y^{(\alpha)}), \mathcal{L}(Y)\right) = 0,
$$

thereby completing the proof of the first step of Theorem 5.

### G.1.1 Proof of Lemma 13 on $T_1$

By property (4) in Proposition 5, we obtain

$$
\begin{aligned}
\mathbb{E}[T_1] \le\, & \underbrace{\alpha \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\mathcal{T}_0(Y_t) - \mathcal{T}_0(Y'_{kt})\|_m]}_{T_{11}} + \underbrace{\frac{\alpha}{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sum_{j=1}^{k} \mathcal{T}_0(Y'_{kt+k-j}) - \mathcal{T}_0(Y'_{kt})\|_m]}_{T_{12}} \\
& + \underbrace{\alpha\gamma \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|DPh(Y_t, \frac{q^*}{\sqrt{\alpha}})\|_m]}_{T_{13}} + \underbrace{\mathcal{O}(\alpha^2)\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt}\|_m]}_{T_{14}} \\
& + \underbrace{\frac{\alpha}{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sum_{j=1}^{k}((1-\frac{\alpha}{k})^{j-1} - 1)\mathcal{T}_0(Y'_{kt+k-j})\|_m]}_{T_{15}} \\
& + \underbrace{\frac{\alpha\gamma}{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|DP\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1} h(Y'_{kt+k-j}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})\|_m]}_{T_{16}}.
\end{aligned}
$$

Below, we bound $T_{11} \sim T_{16}$ separately.

**The $T_{11}$ Term:**

$$
\begin{aligned}
T_{11} &\le \frac{\alpha}{l_{cm}}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\mathcal{T}_0(Y_t) - \mathcal{T}_0(Y'_{kt})\|_c] \\
&\le \frac{\alpha\gamma_0}{l_{cm}}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y_t - Y'_{kt}\|_c] \\
&\le \frac{\alpha\gamma_0 u_{cm}}{l_{cm}}\mathbb{E}[\|Y_t - Y'_{kt}\|_m^2] = \frac{2\alpha\gamma_0 u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(Y_t - Y'_{kt})].
\end{aligned}
$$

**The $T_{12}$ Term:**

$$\begin{aligned}T_{12} &\leq \frac{\alpha}{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \sum_{j=1}^{k} \|\mathcal{T}_0(Y'_{kt+k-j}) - \mathcal{T}_0(Y'_{kt})\|_m] \\ &\leq \frac{\alpha}{kl_{cm}} \sum_{j=1}^{k} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|g(Y'_{kt+k-j}) - g(Y'_{kt})\|_c] \\ &\leq \frac{\alpha\gamma_0}{kl_{cm}} \sum_{j=1}^{k} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt+k-j} - Y'_{kt}\|_c].\end{aligned}$$

By equation (62), we obtain

$$\begin{aligned}Y'_{kt+k-j} - Y'_{kt} =& (1-\frac{\alpha}{k})^{k-j} Y'_{kt} - Y'_{kt} \\ &+ \frac{\alpha}{k}\sum_{i=1}^{k-j}(1-\frac{\alpha}{k})^{i-1}\Big(\mathcal{T}_0(Y'_{kt+k-j-i}) + (\gamma DP + A'_{kt+k-j-i})h(Y'_{kt+k-j-i}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})\Big) \\ &+ \frac{\alpha}{k}\sum_{i=1}^{k-j}(1-\frac{\alpha}{k})^{i-1}\Big(A'_{kt+k-j-i}g(Y'_{kt+k-j-i}) + B'_{kt+k-j-i}Y'_{kt+k-j-i}\Big) \\ &+ \sqrt{\frac{\alpha}{k}}\sum_{i=1}^{k-j}(1-\frac{\alpha}{k})^{i-1}E'_{kt+k-j-i}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}T_{12} \leq& \frac{\alpha\gamma_0}{kl_{cm}}\sum_{j=1}^{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|(1-\frac{\alpha}{k})^{k-j}Y'_{kt} - Y'_{kt}\|_c] && (T_{121}) \\ &+ \frac{\alpha^2\gamma_0}{k^2 l_{cm}}\sum_{j=1}^{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sum_{i=1}^{k-j}(1-\frac{\alpha}{k})^{i-1}\Big(\mathcal{T}_0(Y'_{kt+k-j-i}) + (\gamma DP + A'_{kt+k-j-i})h(Y'_{kt+k-j-i}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})\Big)\|_c] && (T_{122}) \\ &+ \frac{\alpha^2\gamma_0}{k^2 l_{cm}}\sum_{j=1}^{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sum_{i=1}^{k-j}(1-\frac{\alpha}{k})^{i-1}\Big(A'_{kt+k-j-i}g(Y'_{kt+k-j-i}) + B'_{kt+k-j-i}Y'_{kt+k-j-i}\Big)\|_c]. && (T_{123}) \\ &+ \frac{\alpha^{\frac{3}{2}}\gamma_0}{k^{\frac{3}{2}} l_{cm}}\sum_{j=1}^{k}\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\sum_{i=1}^{k-j}(1-\frac{\alpha}{k})^{i-1}E'_{kt+k-j-i}\|_c]. && (T_{124})\end{aligned}$$

By Corollary 2 with $n = 2$ and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$\begin{aligned}T_{121} &\leq \frac{\alpha\gamma_0}{kl_{cm}}\sum_{j=1}^{k}(1-(1-\frac{\alpha}{k})^{k-j})\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|Y'_{kt}\|_c] \\ &\leq \frac{\alpha\gamma_0}{kl_{cm}}\sum_{j=1}^{k}(1-(1-\frac{\alpha}{k})^{k-j})\underbrace{\sqrt{\mathbb{E}[2\|Y_t\|_m^2 + 2\|Y'_{kt}\|_m^2]}\sqrt{\mathbb{E}[\|Y'_{kt}\|_c^2]}}_{\in\mathcal{O}(1)} \\ &\leq \mathcal{O}(1)\cdot\frac{\alpha}{k}\sum_{j=1}^{k}(1-(1-\frac{\alpha}{k})^{k-j}) \overset{\text{(i)}}{\in} \mathcal{O}(\alpha^2),\end{aligned}$$

where (i) holds by equation (42). Turning to the next two terms, we have

$$T_{122} \le \mathcal{O}(1) \cdot \frac{\alpha^2}{k^2} \sum_{j=1}^{k} \sum_{i=1}^{k-j} (1-\frac{\alpha}{k})^{i-1} \underbrace{\mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\mathcal{T}_0(Y'_{kt+k-j-i}) + (\gamma DP + A'_{kt+k-j-i}) h(Y'_{kt+k-j-i}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})\|_c]}_{\in \mathcal{O}(1)}$$

$$\le \mathcal{O}(1) \cdot \frac{\alpha^2}{k^2} \sum_{j=1}^{k} \sum_{i=1}^{k-j} (1-\frac{\alpha}{k})^{i-1} \le \mathcal{O}(1) \cdot \frac{\alpha^2}{k^2} \cdot k^2 \in \mathcal{O}(\alpha^2).$$

Similarly, we have $T_{123} \in \mathcal{O}(\alpha^2)$

$$\begin{aligned}
T_{123} &\overset{\text{(i)}}{\le} \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=1}^{k} \underbrace{\mathbb{E}[\|Y_t - Y'_{kt}\|_m]}_{\in \mathcal{O}(1)} \mathbb{E}[\| \sum_{i=1}^{k-j} (1-\frac{\alpha}{k})^{i-1} E'_{kt+k-j-i}\|_c] \\
&\le \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=1}^{k} \mathbb{E}[\| \sum_{i=1}^{k-j} (1-\frac{\alpha}{k})^{i-1} E'_{kt+k-j-i}\|_2] \\
&\le \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=1}^{k} \sqrt{\sum_{i=1}^{k-j} \mathbb{E}[\|(1-\frac{\alpha}{k})^{i-1} E'_{kt+k-j-i}\|_2^2]} \\
&\le \mathcal{O}(1) \cdot \frac{\alpha^{\frac{3}{2}}}{k^{\frac{3}{2}}} \sum_{j=1}^{k} \sqrt{k-j} \in \mathcal{O}(\alpha^{\frac{3}{2}}),
\end{aligned}$$

where (i) holds because $Y_t$ and $Y'_{kt}$ are independent with $E'_{kt+k-j-i}$ for $j = 1, \dots, k$ and $i = 1, \dots, k-j$.

Combining the bounds for $T_{121}$, $T_{122}$ and $T_{123}$ together, we obtain $T_{12} \in \mathcal{O}(\alpha^{\frac{3}{2}})$.

