\textbf{N-Timescale Stochastic Approximation: Stability and Convergence}

\begin{flushright}
Rohan Deb  
Department of Computer Science and Automation,  
Indian Institute of Science, Bangalore  
rohandeb@iisc.ac.in

Shalabh Bhatnagar  
Department of Computer Science and Automation,  
Indian Institute of Science, Bangalore  
shalabh@iisc.ac.in
\end{flushright}

\textbf{Abstract}

This paper presents the first sufficient conditions that guarantee the stability and almost sure convergence of \(N\)-timescale stochastic approximation (SA) iterates for any \(N \geq 1\). It extends the existing results on One-timescale and Two-timescale SA iterates to general \(N\)-timescale stochastic recursions using the ordinary differential equation (ODE) method. As an application of our results, we study SA algorithms with an added heavy ball momentum term in the context of Gradient Temporal Difference (GTD) algorithms. We show that, when the momentum parameters are chosen in a certain way, the schemes are stable and convergent to the same solution using our proposed results.

\textbf{Keywords}  
Stochastic approximation algorithms, \(N\)-timescale stochastic approximation, stability and convergence of algorithms, the ODE method.

\section{Introduction}

Stochastic Approximation (SA) algorithms \cite{29} are a class of iterative procedures used to find the zeros of a function \(h : \mathbb{R}^d \rightarrow \mathbb{R}^d\), when the actual function is not known but noisy observations of the same are available. Standard stochastic approximation iterates are as given below:

\[ x_{n+1} = x_n + \alpha_n f(x_n, Y_n), \quad n \geq 0, \]

where, \(x_n \in \mathbb{R}^d, n \geq 0\), is a sequence of parameters updated as in (1) using \(f(x_n, Y_n)\) that are noisy observations of the true function value \(h(x_n)\). Further, \(Y_n, n \geq 0\), is the corresponding noise sequence and \(\alpha_n, n \geq 0\), the given step-size sequence. In this paper, we assume that the noisy observation of the function is of the following form:

\[ f(x_n, Y_n) = h(x_n) + M_{n+1}, \]

where \(h(\cdot)\) is the objective function and \(M_{n+1}, n \geq 0\), is a martingale difference sequence with respect to the increasing sequence of sigma fields \(\mathcal{F}_n = \sigma(x_i, M_i, 0 \leq i \leq n), n \geq 0\). The form (2) of \(f\) is fairly general. For instance, if \(Y_n \neq 0\), are i.i.d., one may let \(h(x_n) = E_Y [f(x_n, Y)]\) and \(M_{n+1} = f(x_n, Y_n) - h(x_n), n \geq 0\), are the corresponding martingale differences. Then (2) can be seen to hold.

In many applications, the objective function \(h\) might require an additional averaging for its evaluation at a given parameter update. The procedure, in such a case, would require an additional recursion to estimate \(h\) and the overall procedure would result in a recursive computation in nested loops. For example, the policy iteration procedure in Markov decision processes \cite{26}, involves two nested loops in which the outer loop performs policy improvement while the inner loop performs policy evaluation for a given policy update. Actor-critic algorithms \cite{21, 6} are stochastic recursive iterates that try to mimic the nested loop structure of policy iteration under noisy sample estimates where the actor recursion performs policy improvement while the critic is responsible for policy evaluation. This is achieved by making use of two different step-size schedules, one of which goes to zero at a rate faster than the other. The recursion governed by the step-size that goes to zero faster ends up being the ‘slower’ timescale recursion while the one governed by the other step-size schedule corresponds to the ‘faster’ timescale update. Thus, even though both recursions are executed simultaneously, at each instant, one sees a similar effect that a nested loop procedure would provide.
In particular, the faster recursion sees the slower update as ‘quasi-static’ while the latter sees the faster recursion as essentially ‘equilibrated’.

Another example along these lines would be of a Markov reward process \((24)\) whose transition probabilities depend on a certain parameter \(\theta\) and the goal is to find the optimal parameter, i.e., the one that maximizes the long-run average reward. Two-timescale stochastic iterates can be used to solve such problems as well \((5)\) where the recursion corresponding to the inner loop, i.e., the one governed by the faster timescale, performs the averaging of the single-stage rewards for a given update of the parameter \(\theta\) while the outer loop recursion or the one governed by the slower timescale performs a maximization of the average rewards over \(\theta\).

Two-timescale stochastic approximation algorithms have the following general form:

\[
\begin{align*}
    x^{(1)}_{n+1} &= x^{(1)}_n + \alpha^{(1)}_n \left( h^{(1)}\left(x^{(1)}_n, x^{(2)}_n\right) + M^{(1)}_{n+1}\right), \quad (3) \\
    x^{(2)}_{n+1} &= x^{(2)}_n + \alpha^{(2)}_n \left( h^{(2)}\left(x^{(1)}_n, x^{(2)}_n\right) + M^{(2)}_{n+1}\right), \quad (4)
\end{align*}
\]

where \(\{M^{(1)}_{n+1}\}\) and \(\{M^{(2)}_{n+1}\}\) are both martingale difference sequences and \(\{\alpha^{(1)}_n\}\) and \(\{\alpha^{(2)}_n\}\) are the step-size sequences, respectively. A specific requirement of the step-sizes for (3)-(4) to be a two-timescale recursion is that \(\frac{\alpha^{(1)}_n}{\alpha^{(2)}_n} \to 0\) as \(n \to \infty\).

These algorithms have been analyzed for their convergence in \((7)\) assuming, in particular, that the iterates remain stable, i.e., that almost surely \(\sup_n \|x^{(1)}_n\| < \infty\) and \(\sup_n \|x^{(2)}_n\| < \infty\), respectively. This is a nontrivial requirement and in fact until recently, there were no known conditions to verify this requirement. In the case of one-timescale algorithms, i.e., when \(N = 1\), sufficient conditions for stability and convergence of stochastic iterates are available in the literature. For instance, in \((8, 10)\), certain conditions are provided on the limiting ODE and its scaling limit that are shown to be sufficient for the stability of the stochastic recursion.

In \((22)\), the stability conditions provided in \((8)\) are generalized to the two-timescale setting and these happen to be the first set of conditions that guarantee stability of general two-timescale stochastic recursions and along with \((7)\), provide a complete set of conditions under which the recursions given by (3) and (4) converge almost surely as \(n \to \infty\). In this paper, we generalize the results of \((22)\) to the case of \(N\)-timescale stochastic approximation for any general \(N \geq 1\). We explain below the motivation for doing so.

In many problems of optimization or control under uncertainty, one encounters algorithms involving three or more timescale recursions. Such examples arise for instance when designing actor-critic algorithms for constrained Markov decision processes \((1)\), see for instance, \((9, 3)\) or hierarchical reinforcement learning \((4)\). In \((9, 3)\), the constraints are relaxed using a Lagrangian and the Lagrange parameter is updated on the slowest timescale, while the actor (policy improvement) update happens on the middle timescale and the critic (policy evaluation) update happens on the fastest timescale. Further, for settings such as hierarchical reinforcement learning \((4)\), one may encounter algorithms with a number of timescales that is proportionate to the number of levels of decision making. Sufficient conditions for stability and convergence of algorithms (with \(N > 2\)) are not available in the literature. The results we present here are general enough and applicable to such settings. In fact, we also study, in this paper, applications involving certain reinforcement learning algorithms with momentum that require three and four timescale recursions whose convergence we prove by verifying that our sufficient conditions hold.

We provide the first set of sufficient conditions that ensure both the stability and convergence of general \(N\)-timescale stochastic approximation recursions. Our conditions are obtained by generalizing the requirements in \((8)\) for one-timescale algorithms as well as in \((22)\) for two-timescale recursions. We provide the full analysis of stability and convergence of a general \(N\)-timescale stochastic recursion using the aforementioned sufficient conditions. We then demonstrate the usefulness of these conditions in the context of three and four timescale gradient temporal difference learning algorithms in reinforcement learning. Such an analysis is made possible because of the sufficient conditions that we provide for the stability and convergence of \(N\)-timescale algorithms for any \(N \geq 1\). Since \(N \geq 1\) is arbitrary, we believe that the sufficient conditions for stability and convergence of \(N\)-timescale stochastic recursions will be extremely useful for algorithm designers who will henceforth be able to prove the stability and convergence of the stochastic recursions through verification of the aforementioned sufficient conditions.