**The $T_{13} \sim T_{16}$ Terms:** By Corollary 2 with $n = 2$, Lemma 12 and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$\begin{aligned}
&T_{13} \le \mathcal{O}(\alpha \sqrt{\mathbb{E}[\|h(Y_t, \frac{q^*}{\sqrt{\alpha}})\|_c^2]}) \in \mathcal{O}\left(\alpha^{\frac{3}{2}}\right), \\
&T_{14} \in \mathcal{O}\left(\alpha^2\right) \\
&T_{15} \le \frac{\alpha}{k} \sum_{j=1}^{k} (1-(1-\frac{\alpha}{k})^{j-1}) \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|\mathcal{T}_0(Y'_{kt+k-j})\|_m] \\
&\quad \le \mathcal{O}\left(\frac{\alpha}{k} \sum_{j=1}^{k} (1-(1-\frac{\alpha}{k})^{j-1})\right) \overset{\text{(i)}}{\in} \mathcal{O}\left(\alpha^2\right),
\end{aligned}$$

where (i) holds by equation (42).

$$\begin{aligned}
T_{16} &\le \frac{\alpha\gamma}{k} \sum_{j=1}^{k} (1-\frac{\alpha}{k})^{j-1} \mathbb{E}[\|Y_t - Y'_{kt}\|_m \|DPh(Y'_{kt+k-j}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})\|_m] \\
&\le \mathcal{O}\left(\frac{\alpha}{k} \sum_{j=1}^{k} \sqrt{\frac{\alpha}{k}}\right) \in \mathcal{O}\left(\alpha^{\frac{3}{2}}\right).
\end{aligned}$$

Therefore, we obtain the bound for $\mathbb{E}[T_1]$:

$$\mathbb{E}[T_1] \le \frac{2\alpha\gamma_0 u_{cm}}{l_{cm}} \mathbb{E}[M_\eta(Y_t - Y'_{kt})] + \mathcal{O}(\alpha^{\frac{3}{2}}),$$

thereby completing the proof of Lemma 13.

### G.1.2 Proof of Lemma 14 on $T_2$

By Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}[T_2] \le 9\Bigg( \underbrace{\mathbb{E}[\|\mathcal{O}(\alpha^2)Y'_{kt}\|_2^2]}_{T_{21}} + \underbrace{\mathbb{E}[\|\alpha(\mathcal{T}_0(Y_t) - \mathcal{T}_0(Y'_{kt}))\|_2^2]}_{T_{22}} + \underbrace{\mathbb{E}[\|\frac{\alpha}{k}\sum_{j=1}^{k}\mathcal{T}_0(Y'_{kt+k-j}) - \mathcal{T}_0(Y'_{kt})\|_2^2]}_{T_{23}}$$

$$+ \underbrace{\mathbb{E}[\|\frac{\alpha}{k}\sum_{j=1}^{k}((1-\frac{\alpha}{k})^{j-1}-1)\mathcal{T}_0(Y'_{kt+k-j})\|_2^2]}_{T_{24}} + \underbrace{\mathbb{E}[\|\alpha(\gamma DP + \frac{A'_{kt} + \cdots + A'_{kt+k-1}}{\sqrt{k}})h(Y_t, \frac{q^*}{\sqrt{\alpha}})\|_2^2]}_{T_{25}}$$

$$+ \underbrace{\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}(\gamma DP + A'_{kt+k-j})h(Y'_{kt+k-j}, \frac{q^*}{\sqrt{\frac{\alpha}{k}}})}_{T_{26}}$$

$$+ \underbrace{\mathbb{E}[\|\alpha\frac{A'_{kt} + \cdots + A'_{kt+k-1}}{\sqrt{k}}g(Y_t) + \alpha\frac{B'_{kt} + \cdots + B'_{kt+k-1}}{\sqrt{k}}Y_t\|_2^2]}_{T_{27}}$$

$$+ \underbrace{\mathbb{E}[\|\frac{\alpha}{k}\sum_{j=1}^{k}(1-\frac{\alpha}{k})^{j-1}(A'_{kt+k-j}g(Y'_{kt+k-j}) + B'_{kt+k-j}Y'_{kt+k-j})\|_2^2]}_{T_{28}}$$

$$+ \underbrace{\mathbb{E}[\|\sqrt{\frac{\alpha}{k}}\sum_{j=1}^{k}(1-(1-\frac{\alpha}{k})^{j-1})E'_{kt+k-j}\|_2^2]}_{T_{29}}\Bigg).$$

By Corollary 2(**2**), Lemma 12 and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$\begin{aligned}
&T_{21} \in \mathcal{O}\left(\alpha^4\right), \quad T_{22} \le \mathcal{O}\left(\alpha^2\right)\cdot \mathbb{E}[M_\eta(Y_t - Y'_{kt})], \quad T_{23} \in \mathcal{O}\left(\alpha^2\right), \quad T_{24} \in \mathcal{O}\left(\alpha^2\right)\\
&T_{25} \in \mathcal{O}\left(\alpha^3\right), \quad T_{26} \le \mathcal{O}\left(\alpha^2\right), \quad T_{27} \le \mathcal{O}\left(\alpha^2\right), \quad T_{28} \le \mathcal{O}\left(\alpha^2\right)\\
&T_{29} \le \mathcal{O}\left(\frac{\alpha}{k}\sum_{j=1}^{k}\left((1-\frac{\alpha}{k})^{j-1}-1\right)^2\right) \le \mathcal{O}\left(\frac{\alpha}{k}\sum_{j=1}^{k}\left(1-(1-\frac{\alpha}{k})^{j-1}\right)\right) \overset{\text{(i)}}{\in} \mathcal{O}\left(\alpha^2\right),
\end{aligned}$$

where (i) holds by equation (42). This completes the proof of Lemma 14.

## G.2 Step 2: General Stepsize

In this subsection, we aim to prove that there exists an $\alpha_0$ such that $\mathcal{L}(Y^\alpha)$ is continuous when $\alpha \in (0, \alpha_0)$ with respect to $W_2$. Here we use another coupling as follows:

$$Y_{t+1} = (1-\alpha)Y_t + \alpha\gamma D'_t P'_t f(Y_t + \frac{q^*}{\sqrt{\alpha}}) + \sqrt{\alpha}(I - D'_t)(\sqrt{\alpha}Y_t + q^*) + \sqrt{\alpha}D'_t r'_t - \sqrt{\alpha}q^*,$$

$$Y'_{t+1} = (1-\alpha')Y'_t + \alpha'\gamma D'_t P'_t f(Y'_t + \frac{q^*}{\sqrt{\alpha'}}) + \sqrt{\alpha'}(I - D'_t)(\sqrt{\alpha'}Y'_t + q^*) + \sqrt{\alpha'}D'_t r'_t - \sqrt{\alpha'}q^*.$$

Then, we obtain

$$\begin{aligned}
Y_{t+1} - Y'_{t+1} =& (1-\alpha)(Y_t - Y'_t)\\
&+ \alpha\Big(\gamma D'_t P'_t(f(Y_t + \frac{q^*}{\sqrt{\alpha'}}) + (I - D'_t)(Y_t + \frac{q^*}{\sqrt{\alpha'}}) - \gamma D'_t P'_t(f(Y'_t + \frac{q^*}{\sqrt{\alpha'}}) - (I - D'_t)(Y'_t + \frac{q^*}{\sqrt{\alpha'}})\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\alpha\gamma D_t'P_t'(f(Y_t+\frac{q^*}{\sqrt{\alpha}})-f(Y_t+\frac{q^*}{\sqrt{\alpha'}}))\\
&+(\alpha-\alpha')\gamma D_t'P_t'f(Y_t'+\frac{q^*}{\sqrt{\alpha'}})-(\alpha-\alpha')D_t'Y_t'+(\sqrt{\alpha}-\sqrt{\alpha'})D_t'(r_t'-q^*)\\
:=&(1-\alpha)(Y_t-Y_t')+A.
\end{aligned}
$$

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of above equation and by property (1) in Proposition 5, we obtain

$$
\mathbb{E}[M_\eta(Y_{t+1}-Y_{t+1}')]\le(1-\alpha)^2\mathbb{E}[M_\eta(Y_t-Y_t')]+(1-\alpha)\underbrace{\mathbb{E}\langle\nabla M_\eta(Y_t-Y_t'),A\rangle}_{T_1}+\frac{1}{2\eta}\underbrace{\mathbb{E}\|A\|_2^2}_{T_2}.
$$

Below we separately bound the $T_1$ and $T_2$ terms.

**Bounding the $T_1$ Term:** By property (4) in Proposition 5 and equation (54), we obtain

$$
\begin{aligned}
T_1\le&\underbrace{\alpha\gamma\mathbb{E}[\|Y_t-Y_t'\|_m\|\mathcal{T}(Y_t+\frac{q^*}{\sqrt{\alpha'}})-\mathcal{T}(Y_t'+\frac{q^*}{\sqrt{\alpha'}})\|_m]}_{T_{11}}\\
&+\underbrace{\alpha\gamma\mathbb{E}[\|Y_t-Y_t'\|_m\|DP(f(Y_t+\frac{q^*}{\sqrt{\alpha}})-f(Y_t+\frac{q^*}{\sqrt{\alpha'}}))\|_m]}_{T_{12}}\\
&+\underbrace{|\alpha-\alpha'|\gamma\mathbb{E}[\|Y_t-Y_t'\|_m\|Pf(Y_t'+\frac{q^*}{\sqrt{\alpha'}})\|_m]}_{T_{13}}+\underbrace{|\alpha-\alpha'|\mathbb{E}[\|Y_t-Y_t'\|_m\|DY_t'\|_m]}_{T_{14}}\\
&+\underbrace{|\sqrt{\alpha}-\sqrt{\alpha'}|\mathbb{E}[\|Y_t-Y_t'\|_m\|D(r-q^*)\|_m]}_{T_{15}}.
\end{aligned}
$$

By Corollary 2 with $n=2$, Lemma 12 and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$
T_{11}\le\frac{2u_{cm}\gamma_0\alpha}{l_{cm}}\mathbb{E}[M_\eta(Y_t-Y_t')]\le2\alpha\sqrt{\gamma}\mathbb{E}[M_\eta(Y_t-Y_t')],
$$

where the last inequality holds because we can always choose a proper $\eta$ such that $\frac{u_{cm}}{l_{cm}}\le\frac{1}{\sqrt{\gamma}}$.

Let $\delta=|\alpha-\alpha'|\le\frac{\alpha}{2}$, we obtain

$$
\begin{aligned}
&T_{12}\in\mathcal{O}\left(\alpha|\frac{1}{\sqrt{\alpha}}-\frac{1}{\sqrt{\alpha'}}|\right)\in\left(\frac{\alpha\delta}{\sqrt{\alpha}\sqrt{\alpha'}(\sqrt{\alpha}+\sqrt{\alpha'})}\right)\in\mathcal{O}\left(\frac{\alpha\delta}{\min(\alpha,\alpha')^{\frac{3}{2}}}\right).\\
&T_{13}=\mathcal{O}(1)\cdot\delta\mathbb{E}[\|Y_t-Y_t'\|_m\|Pf(Y_t'+\frac{q^*}{\sqrt{\alpha'}})\|_c]\\
&\quad\le\mathcal{O}(1)\cdot\delta\mathbb{E}[\|Y_t-Y_t'\|_m\|P(f(Y_t'+\frac{q^*}{\sqrt{\alpha'}})-f(\frac{q^*}{\sqrt{\alpha'}}))\|_c]+\mathcal{O}(1)\cdot\delta\mathbb{E}[\|Y_t-Y_t'\|_m\|Pf(\frac{q^*}{\sqrt{\alpha'}})\|_c]\\
&\quad\le\mathcal{O}(\delta)+\mathcal{O}\Big(\frac{\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\Big)\in\mathcal{O}\Big(\frac{\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\Big).\\
&T_{14}\in\mathcal{O}(\delta),\quad T_{15}\in\mathcal{O}\Big(\frac{\delta}{\min(\alpha,\alpha')^{\frac{1}{2}}}\Big).
\end{aligned}
$$

**Bounding the $T_2$ Term:** By Cauchy-Schwarz inequality, we obtain

$$
T_2\le5\Bigg(\underbrace{\alpha^2\mathbb{E}\|\gamma D_t'P_t'(f(Y_t+\frac{q^*}{\sqrt{\alpha'}})+(I-D_t')(Y_t+\frac{q^*}{\sqrt{\alpha'}})-\gamma D_t'P_t'(f(Y_t'+\frac{q^*}{\sqrt{\alpha'}})-(I-D_t')(Y_t'+\frac{q^*}{\sqrt{\alpha'}})\|_2^2}_{T_{21}}
$$

$$+\underbrace{\alpha^2\gamma^2\mathbb{E}\|D_t'P_t'(f(Y_t+\frac{q^*}{\sqrt{\alpha}})-f(Y_t+\frac{q^*}{\sqrt{\alpha'}}))\|_2^2}_{T_{22}}$$

$$+\underbrace{\delta^2\gamma^2\mathbb{E}\|D_t'P_t'f(Y_t'+\frac{q^*}{\sqrt{\alpha'}})\|_2^2}_{T_{23}}+\underbrace{\delta^2\mathbb{E}[\|D_t'Y_t'\|_2^2]}_{T_{24}}+\mathcal{O}\Big(\frac{\delta^2}{\min(\alpha,\alpha')}\Big)\Bigg).$$

By Corollary 2 with $n=2$, Lemma 12 and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$T_{21}\leq\mathcal{O}\left(\alpha^2\right)\cdot\mathbb{E}[M_\eta(Y_t-Y_t')].$$

$$T_{22}\in\mathcal{O}\left(\alpha^2|\frac{1}{\sqrt{\alpha}}-\frac{1}{\sqrt{\alpha'}}|^2\right)\in\mathcal{O}\left(\frac{\alpha^2\delta^2}{\alpha\alpha'(\sqrt{\alpha}+\sqrt{\alpha'})^2}\right)\in\mathcal{O}\left(\frac{\alpha\delta^2}{\min(\alpha,\alpha')^2}\right).$$

$$T_{23}\in\mathcal{O}\left(\frac{\delta^2}{\min(\alpha,\alpha')}\right),T_{24}\in\mathcal{O}\left(\delta^2\right).$$

Combining the above analysis together, there exist an $\alpha_0$ such that $0<\left(1-2(1-\sqrt{\gamma_0})\alpha_0+\mathcal{O}(\alpha_0^2)\right)<1$ and for any $\alpha\leq\alpha_0$, there exist $t_\alpha$ such that for any $t\geq t_\alpha$, we obtain

$$\mathbb{E}[M_\eta(Y_{t+1}-Y_{t+1}')]\leq\left(1-2(1-\gamma_0^{\frac{1}{4}})\alpha+\mathcal{O}(\alpha^2)\right)\mathbb{E}[M_\eta(Y_t-Y_t')]+\mathcal{O}\left(\frac{\alpha\delta}{\min(\alpha,\alpha')^{\frac{3}{2}}}\right).$$

Then, we obtain

$$\lim_{t\to\infty}\mathbb{E}[M_\eta(Y_t-Y_t')]\in\mathcal{O}(\frac{\delta}{\min(\alpha,\alpha')^{\frac{3}{2}}}).$$

Then,

$$\begin{aligned}W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y^{(\alpha')})\right)&\leq\lim_{t\to\infty}W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y_t)\right)+W_2\left(\mathcal{L}(Y_t),\mathcal{L}(Y_t')\right)+W_2\left(\mathcal{L}(Y_t'),\mathcal{L}(Y^{(\alpha')})\right)\\&\leq\lim_{t\to\infty}\sqrt{\mathbb{E}[\|Y_t-Y_t'\|_c^2]}\leq\lim_{t\to\infty}\sqrt{2u_{cm}^2\mathbb{E}[M_\eta(Y_t-Y_t')]}\leq\frac{c\sqrt{\delta}}{\min(\alpha,\alpha')^{\frac{3}{4}}},\end{aligned}$$

where $c$ is a universal constant that is independent with $\alpha,\alpha'$.

Then, for $\forall\epsilon>0$, given $\alpha>0$, we can choose a sufficient small $\delta_\epsilon$ such that

$$\frac{c\sqrt{\delta_\epsilon}}{(\alpha-\delta_\epsilon)^{\frac{3}{4}}}\leq\epsilon\text{ and }0<\delta_\epsilon<\frac{\alpha}{2}.$$

Then, when $\alpha'$ is selected with $|\alpha-\alpha'|\leq\delta_\epsilon$, we obtain

$$W_2\left(\mathcal{L}(x^{(\alpha)}),\mathcal{L}(x^{(\alpha')})\right)\leq\epsilon.$$

Therefore, we complete the proof of continuity of $\mathcal{L}(x^\alpha)$ w.r.t $W_2$. Then, by the same argument at the end of Section C.2, we obtain $\lim_{\alpha\to 0}W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\right)=0$, thereby completing the second step of the proof of Theorem 5.

## G.3 Step 2.5: Convergence Rate under Gaussian Noise

By triangle inequality, we obtain

$$\begin{aligned}W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\right)&\leq W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y^{(\alpha/k)})\right)+W_2\left(\mathcal{L}(Y^{(\alpha/k)}),\mathcal{L}(Y)\right)\\&\leq\mathcal{O}(\alpha^{\frac{1}{4}})+W_2\left(\mathcal{L}(Y^{(\alpha/k)}),\mathcal{L}(Y)\right)\\&\leq\lim_{k\to\infty}\mathcal{O}(\alpha^{\frac{1}{4}})+W_2\left(\mathcal{L}(Y^{(\alpha/k)}),\mathcal{L}(Y)\right)\in\mathcal{O}(\alpha^{\frac{1}{4}}),\end{aligned}\tag{67}$$

which gives the convergence rate.

### G.4 Step 3: General Noise

By Section G.1, G.2 and G.3, we prove that under the noise with normal distribution, there exists a unique random variable $Y$ such that $Y^{(\alpha)}$ converge to $Y$ with respect to $W_2$. In this subsection, we aim to prove that under general i.i.d zero mean noise with the same variance, the convergence result still holds and the limit is still $Y$.