The rest of the paper is organized as follows: In Section \((2)\) we present the notation used in the paper as well as our assumptions. Here, for improved clarity, we first present our assumptions for the 3-timescale case followed by assumptions for the 4-timescale setting before we present the general assumptions for \(N\)-timescale stochastic iterates. In Section \((3)\) we present our main results on stability and convergence of \(N\)-timescale SA and in Section \((4)\) we provide the complete analysis of stability and convergence of these recursions. In Section \((5)\) we present two gradient temporal difference learning algorithms with momentum that involve three and four timescale updates respectively. We show here
the verification of our assumptions in Section 2.2 for the three and four timescale recursions and also show empirically that the proposed momentum methods outperform their vanilla counterparts. Finally, we present concluding remarks as well as potential directions for future research in Section 6.

2 Notations and Assumptions

In this section, we carefully setup the notation followed in the entire paper and then explicitly provide the assumptions under which the iterates are stable and convergent. The $N$ stochastic recursions are given below:

$$
x_{n+1}^{(1)} = x_n^{(1)} + \alpha_n^{(1)}(h^{(1)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(1)}),
$$

$$
x_{n+1}^{(2)} = x_n^{(2)} + \alpha_n^{(2)}(h^{(2)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(2)}),
$$

$$
\vdots
$$

$$
x_{n+1}^{(N)} = x_n^{(N)} + \alpha_n^{(N)}(h^{(N)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(N)}).
$$

Here, $\{x_n^{(j)} \in \mathbb{R}^{d_j}\}_{1 \leq j \leq N}$ are the $N$ parameters that are updated at each time-step and the subscript $n$ denotes the time-step or iteration-index of the update. $\{\alpha_n^{(j)}, n \geq 0\}, 1 \leq j \leq N$ are $N$ different step-size sequences. The functions $h^{(j)} : \mathbb{R}^{d_1+d_2+\cdots+d_N} \to \mathbb{R}^{d_j}, 1 \leq j \leq N$ could potentially depend on all the $N$ parameters. Also, $\{M_{n+1}^{(j)}\}, 1 \leq j \leq N$ are the associated martingale difference noise sequences.

The following are standard assumptions made while analyzing stochastic recursions:

(A:1) $h^{(j)}, 1 \leq j \leq N$ are Lipschitz continuous functions.

(A:2) $\{M_{n+1}^{(j)}\}, 1 \leq j \leq N$ are $N$ martingale difference sequences with respect to the increasing sequence of sigma fields $\{\mathcal{F}_n\}$, where $\mathcal{F}_n = \sigma\left(x_n^{(j)}, M_{n+1}^{(j)}, 1 \leq j \leq N, i \leq n\right)$, $n \geq 0$, with $\mathbb{E}[\|M_{n+1}^{(j)}\|^2 | \mathcal{F}_n] \leq K^{(i)}(1 + \sum_{j=1}^N \|x_n^{(j)}\|^2), n \geq 0$, for some constant $K^{(i)} > 0, 1 \leq i \leq N$.

(A:3) $\{\alpha_n^{(j)}\}_{1 \leq j \leq N}$ are $N$ step-size schedules that satisfy the following:

(i) $\alpha_n^{(j)} > 0, \forall n, \sum_{n=0}^\infty \alpha_n^{(j)} = \infty, 1 \leq j \leq N$,

(ii) $\sum_{n=0}^\infty \sum_{j=1}^N (\alpha_n^{(j)})^2 \leq \infty$,

(iii) $\frac{n^{(j)}}{\alpha_n^{(j)}} \to 0$, as $n \to \infty, 1 \leq j \leq N$.

Assumptions (A:1)-(A:3) along with stability of the iterates, i.e., $\sup_n(x_n^{(j)}) < \infty, \forall 1 \leq j \leq N$ can be shown to ensure convergence. The next set of assumptions along with (A:1)-(A:3) will provide sufficient conditions for the stability of the iterates. For the case of one-timescale stochastic update recursions, [8] provided verifiable sufficient conditions for stability of the iterate sequence (see (A1) in [8] or (A5) in Chapter 3 of [10]). By considering a scaled version of the original recursion and a scaled ODE associated with the scaled version, [8] shows that under sufficiently general conditions, the original recursion remains stable.

For two-timescale recursions, [22] provided conditions for stability (see (A4) and (A5) there). Since the iterates, in this case, evolve along both the timescales simultaneously, [22] analyzed the rescaled trajectories of both the iterates and presented the aforementioned conditions (cf. (A4) and (A5) therein), one for each timescale, to ensure stability. We extend these ideas to the general $N$-timescale regime, where we analyze the $N$ rescaled iterates and come up with a set of $N$ conditions on the behaviour of the limiting ODEs corresponding to the rescaled recursions. For improved clarity we first state these conditions in the 3-Timescale and 4-Timescale settings in Sections 2.1 and 2.2 respectively. In Section 2.3 we analyze the stability and convergence of two reinforcement learning algorithms with momentum that respectively involve three and four timescale updates by verifying these conditions in Sections 2.1 and 2.2 respectively. We state the conditions for general $N$-timescale stochastic recursions in Section 2.3. The assumptions in what follows will be numbered as (B:ji) and (C:ji) respectively. Here the B-assumptions concern the rescaled ODEs while the C-assumptions are for the ODEs corresponding to the original recursions. Further, the index $j$ refers to the number of timescales used in the algorithm and $i$ takes values in general between $1$ and $j$. 

3
2.1 The 3-Timescale Case

We state first the conditions that ensure stability and convergence for the $N = 3$ case. In Section 5 we discuss an application of our results to the case of stochastic approximation with momentum where the iterates can be analyzed using the results for the 3-Timescale case.

(B.3.1) For $c \geq 1$, define the following scaled functions based on $h^{(1)}$:

$$h_c^{(1)}(x^{(1)}, x^{(2)}, x^{(3)}) = \frac{h^{(1)}(c x^{(1)}, c x^{(2)}, c x^{(3)})}{c}.$$  

Further, $h_c^{(1)} = h^{(1)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$\dot{x}^{(1)}(t) = h^{(1)}_{\infty}(x^{(1)}(t), x^{(2)}(t), x^{(3)}(t)),$$

for $x^{(j)} \in \mathbb{R}^{d_j}$, $j = 2, 3$, has a unique globally asymptotically stable equilibrium $\lambda^{(1)}_{\infty}(x^{(2)}, x^{(3)})$, where $\lambda^{(1)}_{\infty} : \mathbb{R}^{d_2+d_3} \to \mathbb{R}^{d_2}$ is Lipschitz continuous. Further $\lambda^{(1)}_{\infty}(0, 0) = 0$.

(B.3.2) For $c \geq 1$, define the following scaled functions based on $h^{(2)}$:

$$h_c^{(2)}(x^{(2)}, x^{(3)}) = \frac{h^{(2)}(c \lambda^{(1)}_{\infty}(x^{(2)}, x^{(3)}), c x^{(2)}, c x^{(3)})}{c}.$$  

Further, $h_c^{(2)} = h^{(2)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$\dot{x}^{(2)}(t) = h^{(2)}_{\infty}(x^{(2)}(t), x^{(3)}(t)),$$

for $x^{(j)} \in \mathbb{R}^{d_j}$, $j = 3, 4$, has a unique globally asymptotically stable equilibrium $\lambda^{(2)}_{\infty}(x^{(3)})$, where $\lambda^{(2)}_{\infty} : \mathbb{R}^{d_3} \to \mathbb{R}^{d_2}$ is Lipschitz continuous. Further $\lambda^{(2)}_{\infty}(0) = 0$.

(B.3.3) For $c \geq 1$, define the following scaled functions based on $h^{(3)}$:

$$h_c^{(3)}(x^{(3)}) = \frac{h^{(3)}(c \lambda^{(1)}_{\infty}(\lambda^{(2)}_{\infty}(x^{(3)}), x^{(3)}), c \lambda^{(2)}_{\infty}(x^{(3)}), c x^{(3)})}{c}.$$  

Further, $h_c^{(3)} = h^{(3)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$\dot{x}^{(3)}(t) = h^{(3)}_{\infty}(x^{(3)}(t))$$

has the origin in $\mathbb{R}^{d_3}$ as its unique globally asymptotically stable equilibrium.