By equation (61), we consider the following coupling:

$$
\begin{aligned}
Y_{t+1} &= (1-\alpha)Y_t + \alpha\Big(\mathcal{T}(Y_t + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(\frac{q^*}{\sqrt{\alpha}})\Big) + \alpha A_t f(Y_t + \frac{q^*}{\sqrt{\alpha}}) + \alpha B_t Y_t + \sqrt{\alpha} H_t,\\
Y'_{t+1} &= (1-\alpha)Y'_t + \alpha\Big(\mathcal{T}(Y'_t + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(\frac{q^*}{\sqrt{\alpha}})\Big) + \alpha A'_t f(Y'_t + \frac{q^*}{\sqrt{\alpha}}) + \alpha B'_t Y'_t + \sqrt{\alpha} H'_t,
\end{aligned}
$$

where $H'_t = B'_t q^* + C'_t$ and $\{(A_t, B_t, H_t)\}_{t\geq 0}$ and $\{(A'_t, B'_t, H'_t)\}_{t\geq 0}$ have zero mean and the same variance. Here $(A_t, B_t, H_t)$ and $(A'_t, B'_t, H'_t)$ are not necessary independent with each other, $\{(A'_t, B'_t, H'_t)\}_{t\geq 0}$ are normal distributed and we assume that $\{(A_t, B_t, H_t)\}_{t\geq 0}$ have finite fourth moments.

Let $\kappa = \lfloor \alpha^{-\frac{1}{2}} \rfloor$. We obtain

$$
\begin{aligned}
Y_{\kappa t+\kappa} =& (1-\alpha)^{\kappa} Y_{\kappa t} + \alpha \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} (\mathcal{T}(Y_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(\frac{q^*}{\sqrt{\alpha}}))\\
&+ \alpha \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} \Big(A_{\kappa t+\kappa-j}(f(Y_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - f(\frac{q^*}{\sqrt{\alpha}})) + B_{\kappa t+\kappa-j} Y_{\kappa t+\kappa-j}\Big)\\
&+ \sqrt{\alpha} \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} \left(A_{\kappa t+\kappa-j} f(q^*) + H_{\kappa t+\kappa-j}\right).
\end{aligned}
$$

and

$$
\begin{aligned}
Y'_{\kappa t+\kappa} =& (1-\alpha)^{\kappa} Y'_{\kappa t} + \alpha \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} (\mathcal{T}(Y'_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(\frac{q^*}{\sqrt{\alpha}}))\\
&+ \alpha \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} \Big(A'_{\kappa t+\kappa-j}(f(Y'_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - f(\frac{q^*}{\sqrt{\alpha}})) + B'_{\kappa t+\kappa-j} Y'_{\kappa t+\kappa-j}\Big)\\
&+ \sqrt{\alpha} \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} \left(A'_{\kappa t+\kappa-j} f(q^*) + H'_{\kappa t+\kappa-j}\right).
\end{aligned}
$$

Taking the difference of the last two equations, we get

$$
\begin{aligned}
Y_{\kappa t+\kappa} - Y'_{\kappa t+\kappa} =& (1-\alpha)^{\kappa}(Y_{\kappa t} - Y'_{\kappa t}) + \kappa\alpha(\mathcal{T}(Y_{\kappa t} + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(Y'_{\kappa t} + \frac{q^*}{\sqrt{\alpha}}))\\
&+ \alpha \sum_{j=1}^{\kappa} (\mathcal{T}(Y_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(Y_{\kappa t} + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(Y'_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) + \mathcal{T}(Y'_{\kappa t} + \frac{q^*}{\sqrt{\alpha}}))\\
&+ \alpha \sum_{j=1}^{\kappa} ((1-\alpha)^{j-1} - 1)(\mathcal{T}(Y_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - \mathcal{T}(Y'_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}))\\
&+ \alpha \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} \Big(A_{\kappa t+\kappa-j}(f(Y_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - f(\frac{q^*}{\sqrt{\alpha}})) + B_{\kappa t+\kappa-j} Y_{\kappa t+\kappa-j}\Big)\\
&- \alpha \sum_{j=1}^{\kappa} (1-\alpha)^{j-1} \Big(A'_{\kappa t+\kappa-j}(f(Y'_{\kappa t+\kappa-j} + \frac{q^*}{\sqrt{\alpha}}) - f(\frac{q^*}{\sqrt{\alpha}})) + B'_{\kappa t+\kappa-j} Y'_{\kappa t+\kappa-j}\Big)\\
&+ \sqrt{\alpha} \sum_{j=1}^{\kappa} \left(A_{\kappa t+\kappa-j} f(q^*) - A'_{\kappa t+\kappa-j} f(q^*) + H_{\kappa t+\kappa-j} - H'_{\kappa t+\kappa-j}\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \sqrt{\alpha}\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})\left(A_{\kappa t+\kappa-j}f(q^*)+H_{\kappa t+\kappa-j}\right)\\
&+ \sqrt{\alpha}\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})\left(A'_{\kappa t+\kappa-j}f(q^*)+H'_{\kappa t+\kappa-j}\right)\\
:=&(1-\alpha)^{\kappa}(Y_{\kappa t}-Y'_{\kappa t})+A.
\end{aligned}
$$

Applying the generalized Moreau envelope $M_\eta(\cdot)$ defined in equation (21) to both sides of above equation and by property (1) in Proposition 5, we obtain

$$
\mathbb{E}[M(Y_{\kappa t+\kappa}-Y'_{\kappa t+\kappa})] \le (1-\alpha)^{2\kappa}\mathbb{E}[M(Y_{\kappa t}-Y'_{\kappa t})]+(1-\alpha)^{\kappa}\underbrace{\mathbb{E}\langle\nabla M(Y_{\kappa t}-Y'_{\kappa t}),A\rangle}_{T_1}+\frac{1}{2\eta}\underbrace{\mathbb{E}\|A\|_2^2}_{T_2}. \tag{68}
$$

The following lemmas, proved in Sections G.4.1 and G.4.2 to follow, control the $T_1$ and $T_2$ terms above.

**Lemma 15.** *Under the setting of Theorem 5, we have*

$$
T_1 \le 2\alpha\kappa\sqrt{\gamma_0}\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]+\mathcal{O}(\alpha^{\frac{3}{2}}\kappa^{\frac{3}{2}}).
$$

**Lemma 16.** *Under the setting of Theorem 5 and some proper couplings between $\{(A_t,B_t,H_t)\}_{t\ge0}$ and $\{(A'_t,B'_t,H'_t)\}_{t\ge0}$, we have*

$$
T_2 \le \mathcal{O}(\alpha^2\kappa^2)\cdot\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]+\mathcal{O}(\alpha).
$$

Plugging the above bounds for $T_1$ and $T_2$ into equation (68), there exist an $\alpha_0$ such that for any $\alpha\le\alpha_0$, there exist $t_\alpha$ such that for any $t\ge t_\alpha$, we obtain

$$
\begin{aligned}
\mathbb{E}[M_\eta(Y_{\kappa t+\kappa}-Y'_{\kappa t+\kappa})] &\le \left((1-\alpha)^{\kappa}+2\alpha\kappa\sqrt{\gamma}+\mathcal{O}(\alpha^2)\right)\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]+\mathcal{O}(\alpha)\\
&\le (1-(1-\sqrt{\gamma})\alpha\kappa)\,\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]+\mathcal{O}(\alpha).
\end{aligned}
$$

Therefore, we obtain

$$
\lim_{t\to\infty}\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]\in\mathcal{O}(\alpha^{\frac{1}{2}}).
$$

By triangle inequality, we have

$$
\begin{aligned}
W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}((Y')^{(\alpha)})\right) &\le \lim_{t\to\infty}W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y_{\kappa t})\right)+W_2\left(\mathcal{L}(Y_{\kappa t}),\mathcal{L}(Y'_{\kappa t})\right)+W_2\left(\mathcal{L}(Y'_{\kappa t}),\mathcal{L}((Y')^{(\alpha)})\right)\\
&\le \lim_{t\to\infty}\sqrt{\mathbb{E}[\|Y_{\kappa t}-Y'_{\kappa t}\|_c^2]}\le\lim_{t\to\infty}\sqrt{2u_{cm}^2\mathbb{E}[M(Y_{\kappa t}-Y'_{\kappa t})]}\in\mathcal{O}(\alpha^{\frac{1}{4}}).
\end{aligned}
$$

Therefore, by equation (67), we obtain

$$
W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y)\right)\le W_2\left(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}((Y')^{(\alpha)})\right)+W_2\left(\mathcal{L}({Y'}^{(\alpha)}),\mathcal{L}(Y)\right)\in\mathcal{O}(\alpha^{\frac{1}{4}}),
$$

which implies

$$
\lim_{\alpha\to0}W_2(\mathcal{L}(Y^{(\alpha)}),\mathcal{L}(Y))=0.
$$

This completes the proof of the last step of Theorem 5, thereby finishing the proof of Theorem 5.

### G.4.1 Proof of Lemma 15 on $T_1$

By property (4) in Proposition 5, we obtain the bound

$$
T_1\le T_{11}+T_{12}+T_{13},
$$

where

$$T_{11} =\alpha\kappa\mathbb{E}[\|Y_{\kappa t}-Y'_{\kappa t}\|_m\|\mathcal{T}(Y_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y'_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})\|_m],$$

$$T_{12} =\alpha\mathbb{E}[\|Y_{\kappa t}-Y'_{\kappa t}\|_m\|\sum_{j=1}^{\kappa}\mathcal{T}(Y_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y'_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})+\mathcal{T}(Y'_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})\|_m],$$

$$T_{13} =\alpha\mathbb{E}[\|Y_{\kappa t}-Y'_{\kappa t}\|_m\|\sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)(\mathcal{T}(Y_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y'_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}}))\|_m].$$

Below, we bound $T_{11}\sim T_{13}$ separately. By Corollary 2(**2**), Lemma 12 and the equivalence of all norms on $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$T_{11}\leq\frac{2\alpha\kappa\gamma_0 u_{cm}}{l_{cm}}\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]\leq 2\alpha\kappa\sqrt{\gamma_0}\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})].$$

where the last inequality holds because we can always choose a proper $\eta$ such that $\frac{u_{cm}}{l_{cm}}\leq\frac{1}{\sqrt{\gamma_0}}$.