(C.3.1) The ODE

$$\dot{x}^{(1)}(t) = h^{(1)}(x^{(1)}(t), x^{(2)}(t), x^{(3)}(t)),$$

$x^{(j)} \in \mathbb{R}^{d_j}$, $j = 2, 3$, has a unique globally asymptotically stable equilibrium $\lambda^{(1)}(x^{(2)}, x^{(3)})$ where $\lambda^{(1)} : \mathbb{R}^{d_2+d_3} \to \mathbb{R}^{d_2}$ is Lipschitz continuous.

(C.3.2) The ODE

$$\dot{x}^{(2)}(t) = h^{(2)}(\lambda^{(1)}(x^{(2)}(t), x^{(3)}(t)), x^{(2)}(t), x^{(3)}(t)),$$

with $x^{(3)} \in \mathbb{R}^{d_3}$, has a unique globally asymptotically stable equilibrium $\lambda^{(2)}(x^{(3)})$ and $\lambda^{(2)} : \mathbb{R}^{d_3} \to \mathbb{R}^{d_2}$ is Lipschitz continuous.

(C.3.3) The ODE

$$\dot{x}^{(3)}(t) = h^{(3)}(\lambda^{(1)}(\lambda^{(2)}(x^{(3)}(t)), x^{(3)}(t)), \lambda^{(2)}(x^{(3)}(t)), x^{(3)}(t)),$$

has a unique globally asymptotically stable equilibrium $x^{(3)}_s \in \mathbb{R}^{d_3}$.  

4
2.2 The 4-Timescale case

Next, we state the conditions for the $N = 4$ case for purposes of clarity. The second application we discuss in Section 5 considers a reinforcement learning algorithm that uses a 4-timescale update rule where we show the verification of these conditions and thereby show that the stochastic iterates are stable and almost surely convergent.

**(B.4.1)** For $c \geq 1$, define the following scaled functions based on $h^{(1)}$:

$$
h^{(1)}_{c}(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}) = \frac{h^{(1)}(c \lambda^{(1)}(x^{(1)}), c \lambda^{(2)}(x^{(2)}), c \lambda^{(3)}(x^{(3)}), c \lambda^{(4)}(x^{(4)}))}{c}.
$$

Further, $h^{(1)}_{c} \to h^{(1)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$
\dot{x}^{(1)}(t) = h^{(1)}_{\infty}(x^{(1)}(t), x^{(2)}, x^{(3)}, x^{(4)}),
$$

for $x^{(j)} \in \mathbb{R}^{d_{j}}$, $j = 2, 3, 4$, has a unique globally asymptotically stable equilibrium $\lambda^{(1)}_{\infty}(x^{(2)}, x^{(3)}, x^{(4)})$, where $\lambda^{(1)}_{\infty} : \mathbb{R}^{d_{2}+d_{3}+d_{4}} \to \mathbb{R}^{d_{1}}$ is Lipschitz continuous. Further $\lambda^{(1)}_{\infty}(0, 0, 0) = 0$.

**(B.4.2)** For $c \geq 1$, define the following scaled functions based on $h^{(2)}$:

$$
h^{(2)}_{c}(x^{(2)}, x^{(3)}, x^{(4)}) = \frac{h^{(2)}(c \lambda^{(1)}(x^{(1)}), x^{(2)}, x^{(3)}, x^{(4)}), c \lambda^{(2)}(x^{(2)}), c \lambda^{(3)}(x^{(3)}), c \lambda^{(4)}(x^{(4)}))}{c}.
$$

Further, $h^{(2)}_{c} \to h^{(2)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$
\dot{x}^{(2)}(t) = h^{(2)}_{\infty}(x^{(2)}(t), x^{(3)}, x^{(4)}),
$$

for $x^{(j)} \in \mathbb{R}^{d_{j}}$, $j = 3, 4$, has a unique globally asymptotically stable equilibrium $\lambda^{(2)}_{\infty}(x^{(3)}, x^{(4)})$, where $\lambda^{(2)}_{\infty} : \mathbb{R}^{d_{3}+d_{4}} \to \mathbb{R}^{d_{2}}$ is Lipschitz continuous. Further $\lambda^{(2)}_{\infty}(0, 0) = 0$.

**(B.4.3)** For $c \geq 1$, define the following scaled functions based on $h^{(3)}$:

$$
h^{(3)}_{c}(x^{(3)}, x^{(4)}) = \frac{1}{c} h^{(3)}_{\infty}(c \lambda^{(2)}(x^{(2)}), x^{(3)}, x^{(4)}), c \lambda^{(2)}(x^{(2)}), c \lambda^{(3)}(x^{(3)}), c \lambda^{(4)}(x^{(4)})).
$$

Further, $h^{(3)}_{c} \to h^{(3)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$
\dot{x}^{(3)}(t) = h^{(3)}_{\infty}(x^{(3)}(t), x^{(4)}),
$$

for $x^{(4)} \in \mathbb{R}^{d_{4}}$ has a unique globally asymptotically stable equilibrium $\lambda^{(3)}_{\infty}(x^{(4)})$, where $\lambda^{(3)}_{\infty} : \mathbb{R}^{d_{3}} \to \mathbb{R}^{d_{4}}$ is Lipschitz continuous. Further $\lambda^{(3)}_{\infty}(0) = 0$.

**(B.4.4)** For $c \geq 1$, define the following scaled functions based on $h^{(4)}$:

$$
h^{(4)}_{c}(x^{(4)}) = \frac{1}{c} h^{(4)}_{\infty}(c \lambda^{(2)}(x^{(2)}), x^{(4)}), c \lambda^{(2)}(x^{(2)}), x^{(4)}), c \lambda^{(3)}(x^{(3)}), x^{(4)}), c \lambda^{(4)}(x^{(4)}).
$$

Further, $h^{(4)}_{c} \to h^{(4)}_{\infty}$ as $c \to \infty$ uniformly on compacts. The ODE

$$
\dot{x}^{(4)}(t) = h^{(4)}_{\infty}(x^{(4)}(t)),
$$

for $x^{(4)} \in \mathbb{R}^{d_{4}}$ has the origin in $\mathbb{R}^{d_{4}}$ as its globally asymptotically stable equilibrium.

**(C.4.1)** The ODE

$$
\dot{x}^{(1)}(t) = h^{(1)}(x^{(1)}(t), x^{(2)}, x^{(3)}, x^{(4)}),
$$

$x^{(j)} \in \mathbb{R}^{d_{j}}$, $j = 2, 3, 4$, has a unique globally asymptotically stable equilibrium $\lambda^{(1)}(x^{(2)}, x^{(3)}, x^{(4)})$, where $\lambda^{(1)} : \mathbb{R}^{d_{2}+d_{3}+d_{4}} \to \mathbb{R}^{d_{1}}$ is Lipschitz continuous.

**(C.4.2)** The ODE

$$
\dot{x}^{(2)}(t) = h^{(2)}(\lambda^{(1)}(x^{(2)}(t), x^{(3)}, x^{(4)}), x^{(2)}(t), x^{(3)}, x^{(4)}),
$$

$x^{(j)} \in \mathbb{R}^{d_{j}}$, $j = 3, 4$, has a unique globally asymptotically stable equilibrium $\lambda^{(2)}(x^{(3)}, x^{(4)})$, where $\lambda^{(2)} : \mathbb{R}^{d_{3}+d_{4}} \to \mathbb{R}^{d_{2}}$ is Lipschitz continuous.
Chapter 6 of [10] makes two more assumptions namely (A1) and (A2) there regarding the global asymptotic stable C-assumptions, respectively. In addition, we shall also require Assumptions A:1-A:3 stated previously.

Next we generalize these assumptions to the N-timescale regime. The assumptions on the timescales 1, 2, ..., N are compactly stated as assumptions (B.N.i) and (C.N.i), respectively, where in fact the index i takes N values, i.e., i = 1, ..., N. Thus, these conditions encapsulate a total of 2N assumptions, with N B-assumptions and N C-assumptions, respectively. In addition, we shall also require Assumptions A:1-A:3 stated previously.

(B.N.i) For c ≥ 1, define the following scaled functions based on h(i):

\[ h_c^{(i)}(x^{(i)}, x^{(i+1)}, \ldots, x^{(N)}) = \frac{h^{(i)}(cy^{(i)}, cy^{(2)}, \ldots, cy^{(N)})}{c}, \]

where, for j < i

\[ y^{(j)} = \lambda^{(i-j)}_{\infty} \left( \lambda^{(i-j+1)}_{\infty} \left( \cdots \lambda^{(i-3)}_{\infty} \left( \lambda^{(i-2)}_{\infty} \left( \lambda^{(i-1)}_{\infty}(x^{(i)}, \ldots, x^{(N)}), \right), \right), \right), \right), \]

\[ x^{(i)}, \ldots, x^{(N)} \]

and for j ≥ i

\[ y^{(j)} = cx^{(j)}. \]

Further, \( h_c^{(i)} \to h^{(i)}_{\infty} \) as \( c \to \infty \) uniformly on compacts. The ODE

\[ \dot{x}^{(i)}(t) = h^{(i)}_{\infty}(x^{(i)}(t), x^{(i+1)}, \ldots, x^{(N)}), \]

for \( x^{(j)} \in \mathbb{R}^{d_j}, j \in \{i + 1, \ldots, N\} \) has

(i) a unique globally asymptotically stable equilibrium \( \lambda^{(i)}_{\infty}(x^{(i+1)}, \ldots, x^{(N)}), i = 1, \ldots, N - 1, \)

where each \( \lambda^{(i)}_{\infty} : \mathbb{R}^{d_{i+1} + d_{i+2} + \cdots + d_N} \to \mathbb{R}^{d_i} \) is Lipschitz continuous. Further, \( \lambda^{(i)}_{\infty}(0, 0, \ldots, 0) = 0, \forall i = 1, \ldots, N - 1, \) and

(ii) for \( i = N, \) the origin in \( \mathbb{R}^{d_N} \) is the unique globally asymptotically stable equilibrium of (6).

Chapter 6 of [10] makes two more assumptions namely (A1) and (A2) there regarding the global asymptotic stable equilibria of the ODE trajectories along the two timescales. We make \( N \) such assumptions here (to account for the \( N \) different timescales in our recursions) on the ODE trajectories and compactly write these as (C.N.i), \( 1 \leq i \leq N. \)

(C.N.i) The ODE

\[ \dot{x}^{(i)}(t) = h^{(i)}(z^{(1)}(t), z^{(2)}(t), \ldots, z^{(N)}(t)), \]

has a globally asymptotically stable equilibrium:

(i) \( \lambda(x^{i+1}, \ldots, x^{N}), x^{(j)} \in \mathbb{R}^{d_j}, i + 1 \leq j \leq N \) with \( 1 \leq i \leq N - 1, \) and
(ii) \( x_n^{(N)} \in \mathbb{R}^{dN} \).

Here, for \( j < i \),
\[
\lambda^{(j)}(t) = \lambda^{(j+1)} \left( \lambda^{(j)} \left( \lambda^{(i-3)} \left( \ldots \lambda^{(i-2)} \left( \lambda^{(i-1)} (x^{(i)}(t), x^{(i+1)}, \ldots, x^{(N)}) \right) \ldots \right) \right) \right)
\]
and, for \( j \geq i \),
\[
\lambda^{(j)}(t) = x^{(j)}.
\]

### 3 Main Results

We state in this section (and later prove) our results for general \( N \)-timescale stochastic approximation algorithms as given in Section 2.3. Define for \( 1 \leq j \leq N - 2 \),
\[
\lambda^{(j;N-1)}(x) \overset{\Delta}{=} \lambda^{(j)} \left( \lambda^{(j+1)} \left( \lambda^{(N-2)}(x), x, \ldots, x \right) \right).
\]

**Theorem 3.1.** Under the assumptions (A:1)-(A:3), \( (B.N.i)_{1 \leq i \leq N} \) and \( (C.N.i)_{1 \leq i \leq N} \),
\[
\begin{align*}
x_n^{(1)} &\to x^{(1)} = \lambda^{(1;N-1)}(x^{(N)}), \\
x_n^{(2)} &\to x^{(2)} = \lambda^{(2;N-1)}(x^{(N)}), \\
\vdots \\
x_n^{(N-1)} &\to x^{(N-1)} = \lambda^{(N-1;N-1)}(x^{(N)}), \\
x_n^{(N)} &\to x^{(N)}.
\end{align*}
\]

We first prove that the \( N \) iterates are convergent under the assumption that the \( N \) iterates are stable. Specifically, we first prove a similar result (see Theorem 3.2) under the following assumption instead of \( (B.N.i) \), \( 1 \leq i \leq N \).

**(B.N.N+1)** The following holds:
\[
\sup_n \left( \|x_n^{(1)}\| + \|x_n^{(2)}\| + \cdots + \|x_n^{(N)}\| \right) < \infty \quad \text{a.s.}
\]

**Theorem 3.2.** Under the assumptions (A:1)-(A:3), \( (B.N.N+1) \) and \( (C.N.i)_{1 \leq i \leq N} \),
\[
\begin{align*}
x_n^{(1)} &\to x^{(1)} = \lambda^{(1;N-1)}(x^{(N)}), \\
x_n^{(2)} &\to x^{(2)} = \lambda^{(2;N-1)}(x^{(N)}), \\
\vdots \\
x_n^{(N-1)} &\to x^{(N-1)} = \lambda^{(N-1;N-1)}(x^{(N)}), \\
x_n^{(N)} &\to x^{(N)}.
\end{align*}
\]

Next, we shall show that the \( N \) recursions are a.s. stable under the assumptions \( (B.N.i)_{1 \leq i \leq N} \). In other words, Assumption \( (B.N.N+1) \) holds under \( (B.N.i)_{1 \leq i \leq N} \).

**Theorem 3.3.** Under the assumptions (A:1)-(A:3), \( (B.N.i)_{1 \leq i \leq N} \) and \( (C.N.i)_{1 \leq i \leq N} \),
\[
\sup_n \left( \|x_n^{(1)}\| + \|x_n^{(2)}\| + \cdots + \|x_n^{(N)}\| \right) < \infty.
\]
Theorem 3.3 with Theorem 3.2 together imply that the recursions converge a.s. under the assumptions in Theorem 3.1.

Finally, we consider the $N$-timescale recursions where each iterate contains a small additive perturbation term as follows:

\[ x_{n+1}^{(1)} = x_n^{(1)} + \alpha_n^{(1)}(h^{(1)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(1)}), \]
\[ x_{n+1}^{(2)} = x_n^{(2)} + \alpha_n^{(2)}(h^{(2)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(2)}), \]
\[ \vdots \]
\[ x_{n+1}^{(N)} = x_n^{(N)} + \alpha_n^{(N)}(h^{(N)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(N)}), \]

(8)

**Theorem 3.4.** Assume (A:1)-(A:3), (B.N1)_{1 \leq i \leq N} and (C.N1)_{1 \leq i \leq N} hold. Further assume that $\sum_{i=1}^{N} \varepsilon_i^{(i)} \to 0$ as $n \to \infty$. Then $x_n^{(j)}, 1 \leq j \leq N$ converges almost surely to the same solution as in Theorem 3.1.

Since the additional error terms are $o(1)$, their contribution is asymptotically negligible. See arguments in the third extension of (Chapter 2, pp. 17 of [10]) that handles this case for one-timescale iterates. Using similar arguments along with our analysis, this extension can be easily obtained.