Similarly to the bound for $T_{12}$ in Section G.1, we obtain

$$\begin{aligned}
&T_{12}\in\mathcal{O}(\alpha^{\frac{3}{2}}\kappa^{\frac{3}{2}}),\\
&T_{13}\leq\mathcal{O}\left(\alpha\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})\right)\leq\mathcal{O}\left(1-(1-\alpha)^{\kappa}-\alpha\kappa\right)\overset{\text{(i)}}{\in}\mathcal{O}(\alpha^2\kappa^2),
\end{aligned}$$

where (i) holds by equation (42).

Combining the bound for $T_{11}\sim T_{13}$ together, we obtain

$$T_1\leq 2\alpha\kappa\sqrt{\gamma_0}\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]+\mathcal{O}(\alpha^{\frac{3}{2}}\kappa^{\frac{3}{2}}),$$

thereby completing the proof of Lemma 15.

### G.4.2 Proof of Lemma 16 on $T_2$

By Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
T_2\leq 8\Bigg(&\alpha^2\kappa^2\mathbb{E}\|\mathcal{T}(Y_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y'_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})\|_2^2 && (T_{21})\\
&+\alpha^2\mathbb{E}\|\sum_{j=1}^{\kappa}\mathcal{T}(Y_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y'_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})+\mathcal{T}(Y'_{\kappa t}+\frac{q^*}{\sqrt{\alpha}})\|_2^2 && (T_{22})\\
&+\alpha^2\mathbb{E}\|\sum_{j=1}^{\kappa}((1-\alpha)^{j-1}-1)\mathcal{T}(Y_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})-\mathcal{T}(Y'_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})\|_2^2 && (T_{23})\\
&+\alpha^2\mathbb{E}\|\sum_{j=1}^{\kappa}(1-\alpha)^{j-1}\Big(A_{\kappa t+\kappa-j}(f(Y_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})-f(\frac{q^*}{\sqrt{\alpha}}))+B_{\kappa t+\kappa-j}Y_{\kappa t+\kappa-j}\Big)\|_2^2 && (T_{24})\\
&+\alpha^2\mathbb{E}\|\sum_{j=1}^{\kappa}(1-\alpha)^{j-1}\Big(A'_{\kappa t+\kappa-j}(f(Y'_{\kappa t+\kappa-j}+\frac{q^*}{\sqrt{\alpha}})-f(\frac{q^*}{\sqrt{\alpha}}))+B'_{\kappa t+\kappa-j}Y'_{\kappa t+\kappa-j}\Big)\|_2^2 && (T_{25})\\
&+\alpha\mathbb{E}\|\sum_{j=1}^{\kappa}\left(A_{\kappa t+\kappa-j}f(q^*)-A'_{\kappa t+\kappa-j}f(q^*)+H_{\kappa t+\kappa-j}-H'_{\kappa t+\kappa-j}\right)\|_2^2 && (T_{26})\\
&+\alpha\mathbb{E}\|\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})(A_{\kappa t+\kappa-j}f(q^*)+H_{\kappa t+\kappa-j})\|_2^2 && (T_{27})
\end{aligned}$$

$$+\alpha\mathbb{E}\|\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})(A'_{\kappa t+\kappa-j}f(q^*)+H'_{\kappa t+\kappa-j})\|_2^2\Bigg). \tag{$T_{28}$}$$

By Corollary 2(**2**), Lemma 12 and equivalence of norms in $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain the following bounds:

$$T_{21}\leq\mathcal{O}(\alpha^2\kappa^2)\cdot\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})],\quad T_{22}\in\mathcal{O}(\alpha^2\kappa^2)$$

$$T_{23}\leq\mathcal{O}(\alpha^2\kappa\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})^2)\in\mathcal{O}(\alpha^3\kappa^3),\quad T_{24}\leq\mathcal{O}(\alpha^2\kappa\sum_{j=1}^{\kappa}(1-\alpha)^{2j-2})\in\mathcal{O}(\alpha^2\kappa^2)$$

$$T_{25}\leq\mathcal{O}(\alpha^2\kappa\sum_{j=1}^{\kappa}(1-\alpha)^{2j-2})\in\mathcal{O}(\alpha^2\kappa^2),\quad T_{27}\leq\mathcal{O}(\alpha\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})^2)\in\mathcal{O}(\alpha^2\kappa^2)$$

$$T_{28}\leq\mathcal{O}(\alpha\sum_{j=1}^{\kappa}(1-(1-\alpha)^{j-1})^2)\in\mathcal{O}(\alpha^2\kappa^2).$$

For $T_{26}$, we can notice that $(A'_{\kappa t+\kappa-j}f(q^*),H'_{\kappa t+\kappa-j})$ is normal distributed. Then, similarly to the analysis of $T_{24}$ in Section C.4, we can find a coupling between $\{A_t,B_t,H_t\}_{t\geq 0}$ and $\{A'_t,B'_t,H'_t\}_{t\geq 0}$ such that $T_{26}\in\mathcal{O}(\alpha)$.

Recall that $\kappa=\lfloor\alpha^{-\frac{1}{2}}\rfloor$, we obtain

$$T_2\leq\mathcal{O}(\alpha^2\kappa^2)\cdot\mathbb{E}[M_\eta(Y_{\kappa t}-Y'_{\kappa t})]+\mathcal{O}(\alpha),$$

thereby completing the proof of Lemma 16.

# H Proof of Theorem 6

By Theorem 4 and equation (62), we have the following equalities in distribution:

$$\begin{aligned}Y^{(\alpha)}\overset{\mathrm{d}}{=}&(1-\alpha)Y^{(\alpha)}+\alpha\mathcal{T}_0(Y^{(\alpha)})+\alpha\gamma DPh(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})+\alpha A_0h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})\\&+\alpha A_0g(Y^{(\alpha)})+\alpha B_0Y^{(\alpha)}+\sqrt{\alpha}E_0,\end{aligned}\tag{69}$$

where $\mathcal{T}_0(q)=\gamma DPg(q)+(I-D)q$.

Taking expectation to both sides of the above equation, we obtain

$$\mathbb{E}[Y^{(\alpha)}]=\mathbb{E}[\mathcal{T}_0(Y^{(\alpha)})+\gamma DPh(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})].$$

Rearranging terms, we obtain the equality

$$\mathbb{E}[Y^{(\alpha)}]=\mathbb{E}[\gamma Pg(Y^{(\alpha)})+\gamma Ph(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})].$$

By Fatou's lemma [Dur19, Exercise 3.2.4] and Lemma 12, we obtain

$$\|E[h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})]\|_c\leq E[\|h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})\|_c]\leq\sqrt{E[\|h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})\|_c^2]}\in\mathcal{O}(\sqrt{\alpha}).$$

It is well known that weak convergence in $\mathcal{P}_2(\mathbb{R}^{|\mathcal{S}||\mathcal{A}|})$ is equivalent to convergence in distribution and the convergence of the first two moments. By [Dur19, Exercise 3.2.5] and the Lipschitz continuity of $\mathcal{T}_0(\cdot)$, we obtain

$$\lim_{\alpha\to 0}\mathbb{E}[Y^{(\alpha)}]=\mathbb{E}[Y]\text{ and }\lim_{\alpha\to 0}\mathbb{E}[\mathcal{T}_0(Y^{(\alpha)})]=\mathbb{E}[\mathcal{T}_0(Y)].$$

Therefore, we have

$$\mathbb{E}[Y]=\gamma P\mathbb{E}[g(Y)].\tag{70}$$

Below, we discuss the $\mathbb{E}[Y]$ in three cases.

**Case 1:** If there exists a state $s'$ that is both tied and not rooted. Then, if $\mathbb{E}[Y]=0$, then because $g_s(Y)=\max_{a\in\mathcal{A}^*(s)} Y(s,a)$, we obtain

$$\mathbb{E}[g_s(Y)] \geq \mathbb{E}[Y(s,a)] = 0, \quad \forall a \in \mathcal{A}^*(s),$$

which implies that $\mathbb{E}[g(Y)] \geq 0$.

Because $s'$ is a non-rooted state, $\exists s, a$ such that $P(s'|s,a)>0$. Therefore, by $\mathbb{E}[g(Y)]\geq 0$ and equation (70), we have

$$0 = \mathbb{E}[Y(s,a)] \geq \gamma P(s'|s,a)\mathbb{E}[g_{s'}(Y)].$$

Then, we have $\mathbb{E}[g_{s'}(Y)]=0$.