4 Proof of the Main Results

4.1 Showing convergence by assuming Stability (Theorem 3.2)

We start by characterizing the set to which the iterate-vector $(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)})$ converges. Extending the arguments in Lemma 1, Chapter 6 of [10], we first consider the timescale of $\{x_n^{(1)}\}$ and rewrite the $N$ iterates as follows:

\[ x_{n+1}^{(1)} = x_n^{(1)} + \alpha_n^{(1)}(h^{(1)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(1)}), \]
\[ x_{n+1}^{(2)} = x_n^{(2)} + \alpha_n^{(2)}(h^{(2)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(2)}), \]
\[ \vdots \]
\[ x_{n+1}^{(N)} = x_n^{(N)} + \alpha_n^{(N)}(h^{(N)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) + M_{n+1}^{(N)}), \]

where, we define $\forall 2 \leq j \leq N$,

\[ \varepsilon_n^{(j)} = \frac{\alpha_n^{(j)}}{\alpha_n^{(1)}} h^{(j)}(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}), \quad M_{n+1}^{(j)} = \frac{\alpha_n^{(j)}}{\alpha_n^{(1)}} M_{n+1}^{(j)}. \]

From assumption (A:3)(iii), $\frac{\alpha_n^{(j)}}{\alpha_n^{(1)}} \to 0$, as $n \to \infty$, $\forall 2 \leq j \leq N$ and thus $\varepsilon_n^{(j)} = o(1)$ a.s. Using now the third extension from Chapter 2 of [10], $(x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)})$ converges to an internally chain transitive invariant set of the ODE

\[ \dot{x}^{(1)}(t) = h^{(1)}(x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t)), \]
\[ \dot{x}^{(2)}(t) = 0, \]
\[ \dot{x}^{(3)}(t) = 0, \]
\[ \vdots \]
\[ \dot{x}^{(N)}(t) = 0. \]
For initial conditions $x^{(j)} \in \mathbb{R}^d, 1 \leq j \leq N$, the internally chain transitive invariant set of the above ODE system is $\left\{ \left( \chi^{(1)}(x^{(2)}, x^{(3)}, \ldots, x^{(N)}), x^{(2)}, x^{(3)}, \ldots, x^{(N)} \right) \right\}$. Therefore,

$$
\begin{align*}
\left( \frac{d}{dt} x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t) \right) &
\to
\left\{ \left( \chi^{(1)}(x^{(2)}, x^{(3)}, \ldots, x^{(N)}), x^{(2)}, x^{(3)}, \ldots, x^{(N)} \right),
\ldots,
\chi^{(N)}(x^{(N)}, x^{(N)}), x^{(N)}, \ldots, x^{(N)} \right\},
\end{align*}
$$

$$
\begin{align*}
\chi^{(N)}(x^{(N)}, x^{(N)}) &
\to
\left\{ \chi^{(N)}(x^{(N)}, x^{(N)}), x^{(N)}, \ldots, x^{(N)} \right\},
\end{align*}
$$

$$
x^{(j)} \in \mathbb{R}^d, 2 \leq j \leq N \right\}
$$

Next, we consider the timescale $\alpha^{(m)}_n 2 \leq m \leq N - 1,$

$$
x^{(m)}_{n+1} = x^{(m)}_n + \alpha^{(m)}_n \left( h^{(m)}(x^{(1)}_n, x^{(2)}_n, \ldots, x^{(N)}_n) + M^{(m)}_{n+1} \right),
$$

$$
x^{(m+1)}_{n+1} = x^{(m+1)}_n + \alpha^{(m+1)}_n \left( e^{(m+1)}_n + \bar{M}^{(m+1)}_{n+1} \right),
$$

$$
\vdots
$$

$$
x^{(N)}_{n+1} = x^{(N)}_n + \alpha^{(N)}_n \left( e^{(N)}_n + \bar{M}^{(N)}_{n+1} \right),
$$

where we re-define for $m + 1 \leq j \leq N$ (note the abuse of notation)

$$
\begin{align*}
\epsilon^{(j)}_n &
= \frac{\alpha^{(j)}_n}{\alpha^{(m)}_n} h^{(j)}(x^{(1)}_n, x^{(2)}_n, \ldots, x^{(N)}_n),
M^{(j)}_{n+1} = \frac{\alpha^{(j)}_n}{\alpha^{(m)}_n} M^{(j)}_{n+1}.
\end{align*}
$$

From assumption (A:3)(ii), \[\frac{\epsilon^{(j)}_n}{\alpha^{(m)}_n} \to 0 as n \to \infty\] and therefore \[\epsilon^{(j)}_n = o(1) \forall m + 1 \leq j \leq N\]. Using the third extension from Chapter 2 of [10], \(x^{(m+1)}, x^{(m+2)}, \ldots, x^{(N)}_n\) converges to an internally chain transitive invariant set of the ODE

$$
\begin{align*}
\frac{dx^{(m+1)}}{dt} &
= h^{(m+1)}(x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t)),
\frac{dx^{(m+2)}}{dt} = 0,
\vdots
\frac{dx^{(N)}}{dt} = 0.
\end{align*}
$$

For initial conditions $x^{(j)} \in \mathbb{R}^d, m \leq j \leq N$, the internally chain transitive invariant set of the ODE system is $\left\{ \left( \chi^{(m)}(x^{(m+1)}, \ldots, x^{(N)}), x^{(m+1)}, x^{(m+2)}, \ldots, x^{(N)} \right) \right\}$. Therefore,

$$
\begin{align*}
\left( \frac{d}{dt} x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t) \right) &
\to
\left\{ \left( \chi^{(m)}(x^{(m+1)}, x^{(m+2)}, \ldots, x^{(N)}), x^{(m+1)}, x^{(m+2)}, \ldots, x^{(N)} \right),
\ldots,
\chi^{(N)}(x^{(N)}), x^{(N)} \right\},
\end{align*}
$$

$$
\begin{align*}
\chi^{(N)}(x^{(N)}) &
\to
\left\{ \chi^{(N)}(x^{(N)}), x^{(N)} \right\},
\end{align*}
$$

$$
x^{(j)} \in \mathbb{R}^d, m + 1 \leq j \leq N \right\}
$$

We now infer the following from [10]:

$$
\begin{align*}
\left( \frac{d}{dt} x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t) \right) &
\to
\left\{ \left( \chi^{(1)}(\lambda(1)(\ldots, \lambda^{(N-2)}(\lambda^{(N-1)}(x^{(N)}), x^{(N)}), \ldots, x^{(N)})) \right),
\chi^{(1)}\lambda(2)(\ldots, \lambda^{(N-2)}(\lambda^{(N-1)}(x^{(N)}), x^{(N)}), x^{(N)}), \ldots,
\ldots, \chi^{(1)}(\lambda(1)(\ldots, \lambda^{(N-2)}(\lambda^{(N-1)}(x^{(N)}), x^{(N)}), x^{(N)}), x^{(N)}), x^{(N)} \right\} \in \mathbb{R}^d a.s.
\end{align*}
$$
Recall now that
\[
\lambda^{(j:N-1)}(x) \triangleq \lambda^{(j)} \left( \lambda^{(j+1)} \left( \ldots \lambda^{(N-2)}(\lambda^{(N-1)}(x), x), \ldots, x \right), x \right),
1 \leq j \leq N - 2.
\]
Then, \( \forall 1 \leq k \leq N - 1, \)
\[
\|\lambda^{(k:N-1)}(x^{(N)}_n) - x^{(k)}_n\| \to 0 \text{ a.s. as } n \to \infty.
\] (12)
Finally, we consider the slowest timescale, i.e., the one corresponding to \( \{\alpha^{(N)}_n\} \). We define the piece-wise linear continuous interpolation of the iterates \( x^{(N)}_n \) as follows:
\[
\bar{x}^{(N)}(t(n)) = x^{(N)}_n,
\]
\[
\bar{x}^{(N)}(t) = x^{(N)}_n + \left( x^{(N)}_{n+1} - x^{(N)}_n \right) \frac{t - t(n)}{t(n + 1) - t(n)},
\]
where, \( t(n) = \sum_{m=0}^{n-1} \alpha^{(N)}_m, n \geq 1 \) with \( t(0) = 0 \). Also, let \( x^{(N),s}(t), t \geq s, \) denote the unique solution of the below ODE starting at \( s \in \mathbb{R} : \)
\[
\dot{x}^{(N),s}(t) = h^{(N)} \left( \lambda^{(1:N-1)}(x^{(N),s}(t)), \lambda^{(2:N-1)}(x^{(N),s}(t)), \ldots, \lambda^{(N-2:N-1)}(x^{(N),s}(t)), \lambda^{(N-1)}(x^{(N),s}(t)), x^{(N),s}(t) \right),
\]
with \( x^{(N),s}(s) = \bar{x}^{(N)}(s) \). Let
\[
\psi^{(N)}_n = \sum_{i=0}^{n-1} \alpha^{(N)}_1 M^{(N)}_{i+1}, n \geq 1.
\]
It is easy to see that \( \{\psi^{(N)}_n, \mathcal{F}_n\} \) is a zero-mean square integrable martingale. Further,
\[
\sum_{n \geq 0} E[\|\psi^{(N)}_{n+1} - \psi^{(N)}_n\|^2|\mathcal{F}_n] = \sum_{n \geq 0} (\alpha^{(N)}_1)^2 E[\|M^{(N)}_{n+1}\|^2|\mathcal{F}_n] < \infty,
\]
a.s., from (A:2), (A:3) and (B.N.N+1). From the martingale convergence theorem, \( \{\psi^{(N)}_n, \mathcal{F}_n\} \) is an almost surely convergent martingale sequence. Let \( [t] = \max\{t(n) : t(n) \leq t\}, t \geq 0. \) Then for \( n, m \geq 0, \)
\[
\bar{x}^{(N)}(t(n + m)) = \bar{x}^{(N)}(t(n)) + \int_{t(n)}^{t(n+m)} \dot{h}^{(N)} \left( \lambda^{(1:N-1)}(\bar{x}^{(N)}(t)), \lambda^{(2:N-1)}(\bar{x}^{(N)}(t)), \ldots, \lambda^{(N-2:N-1)}(\bar{x}^{(N)}(t)), \lambda^{(N-1)}(\bar{x}^{(N)}(t)), \bar{x}^{(N)}(t) \right)
\]
\[+ I + II + III,
\]
where,
\[
I = \left\| \int_{t(n)}^{t(n+m)} \dot{h}^{(N)} \left( \lambda^{(1:N-1)}(\bar{x}^{(N)}(t)), \lambda^{(2:N-1)}(\bar{x}^{(N)}(t)), \ldots, \lambda^{(N-2:N-1)}(\bar{x}^{(N)}(t)), \lambda^{(N-1)}(\bar{x}^{(N)}(t)), \bar{x}^{(N)}(t) \right) dt \right\|
\]
\[
II = \left\| \sum_{j=0}^{m-1} \alpha^{(N)}_j \left( \dot{h}^{(N)}(x^{(1)}_{n+j}, x^{(2)}_{n+j}, \ldots, x^{(N)}_{n+j}) - h^{(N)}(x^{(1:N-1)}_{n+j}) \right) \right\|
\]
\[
III = \left\| \left( \psi^{(N)}_{n+m+1} - \psi^{(N)}_n \right) \right\|.
\]
Now, as in Lemma 2 of Chapter 2 in [10], using Gronwall inequality we have:

\[
\sup_{t \in [s, s+T]} \| \bar{x}^N(t) - x^{(N), s}(t) \| \leq K_T \left( I + II + III \right),
\]

where, \( K_T > 0 \) is a constant that depends on \( T \). We next show that all the three terms on the RHS above go to 0 as \( n \to \infty \). Using Lipschitz continuity of \( h^{(N)} \), we have

\[
\| I \| \leq \int_{t(n)}^{t(n+m)} L \left( \| \lambda^{(1:N-1)}(\bar{x}^{(N)}([t])) - \lambda^{(1:N-1)}(\bar{x}^{(N)}(t)) \| 
\right.
\]

\[
+ \| \lambda^{(2:N-1)}(\bar{x}^{(N)}([t])) - \lambda^{(2:N-1)}(\bar{x}^{(N)}(t)) \| + \cdots
\]

\[
+ \| \bar{x}^{(N)}([t]) - \bar{x}^{(N)}(t) \| \right)
\]

\[
\leq \sum_{j=0}^{m-1} K_L \int_{t(n+j)}^{t(n+j+1)} \| \bar{x}^{(N)}(t(n+j)) - \bar{x}^{(N)}(t) \| dt,
\]

where \( K_L = L(L^{N-1} + L^{N-2} + \cdots + L + 1) \) and we have used the fact that \( \lambda^{(j:N)} \) is Lipschitz \( \forall 1 \leq j \leq N - 2 \) and \( \| t \| = t(n+j) \) for \( t \in [t(n+j), t(n+j+1)] \), \( \forall 1 \leq j \leq m - 1 \). Now,

\[
\| \bar{x}^{(N)}(t(n+j)) - \bar{x}^{(N)}(t) \| \leq \frac{\| x^{(N)}_{n+j+1} - x^{(N)}_{n+j} \| t}{t(n+j+1) - t(n+j)}.
\]

Therefore,

\[
\| I \| \leq \sum_{j=0}^{m-1} O(\alpha^{(N)}_{n+j}) \int_{t(n+j)}^{t(n+j+1)} \frac{t}{t(n+j+1) - t(n+j)} dt
\]

\[
\leq \sum_{j=n}^{\infty} O(\alpha^{(N)}_{j})^2 \to 0 \text{ as } n \to \infty. \quad (\text{from A:3 (ii)})
\]

Next, using Lipschitz continuity of \( h^{(N)}(\cdot) \),

\[
\| II \| \leq \left\| \sum_{j=0}^{m-1} \alpha^{(N)}_j L \sum_{k=1}^{N-1} \left( \| x^{(k)}_{n+j} - \lambda^{(k:N-1)}(x^{(N)}_{n+j}) \| \right) \right\|
\]

\[
\leq L \sum_{k=1}^{N} O(\sup_{j \geq n} \left\| x^{(k)}_{j} - \lambda^{(k:N-1)}(x^{(N)}_{j}) \| \right)
\]

\[
\to 0 \text{ a.s. as } n \to \infty \quad (\text{from \ref{12}})
\]

Finally, from the foregoing, \( \| III \| \leq \sup_{j > n} \| \psi_j^{(N)} - \psi_n^{(N)} \| \to 0 \) almost surely as \( n \to \infty \). Now, using arguments as in Lemma 1 of Chapter 2 in [10], it can be shown that

\[
\lim_{s \to \infty} \sup_{t \in [s, s+T]} \| \bar{x}^{(N)}(t) - x^{(N), s}(t) \| = 0 \text{ a.s.}
\]

Since, \( x^{(N)}_1 \) is a globally asymptotically stable equilibrium of \( (7) \) with \( i = N \), using Theorem 2 of Chapter 2 in [10], we get \( x^{(N)}_n \to x^{(N)}_* \) a.s. as \( n \to \infty \). Combining this with \( \ref{11} \), proves Theorem 3.2

### 4.2 Showing Stability of the recursions (Theorem 3.3)

We begin with the fastest timescale recursion governed by the step-size sequence \( \{\alpha^{(1)}_n\} \). We first state some of the notations and definitions used.
(D1) Let
\[ t^{(1)}(n) = \sum_{i=0}^{n-1} \alpha_i^{(1)}, n \geq 1, \text{ with } t^{(1)}(0) = 0. \]

Let \( X_k \triangleq (x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(N)}), k \geq 0. \) For \( t \in [t^{(1)}(n), t^{(1)}(n + 1)], \)
\[ \check{X}^{(1)}(t) = X_n + (X_{n+1} - X_n) \frac{t - t^{(1)}(n)}{t^{(1)}(n + 1) - t^{(1)}(n)}. \]

(D2) Given \( t^{(1)}(n), n \geq 0 \) and a constant \( T > 0 \) define
\[ T_0 = 0, \]
\[ T_n = \min(t^{(1)}(m) : t^{(1)}(m) \geq T_{n-1} + T), n \geq 1. \]

One can find a subsequence \( \{m(n)\} \) such that \( T_n = t^{(1)}(m(n)) \forall n \) and \( m(n) \to \infty \) as \( n \to \infty. \)

(D3) Define a sequence \( r(n) \geq 1, \forall n \) as follows:
\[ r(n) = \max(r(n-1), \|\check{X}^{(1)}(T_n)\|, 1). \]

(D4) Define the scaled iterates (obtained from the stochastic recursion above) for \( m(n) \leq k \leq m(n) + 1 - 1 \) as:
\[ \check{x}_k^{(1)} = \frac{x_k^{(1)}}{r(n)}, \check{x}_k^{(2)} = \frac{x_k^{(2)}}{r(n)}, \ldots, \check{x}_k^{(N)} = \frac{x_k^{(N)}}{r(n)}. \]