Let $\mathcal{A}^*(s')=\{a_1,a_2,\ldots,a_N\}$. Then, we have $\mathbb{E}[Y(s',a_i)]=0$ for all $i\in[N]$ and $\mathbb{E}[\max_{i\in[N]} Y(s',a_i)]=0$. Therefore, we obtain $Y(s',a_i)=Y(s',a_j)$ a.e. for all $i\neq j\in[N]$ by [Dur19, Exercises 1.4.1]. By Fatou's lemma, we have

$$\mathbb{E}[Y(s',a_i)^2] \leq \liminf_{\alpha\to 0} \mathbb{E}[{Y^{(\alpha)}(s',a_i)}^2] = \liminf_{\alpha\to 0}\lim_{t\to\infty} \mathbb{E}[{Y_t^{(\alpha)}(s',a_i)}^2] < +\infty.$$

Because $(Y(s',a_1)-Y(s',a_2))^2 \leq 2Y(s',a_1)^2+2Y(s',a_2)^2$, by dominated convergence theorem, we have

$$\mathbb{E}[(Y(s',a_1)-Y(s',a_2))^2]=0.$$

By equation (69), Corollary 2(**2**), Lemma 12 and equivalence of norms in $\mathbb{R}^{|\mathcal{S}||\mathcal{A}|}$, we obtain

$$\begin{aligned}
&\mathbb{E}[(Y^{(\alpha)}(s',a_1)-Y^{(\alpha)}(s',a_2))^2] \\
&= (1-\alpha)^2\mathbb{E}[(Y^{(\alpha)}(s',a_1)-Y^{(\alpha)}(s',a_2))^2] + o(\alpha) + \alpha\mathbb{E}[(E_0(s',a_1)-E_0(s',a_2))^2].
\end{aligned}$$

Rearranging terms, we obtain the equality

$$(2-\alpha)\mathbb{E}[(Y^{(\alpha)}(s',a_1)-Y^{(\alpha)}(s',a_2))^2] = o(1)+\mathbb{E}[(E_0(s',a_1)-E_0(s',a_2))^2].$$

Letting $\alpha$ go to 0, we obtain

$$2\mathbb{E}[(Y(s',a_1)-Y(s',a_2))^2] = \mathbb{E}[(E_0(s',a_1)-E_0(s',a_2))^2].$$

Recall that $E_0 = A_0 f(q^*)+B_0q^*+C_0 = \gamma D_0P_0f(q^*)+D_0r_0-D_0q^*$ Therefore, have $\mathrm{Var}(E_0)$ is positive definite, which implies $\mathbb{E}[(Y(s',a_1)-Y(s',a_2))^2]>0$ and contradicts with the fact $\mathbb{E}[(Y(s',a_1)-Y(s',a_2))^2]=0$. Then, we can conclude that $\mathbb{E}[Y]\neq 0$.

**Case 2:** If there is no tied state, by definition, $g(\cdot)$ will be a linear function. Recall that

$$\mathbb{E}[Y^{(\alpha)}] = \gamma P\mathbb{E}\left[g(Y^{(\alpha)})+h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})\right].$$

For $n\geq 2$, by equation (63) in Lemma 12, Assumption 4(**n**) and the above equation, we obtain

$$\begin{aligned}
\|\mathbb{E}[Y^{(\alpha)}]\|_c &\leq \|\gamma P\mathbb{E}[g(Y^{(\alpha)})]\|_c + \|\gamma P\mathbb{E}[h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})]\|_c \\
&\leq \gamma\|g(\mathbb{E}[Y^{(\alpha)}])\|_c + \gamma\|\mathbb{E}[h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})]\|_c \\
&\leq \gamma\|\mathbb{E}[Y^{(\alpha)}]\|_c + \gamma\|\mathbb{E}[h(Y^{(\alpha)},\frac{q^*}{\sqrt{\alpha}})]\|_c \\
&\leq \gamma\|\mathbb{E}[Y^{(\alpha)}]\|_c + \gamma\left(\mathbb{E}\|f(Y^{(\alpha)})-g(Y^{(\alpha)})\|_c^{2n}\right)^{\frac{1}{2n}}\cdot\left(\mathbb{P}(\|Y^{(\alpha)}\|_c\geq\frac{\Delta}{2\sqrt{\alpha}})\right)^{\frac{2n-1}{2n}}
\end{aligned}$$

We continue by bounding the second right hand term and obtain

$$
\begin{aligned}
\|\mathbb{E}[Y^{(\alpha)}]\|_c &\overset{(i)}{\leq} \gamma\|\mathbb{E}[Y^{(\alpha)}]\|_c + \mathcal{O}\left(\left(\mathbb{P}(\|Y^{(\alpha)}\|_c \geq \frac{\Delta}{2\sqrt{\alpha}})\right)^{\frac{2n-1}{2n}}\right) \\
&\leq \gamma\|\mathbb{E}[Y^{(\alpha)}]\|_c + \mathcal{O}\left(\left(\frac{\mathbb{E}[\|Y^{(\alpha)}\|_c^{2n}]4^n\alpha^n}{\Delta^{2n}}\right)^{\frac{2n-1}{2n}}\right) \overset{(ii)}{\leq} \gamma\|\mathbb{E}[Y^{(\alpha)}]\|_c + \mathcal{O}\left(\alpha^{\frac{2n-1}{2}}\right),
\end{aligned}
$$

where (i) and (ii) hold by Corollary 2 with $n = 2$ and Fatou's lemma.

Therefore, we have

$$
\mathbb{E}[Y^{(\alpha)}] \in \mathcal{O}\left(\alpha^{\frac{2n-1}{2}}\right),
$$

which implies $\mathbb{E}[Y] = 0$. Furthermore, recall that $\mathbb{E}[Y^{(\alpha)}] = \frac{\mathbb{E}[q^{(\alpha)}]-q^*}{\sqrt{\alpha}}$ and we obtain

$$
\mathbb{E}[q^{(\alpha)}] = q^* + \mathcal{O}(\alpha^n).
$$

**Case 3:** If tied states are always rooted states, the MDP other than all these tied and rooted states will form a new MDP with no tied state. Then, for any state $s$ and action $a$ in the new MDP, we have proved that $\mathbb{E}[Y^{(\alpha)}(s,a)] \in \mathcal{O}\left(\alpha^{\frac{2n-1}{2}}\right)$. We notice that, for any state $s$ that is tied and rooted, the next state $s'$ is always in the new MDP by definition of rooted state. Therefore, for any state $s$ that is tied and rooted and action $a$, we obtain

$$
\begin{aligned}
\mathbb{E}[Y^{(\alpha)}(s,a)] &= \gamma\sum_{s'} P(s'|s,a)\left(E[g(Y^{(\alpha)})][s'] + \mathbb{E}[h(Y^{(\alpha)}, \frac{q^*}{\sqrt{\alpha}})][s']\right) \\
&= \gamma\sum_{s'} P(s'|s,a)\left(g(E[Y^{(\alpha)}])[s'] + \mathbb{E}[h(Y^{(\alpha)}, \frac{q^*}{\sqrt{\alpha}})][s']\right) \in \mathcal{O}\left(\alpha^{\frac{2n-1}{2}}\right).
\end{aligned}
$$

Then, we conclude that $\mathbb{E}[Y] = 0$ and $\mathbb{E}[q^{(\alpha)}] = q^* + \mathcal{O}(\alpha^n)$.

# I Proof of Proposition 3

In this section, we provide the proof of the first and second moment bounds in Proposition 3. Firstly, we provide the following lemma.

**Lemma 17.** *For iterates $\theta_t^{(\alpha)}$ that are generated by equation (14) and satisfy the Condition 1 and 2, there exist two universal constants $C_2$ and $C_3$ such that*

1. $\mathbb{E}[\|\theta^{(\alpha)}\|_2] \in \mathcal{O}(1)$ *and* $\mathrm{Var}(\theta^{(\alpha)}) \in \mathcal{O}(\alpha\tau_\alpha)$.
2. $\|\mathbb{E}[\theta_t^{(\alpha)}] - \mathbb{E}[\theta^{(\alpha)}]\|_2 \leq C_2 \cdot (1-\alpha C_1)^{\frac{t}{2}}, \quad \forall \alpha \leq \bar{\alpha}$ *and* $t \geq \tau_\alpha$.
3. $\|\mathbb{E}[\theta_t^{(\alpha)}{\theta_t^{(\alpha)}}^T] - \mathbb{E}[\theta^{(\alpha)}{\theta^{(\alpha)}}^T]\|_2 \leq C_3 \cdot (1-\alpha C_1)^{\frac{t}{2}}, \quad \forall \alpha \leq \bar{\alpha}$ *and* $t \geq \tau_\alpha$.