Further,
\[ \check{x}_{k+1}^{(1)} = \check{x}_k^{(1)} + \alpha_k^{(1)} \left( \frac{h^{(1)}(c\check{x}_k^{(1)}, \ldots, c\check{x}_k^{(N)})}{c} + \hat{M}_k^{(1)} \right), \]
\[ \check{x}_{k+1}^{(2)} = \check{x}_k^{(2)} + \alpha_k^{(1)} \left( \check{t}_k^{(2)} + \hat{M}_k^{(2)} \right), \]
\[ \vdots \]
\[ \check{x}_{k+1}^{(N)} = \check{x}_k^{(N)} + \alpha_k^{(1)} \left( \check{t}_k^{(N)} + \hat{M}_k^{(N)} \right) \]

where, \( c = r(n), \) and
\[ \check{c}_k^{(j)} = \frac{\alpha_k^{(j)} h^{(j)}(c\check{x}_k^{(1)}, \ldots, c\check{x}_k^{(N)})}{\alpha_k^{(1)} c}, 2 \leq j \leq N. \]
\[ \hat{M}_k^{(1)} = \frac{M_k^{(1)}}{r(n)}, \hat{M}_k^{(j)} = \frac{\check{c}_k^{(j)} M_k^{(j)}}{\check{c}_k^{(1)} r(n)}, 2 \leq j \leq N. \]

(D5) Next we define the linearly interpolated trajectory for the scaled iterates \( \check{X}_n^{(1)} = (\check{x}_n^{(1)}, \check{x}_n^{(2)}, \ldots, \check{x}_n^{(N)}) \) as follows:
\[ \check{X}^{(1)}(t) = \hat{X}_n^{(1)} + (\hat{X}_{n+1}^{(1)} - \hat{X}_n^{(1)}) \frac{t - t^{(1)}(n)}{t^{(1)}(n+1) - t^{(1)}(n)}, \]
\[ t \in [t^{(1)}(n), t^{(1)}(n + 1)]. \]

(D6) Let \( X_n^{(1)}(t) \triangleq (x_n^{(1)}(t), x_n^{(2)}(t), \ldots, x_n^{(N)}(t)), t \in [T_n, T_{n+1}] \) denote the trajectory of the ODE:
\[ \dot{x}^{(1)}(t) = h_{r(n)}(x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t)), \]
\[ \dot{x}^{(2)}(t) = 0, \]
\[ \vdots \]
\[ \dot{x}^{(N)}(t) = 0, \]

with \( x_n^{(j)}(T_n) = \check{x}^{(j)}(T_n), \forall 1 \leq j \leq N. \)
We state a lemma for an ODE with external inputs. Let \( x^{(1)}_c(t), x^{(2)}(t), \ldots, x^{(N)}(t) \) denote the trajectories of the following ODEs:

\[
\dot{x}^{(1)}(t) = h^{(1)}_c \left( x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t) \right),
\]

\[
\dot{x}^{(1)}(t) = h^{(1)}_c \left( x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t) \right),
\]

respectively, with initial condition \( x^{(1)}(t) \in \mathbb{R}^{d_1} \) and the external inputs \( x^{(j)}(t) \in \mathbb{R}^{d_j}, \forall 2 \leq j \leq N \). Let \( B^j(x^{(j)}, r) \triangleq \{ q \in \mathbb{R}^{d_j} \mid \| q - x^{(j)} \| < r \} \) denote a ball of radius \( r \) centred at \( x^{(j)} \).

**Lemma 4.1.** Let \( x^{(1)}(t) \in B^1(0, 1) \subset \mathbb{R}^{d_1}, x^{(j)}(t) \in W^{(j)} \subset \mathbb{R}^{d_j}, \forall 2 \leq j \leq N \) and let (B.N.I) hold. Then given \( \epsilon > 0, \exists c_\epsilon \geq 1, r_\epsilon > 0 \) and \( T_\epsilon > 0 \) such that for any external inputs satisfying \( \tilde{x}^{(j)}(s) \in B^j(x^{(j)}, r_\epsilon), \forall s \in [0, T], \forall c > c_\epsilon, t \geq T_\epsilon, \)

\[
\| x^{(1)}(t), x^{(2)}(t), \ldots, x^{(N)}(t) \| \leq 2\epsilon.
\]

The next lemma uses the convergence result of \( N \) scale iterates under the stability assumption of (B.N.N+1) and shows that the scaled iterates defined in (D4) converge.

**Lemma 4.2. Under (A:1)-(A:3),

(i) For \( 0 \leq k \leq m(n + 1) - m(n), \| \tilde{X}^{(1)}(t(m(n + k))) \| \leq K^{(1)} \) a.s. for some constant \( K^{(1)} > 0 \).

(ii) \( \lim_{n \to \infty} \sup_{t \in [T_n, T_{n+1}]} \| \tilde{X}^{(1)}(t) - X^{(1)}(t) \| = 0 \) a.s.

The proof of the above two Lemmas follows in a similar manner as that of Lemmas 5 and 6, respectively, of \[22\]. In particular, Lemma 4.2(i) shows that along the timescale of \( \{ \alpha^{(1)}_n \} \), between instants \( T_n \) and \( T_{n+1} \), the norm of the scaled iterate can grow at most by a factor \( K^{(1)} \) starting from \( B^1(0, 1) \). Next, Lemma 4.2(ii) shows that the scaled iterate asymptotically tracks the ODE defined in (D6). The next lemma bounds \( \| x^{(1)}(n) \| \) in terms of \( \| x^{(2)}(n) \|, \ldots, \| x^{(N)}(n) \| \).

We define the linearly interpolated trajectories of the \( N \) iterates as follows: \( \tilde{x}^{(j)}(t(n)) = x^{(j)} \) and \( \forall t \in [t(n), t(n+1)] \),

\[
\tilde{x}^{(j)}(t) = x^{(j)} + \left( x^{(j)} - x^{(j)}_{n+1} \right) \frac{t - t^{(1)}(n)}{t^{(1)}(n + 1) - t^{(1)}(n)}.
\]

**Lemma 4.3. Under assumptions (A:1)-(A:3), (B.N.i) \( 1 \leq i \leq N \) and (C.N.i) \( 1 \leq i \leq N \),

(i) For \( n \) large, and \( T = T_n \), if \( \| \tilde{x}^{(1)}(T_n) \| > C_1(1 + \| \tilde{x}^{(2)}(T_n) \| + \cdots + \| \tilde{x}^{(N)}(T_n) \|) \), for some \( C_1 > 0 \) then \( \| \tilde{x}^{(1)}(T_{n+1}) \| \leq \frac{3}{2} \| \tilde{x}^{(1)}(T_n) \| \).

(ii) \( \| \tilde{x}^{(1)}(T_n) \| \leq C_1(1 + \| x^{(2)}(T_n) \| + \cdots + \| x^{(N)}(T_n) \|) \) a.s. for some \( C_1^* \geq 1 \).

(iii) \( \| x^{(1)}(n) \| \leq K_1^* \left( \| x^{(2)}(n) \| + \cdots + \| x^{(N)}(n) \| \right) \), for some \( K_1^* > 0 \).