*Proof of Lemma 17.* By Condition 1, we obtain

$$
\begin{aligned}
\mathbb{E}[\|\theta^{(\alpha)}\|_2^2] &\leq \underbrace{2\mathbb{E}(\|\theta^{(\alpha)} - \theta^*\|_2^2)}_{\in\mathcal{O}(\alpha)} + \underbrace{2\|\theta^*\|_2^2}_{\in\mathcal{O}(1)} \in \mathcal{O}(1), \\
\|\mathrm{Var}(\theta^{(\alpha)})\|_2 &\leq \mathbb{E}[\|\theta^{(\alpha)} - \theta^*\|_2^2] \in \mathcal{O}(\alpha\tau_\alpha).
\end{aligned}
$$

By [Vil09, Theorem 4.1], there exists a coupling between $\theta_t$ and $\theta^{(\alpha)}$ such that

$$
W_2^2(\mathcal{L}(\theta_t), \mathcal{L}(\theta^{(\alpha)})) = \mathbb{E}[\|\theta_t - \theta^{(\alpha)}\|_c^2].
$$

By the above bounds and applying Jensen's inequality twice, we obtain that

$$\begin{aligned}\|\mathbb{E}[\theta_t - \theta^{(\alpha)}]\|_2^2 &\le \Big(\mathbb{E}[\|\theta_t - \theta^{(\alpha)}\|_2]\Big)^2 \le \mathbb{E}[\|\theta_t - \theta^{(\alpha)}\|_2^2] \\ &\le \frac{1}{l_{cs}^2}\mathbb{E}[\|\theta_t - \theta^{(\alpha)}\|_c^2] \le \frac{C_0}{l_{cs}^2}(1-\alpha C_1)^t.\end{aligned}$$

We thus have

$$\|\mathbb{E}[\theta_t] - \mathbb{E}[\theta^{(\alpha)}]\|_2 \le \mathbb{E}[\|\theta_t - \theta^{(\alpha)}\|_2] \le C_2 \cdot (1-\alpha C_1)^{\frac{t}{2}}.$$

For the second moment, by [HCX23b, Equation A.28], we obtain

$$\begin{aligned}\left\|\mathbb{E}\left[\theta_t\theta_t^\top\right] - \mathbb{E}\left[\theta^{(\alpha)}\theta^{(\alpha)\top}\right]\right\|_2 &\le \mathbb{E}\left[\left\|\theta_t - \theta^{(\alpha)}\right\|_2^2\right] + 2\left(\mathbb{E}\left[\left\|\theta_t - \theta^{(\alpha)}\right\|_2^2\right]\mathbb{E}\left[\left\|\theta^{(\alpha)}\right\|_2^2\right]\right)^{1/2} \\ &\le \mathbb{E}\left[\left\|\theta_t - \theta^{(\alpha)}\right\|_2^2\right] + 2\left(\mathbb{E}\left[\left\|\theta_t - \theta^{(\alpha)}\right\|_2^2\right]\mathbb{E}\left[2\left\|\theta^{(\alpha)} - \theta^*\right\|_2^2 + 2\left\|\theta^*\right\|_2^2\right]\right)^{1/2}\end{aligned} \tag{71}$$

Meanwhile, we have

$$\mathbb{E}\left[\left\|\theta_t - \theta^{(\alpha)}\right\|_2^2\right] \le \frac{C_0}{l_{cs}^2}(1-\alpha C_1)^t \quad \text{and} \quad \mathbb{E}\left[\left\|\theta^{(\alpha)} - \theta^*\right\|_2^2\right] \in \mathcal{O}(\alpha\tau_\alpha).$$

Substituting the above bounds into the right-hand side of inequality (71) yields

$$\left\|\mathbb{E}\left[\theta_t\theta_t^\top\right] - \mathbb{E}\left[\theta^{(\alpha)}\theta^{(\alpha)\top}\right]\right\|_2 \le C_3 \cdot (1-\alpha C_1)^{\frac{t}{2}}.$$

□

## I.1 First Moment

First, we have

$$\mathbb{E}\left[\bar{\theta}_{k_0,k}\right] - \theta^* = \left(\mathbb{E}\left[\theta^{(\alpha)}\right] - \theta^*\right) + \frac{1}{k-k_0}\sum_{t=k_0}^{k-1}\mathbb{E}\left[\theta_t - \theta^{(\alpha)}\right].$$

To bound the sum on the right hand side, we use Lemma 17 to obtain

$$\|\mathbb{E}[\theta_k] - \mathbb{E}[\theta^{(\alpha)}]\|_2 \le C_2 \cdot (1-\alpha C_1)^{\frac{k}{2}}.$$

It follows that

$$\begin{aligned}\left\|\sum_{t=k_0}^{k-1}\mathbb{E}\left[\theta_t - \theta^{(\alpha)}\right]\right\|_2 &\le \sum_{t=k_0}^{k-1}\left\|\mathbb{E}\left[\theta_t\right] - \mathbb{E}[\theta^{(\alpha)}]\right\|_2 \\ &\le C_2 \cdot (1-\alpha C_1)^{\frac{k_0}{2}}\frac{1}{1-\sqrt{1-\alpha C_1}} \le C_2 \cdot (1-\alpha C_1)^{\frac{k_0}{2}}\frac{2}{\alpha C_1} \\ &\overset{(i)}{\le} C_2 \cdot \exp\Big(-\frac{\alpha C_1 k_0}{2}\Big)\frac{2}{(1-\sqrt{\gamma})\alpha} \le C \cdot \frac{1}{\alpha}\cdot\exp\Big(-\frac{\alpha C_1 k_0}{2}\Big),\end{aligned}$$

where $(i)$ follows from the inequality that $(1-u)^m \le \exp(-um)$ for $0 < u < 1$.

Together with Condition 2 we have

$$\mathbb{E}\left[\bar{\theta}_{k_0,k}\right] - \theta^* = \alpha^\beta B + o(\alpha^{\beta+\delta}) + \mathcal{O}\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)\right),$$

thereby finishing the proof of Proposition 3 for the first moment.

## I.2 Second Moment

In this subsection, we follow the proof technique in [HCX23b, Section A.6.2] to bound the second moment of the tail-averaged iterate. Here we use the same decomposition:

$$
\begin{aligned}
&\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\theta^*\right)\left(\bar{\theta}_{k_0,k}-\theta^*\right)^\top\right]\\
=&\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\mathbb{E}\left[\theta^{(\alpha)}\right]+\mathbb{E}\left[\theta^{(\alpha)}\right]-\theta^*\right)\left(\bar{\theta}_{k_0,k}-\mathbb{E}\left[\theta^{(\alpha)}\right]+\mathbb{E}\left[\theta^{(\alpha)}\right]-\theta^*\right)^\top\right]\\
=&\underbrace{\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\mathbb{E}\left[\theta^{(\alpha)}\right]\right)\left(\bar{\theta}_{k_0,k}-\mathbb{E}\left[\theta^{(\alpha)}\right]\right)^\top\right]}_{T_1}+\underbrace{\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\mathbb{E}\left[\theta^{(\alpha)}\right]\right)\left(\mathbb{E}\left[\theta^{(\alpha)}\right]-\theta^*\right)^\top\right]}_{T_2}\\
&+\underbrace{\mathbb{E}\left[\left(\mathbb{E}\left[\theta^{(\alpha)}\right]-\theta^*\right)\left(\bar{\theta}_{k_0,k}-\mathbb{E}\left[\theta^{(\alpha)}\right]\right)^\top\right]}_{T_3}+\underbrace{\mathbb{E}\left[\left(\mathbb{E}\left[\theta^{(\alpha)}\right]-\theta^*\right)\left(\mathbb{E}\left[\theta^{(\alpha)}\right]-\theta^*\right)^\top\right]}_{T_4}.
\end{aligned}
$$

For $T_2$, we have

$$
\begin{aligned}
T_2=&\frac{1}{k-k_0}\left(\sum_{t=k_0}^{k-1}\mathbb{E}\left[\theta_t-\theta^{(\alpha)}\right]\right)\left(\mathbb{E}[\theta^{(\alpha)}]-\theta^*\right)^\top\\
=&\mathcal{O}\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)\right)\cdot(\alpha^d B+o(\alpha^{\beta+\delta}))\in\mathcal{O}\left(\frac{\alpha^{\beta-1}}{(k-k_0)}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)\right).
\end{aligned}
$$

The term $T_3$ is similar to $T_2$ and obeys the same bound.

For $T_4$, we have

$$
T_4=(\alpha^\beta B+o(\alpha^{\beta+\delta}))(\alpha^\beta B+o(\alpha^{\beta+\delta}))^T=\alpha^{2\beta}BB^T+o(\alpha^{2\beta+\delta}).
$$

For $T_1$, we have

$$
\begin{aligned}
T_1=&\frac{1}{(k-k_0)^2}\mathbb{E}\Big[\Big(\sum_{t=k_0}^{k-1}\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)\Big)\Big(\sum_{t=k_0}^{k-1}\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)\Big)^\top\Big]\\
=&\frac{1}{(k-k_0)^2}\sum_{t=k_0}^{k-1}\mathbb{E}\left[\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)^\top\right] && (72)\\
&+\frac{1}{(k-k_0)^2}\sum_{t=k_0}^{k-1}\sum_{l=t+1}^{k-1}\mathbb{E}\left[\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)\left(\theta_l-\mathbb{E}[\theta^{(\alpha)}]\right)^\top\right] && (73)\\
&+\frac{1}{(k-k_0)^2}\sum_{t=k_0}^{k-1}\sum_{l=t+1}^{k-1}\mathbb{E}\left[\left(\theta_l-\mathbb{E}[\theta^{(\alpha)}]\right)\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)^\top\right]. && (74)
\end{aligned}
$$

By Lemma 17, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)\left(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\right)^\top\Big]\\
=&\Big(\mathbb{E}\big[\theta_t\theta_t^\top\big]-\mathbb{E}\big[\theta^{(\alpha)}{\theta^{(\alpha)}}^\top\big]\Big)+\Big(\mathbb{E}\big[\theta^{(\alpha)}{\theta^{(\alpha)}}^\top\big]-\mathbb{E}[\theta^{(\alpha)}]\mathbb{E}\big[{\theta^{(\alpha)}}^\top\big]\Big)\\
&-\Big(\mathbb{E}\left[\theta_t\right]\mathbb{E}\big[{\theta^{(\alpha)}}^\top\big]+\mathbb{E}[\theta^{(\alpha)}]\mathbb{E}\left[\theta_t^\top\right]-2\mathbb{E}[\theta^{(\alpha)}]\mathbb{E}\big[{\theta^{(\alpha)}}^\top\big]\Big)\\
=&\Big(\mathbb{E}\left[\theta_t\theta_t^\top\right]-\mathbb{E}\big[\theta^{(\alpha)}{\theta^{(\alpha)}}^\top\big]\Big)+\mathrm{Var}\left(\theta^{(\alpha)}\right)-\mathbb{E}\big[\theta_t-\theta^{(\alpha)}\big]\mathbb{E}\big[{\theta^{(\alpha)}}^\top\big]-\mathbb{E}[\theta^{(\alpha)}]\mathbb{E}\Big[(\theta_t-\theta^{(\alpha)})^\top\Big]\\
\in&\mathcal{O}\left((1-\alpha C_1)^{\frac{t}{2}}+\alpha\tau_\alpha\right).
\end{aligned}
$$

Therefore, for the term in (72), we have

$$
\begin{aligned}
(72) \in & \frac{1}{(k-k_0)^2}\sum_{t=k_0}^{k-1}\mathcal{O}\left((1-\alpha C_1)^{\frac{t}{2}}+\alpha\tau_\alpha\right)\\
\in & \,\mathcal{O}\Bigg(\frac{1}{(k-k_0)^2}\sum_{t=k_0}^{\infty}(1-\alpha C_1)^{\frac{t}{2}}\Bigg)+\mathcal{O}\left(\frac{\alpha\tau_\alpha}{k-k_0}\right)\\
\in & \,\mathcal{O}\Bigg(\frac{1}{\alpha(k-k_0)^2}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)+\frac{\alpha\tau_\alpha}{k-k_0}\Bigg).
\end{aligned}
$$

We then restate the following claim, whose proof is almost the same as the proof of Claim 4 in [HCX23b].

**Claim 1.** *For $t \geq \frac{2}{\alpha C_1}\log\left(\frac{1}{\alpha\tau_\alpha}\right)$ and $l \geq t+\tau_\alpha$, we have*

$$
\left\|\mathbb{E}\Big[\big(\theta_t-\mathbb{E}[\theta^{(\alpha)}]\big)\big(\theta_l-\mathbb{E}[\theta^{(\alpha)}]\big)^\top\Big]\right\| \in \mathcal{O}\left((\alpha\tau_\alpha)\cdot(1-\alpha C_1)^{\frac{(l-t)}{2}}\right).
$$

Then, by [HCX23b, Claim 4], we have term (73) $\in \mathcal{O}\big(\frac{\tau_\alpha}{k-k_0}\big)$. Similarly, we have term (74)$\in \mathcal{O}\big(\frac{\tau_\alpha}{k-k_0}\big)$. Therefore, we have

$$
T_1 \in \mathcal{O}\Bigg(\frac{1}{\alpha(k-k_0)^2}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)+\frac{\tau_\alpha}{k-k_0}\Bigg). \tag{75}
$$

By adding $T_1 \sim T_4$ together, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\left(\bar{\theta}_{k_0,k}-\theta^*\right)\left(\bar{\theta}_{k_0,k}-\theta^*\right)^\top\right] =&\alpha^{2\beta}BB^T+o(\alpha^{2\beta+\delta})+\mathcal{O}\Bigg(\frac{\alpha^{\beta-1}}{(k-k_0)}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)\Bigg)\\
&+\mathcal{O}\Bigg(\frac{1}{\alpha(k-k_0)^2}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)+\frac{\tau_\alpha}{k-k_0}\Bigg)\\
=&\alpha^{2\beta}BB^T+o(\alpha^{2\beta+\delta})+\mathcal{O}\Bigg(\frac{1}{\alpha(k-k_0)^2}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)+\frac{\tau_\alpha}{k-k_0}\Bigg)
\end{aligned}
$$

# J Proof of Proposition 4

We prove the first and second moment bounds in Proposition 4.

## J.1 First Moment

By equation (16), we obtain

$$
\begin{aligned}
\mathbb{E}\left[\tilde{\theta}^{(\alpha)}_{k_0,k}\right]-\theta^* =&\mathbb{E}\left[\frac{2^\beta}{2^\beta-1}\bar{\theta}^{(\alpha)}_{k_0,k}-\frac{1}{2^\beta-1}\bar{\theta}^{(2\alpha)}_{k_0,k}\right]-\theta^*\\
=&\frac{2^\beta}{2^\beta-1}\mathbb{E}\left[\bar{\theta}^{(\alpha)}_{k_0,k}-\theta^*\right]-\frac{1}{2^\beta-1}\mathbb{E}\left[\bar{\theta}^{(2\alpha)}_{k_0,k}-\theta^*\right]\\
=&\frac{2^\beta}{2^\beta-1}\left(\alpha^\beta B+o(\alpha^{\beta+\delta})+\mathcal{O}\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)\right)\right)\\
&-\frac{1}{2^\beta-1}\left((2\alpha)^\beta B+o(\alpha^{\beta+\delta})+\mathcal{O}\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\alpha C_1 k_0\right)\right)\right)\\
\in& o(\alpha^{\beta+\delta})+\mathcal{O}\left(\frac{1}{\alpha(k-k_0)}\exp\left(-\frac{\alpha C_1 k_0}{2}\right)\right).
\end{aligned}
$$

### J.2 Second Moment

Let $u_1 := \bar{\theta}^{(\alpha)}_{k_0,k} - \mathbb{E}\left[\theta^{(\alpha)}\right]$, $u_2 := \bar{\theta}^{(2\alpha)}_{k_0,k} - \mathbb{E}\left[\theta^{(2\alpha)}\right]$ and $v := \frac{2^\beta}{2^\beta-1}\mathbb{E}\left[\theta^{(\alpha)}\right] - \frac{1}{2^\beta-1}\mathbb{E}\left[\theta^{(2\alpha)}\right] - \theta^*$.

With these notations, $\tilde{\theta}_{k_0,k} - \theta^* = \frac{2^\beta}{2^\beta-1}u_1 - \frac{1}{2^\beta-1}u_2 + v$. We then have the following bound

$$\begin{aligned}\left\|\mathbb{E}\left[\left(\tilde{\theta}^{(\alpha)}_{k_0,k} - \theta^*\right)\left(\tilde{\theta}^{(\alpha)}_{k_0,k} - \theta^*\right)^\top\right]\right\|_2 \leq& \theta \left\|\mathbb{E}\left[\left(\tilde{\theta}^{(\alpha)}_{k_0,k} - \theta^*\right)\left(\tilde{\theta}^{(\alpha)}_{k_0,k} - \theta^*\right)^\top\right]\right\|_2 \\ \leq& \mathbb{E}\left[\left\|\frac{2^\beta}{2^\beta-1}u_1 - \frac{1}{2^\beta-1}u_2 + v\right\|_2^2\right] \\ \leq& 3\mathbb{E}\left\|\frac{2^\beta}{2^\beta-1}u_1\right\|_2^2 + 3\mathbb{E}\left\|\frac{1}{2^\beta-1}u_2\right\|_2^2 + 3\|v\|_2^2.\end{aligned}$$

By equation (75), we have

$$\mathbb{E}\left\|u_1\right\|_2^2 = \operatorname{Tr}\left(\mathbb{E}\left[u_1 u_1^\top\right]\right) \in \mathcal{O}\left(\frac{1}{\alpha\left(k-k_0\right)^2}\exp\left(-\frac{\alpha C_1 k_0}{2}\right) + \frac{\tau_\alpha}{k-k_0}\right).$$

Similarly, we have

$$\mathbb{E}\left\|u_2\right\|_2^2 \in \mathcal{O}\left(\frac{1}{\alpha\left(k-k_0\right)^2}\exp\left(-\frac{\alpha C_1 k_0}{2}\right) + \frac{\tau_\alpha}{k-k_0}\right).$$

By Condition 2, we have $\|v\|_2^2 = o(\alpha^{2\beta+2\delta})$.

Combining these bounds together, we have

$$\mathbb{E}\Big[\big(\tilde{\theta}_{k-k_0} - \theta^*\big)\big(\tilde{\theta}_{k-k_0} - \theta^*\big)^\top\Big] \in o(\alpha^{2\beta+2\delta}) + \mathcal{O}\Big(\frac{1}{\alpha\left(k-k_0\right)^2}\exp\big(-\frac{\alpha C_1 k_0}{2}\big) + \frac{2^{2\beta}}{(2^\beta-1)^2}\frac{\tau_\alpha}{k-k_0}\Big).$$