**Proof.** (i) We have \( \| \tilde{x}^{(1)}(T_n) \| > C_1(1 + \| \tilde{x}^{(2)}(T_n) \| + \cdots + \| \tilde{x}^{(N)}(T_n) \|) \). Since, \( r(n) = \max(0, r(n - 1), \| \tilde{X}^{(1)}(T_n) \|, 1) \), \( r(n) \geq \| \tilde{X}^{(1)}(T_n) \| \). Therefore, \( r(n) \geq C_1 \). Next we show that

\[
\| \tilde{x}^{(j)}(T_n) \| < \frac{1}{C_1}, \quad \forall 2 \leq j \leq N.
\]

For \( p \geq 1 \),

\[
\| \tilde{x}^{(j)}(T_n) \|_p = \frac{\| \tilde{x}^{(j)}(T_n) \|_p}{r(n)} \leq \frac{\| \tilde{x}^{(j)}(T_n) \|_p}{\| \tilde{X}^{(1)}(T_n) \|_p} \leq \frac{\| \tilde{x}^{(j)}(T_n) \|_p}{\| x^{(1)}(T_n) \|_p} \frac{1}{\left( \| x^{(1)}(T_n) \|_p^p + \cdots + \| x^{(N)}(T_n) \|_p^p \right)^{\frac{1}{p}}}.
\]

Since, \( \| x^{(1)}(T_n) \|_p \geq C_1(1 + \| x^{(2)}(T_n) \|_p + \cdots + \| x^{(N)}(T_n) \|_p) \),

\[
\| x^{(1)}(T_n) \|_p \geq C_1^p \left( \| x^{(2)}(T_n) \|_p + \cdots + \| x^{(N)}(T_n) \|_p \right)^p \geq C_1^p \left( \| x^{(2)}(T_n) \|_p + \cdots + \| x^{(N)}(T_n) \|_p \right)^p.
\]

(13)
Therefore, 
\[
\|\hat{x}^{(2)}(T_n)\|_* \leq \left( \frac{\|\hat{x}^{(2)}(T_n)\|_p}{C_1^p + 1} \right)^{\frac{1}{p}} \left( \|\hat{x}^{(2)}(T_n)\|_p^p + \cdots + \|\hat{x}^{(N)}(T_n)\|_p^p \right)^{\frac{1}{p}} \\
\leq \frac{1}{(1 + C_1^p)^{\frac{1}{p}}} \\
< \frac{1}{C_1}.
\]

The second inequality follows from the fact that \(\|\hat{x}^{(2)}(T_n)\|_p^p \leq \|\hat{x}^{(2)}(T_n)\|_p^p + \cdots + \|\hat{x}^{(N)}(T_n)\|_p^p\). A similar analysis can be carried out to show that \(\|\hat{x}^{(j)}(T_n)\|_p \leq \frac{1}{C_1}, 1 \leq j \leq N\). Next we show that 
\[
\|\hat{x}^{(1)}(T_n)\|_* > \frac{1}{1 + C_1}.
\]

Here we focus on the case when iterates are blowing up. Therefore let \(r(n) = \hat{X}^{(1)}(T_n)\). Then, 
\[
\|\hat{x}(T_n)\|_* = \frac{\|\hat{x}^{(1)}(T_n)\|_p}{\|\hat{x}^{(1)}(T_n)\|_p} = \left( \frac{\|\hat{x}^{(1)}(T_n)\|_p + \cdots + \|\hat{x}^{(N)}(T_n)\|_p}{\|\hat{x}^{(1)}(T_n)\|_p} \right)^{\frac{1}{p}} \\
> \frac{1}{\left( 1 + \frac{\|\hat{x}^{(2)}(T_n)\|_p + \cdots + \|\hat{x}^{(N)}(T_n)\|_p}{C_1^p(\|\hat{x}^{(2)}(T_n)\|_p + \cdots + \|\hat{x}^{(N)}(T_n)\|_p)} \right)^{\frac{1}{p}}} \\
> \frac{1}{1 + \frac{1}{C_1}}.
\]

Let \(\hat{x}^{(j)}(t - T_n) = \hat{x}^{(j)}(t), \forall 2 \leq j \leq N, \forall t \in [T_n, T_{n+1}]\). From Lemma 4.1, \(\exists r_\frac{1}{4}, c_\frac{1}{4}, T_\frac{1}{4} > 0\) such that 
\[
\|x_r^{(1)}(t), \hat{x}^{(2)}(t), \ldots, \hat{x}^{(N)}(t)\|_* \leq \frac{1}{4}, \forall t \geq T_\frac{1}{4}, \forall c \geq c_\frac{1}{4},
\]
whenever \(\hat{x}^{(j)}(t) \in B^j(0, r_\frac{1}{4})\). Choose \(C_1 > \max(c_\frac{1}{4}, \frac{2}{4})\) and \(T = T_\frac{1}{4}\). Since \(\hat{x}^{(j)}(t) = 0, \forall 2 \leq j \leq N\) for the ODE defined in (D6), \(\hat{x}^{(j)}(t - T_n) = \hat{x}^{(j)}(T_n), 2 \leq j \leq N, \forall t \in [T_n, T_{n+1}]\). From \(\|\hat{x}^{(j)}(T_n)\| < \frac{1}{C_1}, 2 \leq j \leq N\), it follows that \(\hat{x}^{(j)}(s) \in B^j(0, r_\frac{1}{4}), 2 \leq j \leq N, \forall s \in [0, T]\). Using Lemma 4.2 ii), \(\|\hat{x}^{(1)}(T_{n+1}^-) - x^{(1)}(T_{n+1})\| < \frac{1}{4}\) for large enough \(n\). Also observe that \(\|x^{(1)}(T_{n+1})\| = \|x_r^{(1)}, \hat{x}^{(2)}(t), \ldots, \hat{x}^{(N)}(t)\|_* \leq \frac{1}{4}\). Using these, we have \(\|\hat{x}^{(1)}(T_{n+1})\| \leq \|\hat{x}^{(1)}(T_{n+1}) - x^{(1)}(T_{n+1})\| + \|\hat{x}^{(1)}(T_{n+1})\| \leq \frac{1}{2}\), where \(\hat{x}^{(1)}(T_{n+1}) = \hat{x}^{(1)}(T_{n+1})/r(n)\). Finally, since 
\[
\frac{\|\hat{x}^{(1)}(T_{n+1})\|}{\|\hat{x}^{(1)}(T_{n+1})\|} = \frac{\|\hat{x}^{(1)}(T_{n+1})\|}{\|\hat{x}^{(1)}(T_{n+1})\|} = \frac{1}{2},
\]
we have 
\[
\|\hat{x}^{(1)}(T_{n+1})\| < \frac{1}{4} \|\hat{x}^{(1)}(T_{n+1})\|.
\]

Choosing \(C_1 > \max(c_\frac{1}{4}, \frac{2}{4}) > 2\), proves the claim.
The claims in (ii) and (iii) follow in a similar manner as Lemma 6(ii)-(iii) of [22]. We repeat the arguments for the sake of completeness.

(ii) We will prove the claim by contradiction. Suppose there exists a monotonically increasing sequence \( \{n_k\} \) such that \( C_{n_k} \uparrow \infty \) as \( k \to \infty \) and \( \|\tilde{x}^{(1)}(T_n)\| \geq C_{n_k}(\|\tilde{x}^{(2)}(T_n)\| + \cdots + \|\tilde{x}^{(N)}(T_n)\|) \), on a set of positive probability. From Lemma 4.2(i), we know that if \( \|\tilde{x}^{(1)}(T_n)\| > C_1(\|\tilde{x}^{(2)}(T_n)\| + \cdots + \|\tilde{x}^{(N)}(T_n)\|) \), then \( \|\tilde{x}^{(1)}(T_k)\| \) falls at an exponential rate into a ball of radius \( C_1(\|\tilde{x}^{(2)}(T_n)\| + \cdots + \|\tilde{x}^{(N)}(T_n)\|) \) for \( k \geq n \). Therefore, corresponding to the sequence \( \{n_k\} \), there exists another sequence \( \{n_k'\} \) such that \( \|\tilde{x}^{(1)}(T_{n_k'})\| \leq C_1(\|\tilde{x}^{(2)}(T_{n_k'})\| + \cdots + \|\tilde{x}^{(N)}(T_{n_k'})\|) \) for \( n_{k-1} \leq n'_k \leq n_k \), but \( \|\tilde{x}^{(1)}(T_{n_k'})\| > C_{n_k}(\|\tilde{x}^{(2)}(T_{n_k'})\| + \cdots + \|\tilde{x}^{(N)}(T_{n_k'})\|) \). From Lemma 4.2(iii), however, we know that the iterates \( \tilde{x}^{(1)} \) can only grow by a factor of \( K^{(1)} \) between \( m(n_k' - 1) \) and \( m(n_k') \), leading to a contradiction. Therefore, \( \|\tilde{x}^{(1)}(T_n)\| \leq C_1(\|\tilde{x}^{(2)}(T_n)\| + \cdots + \|\tilde{x}^{(N)}(T_n)\|) \) a.s. for some \( C_1 \geq 1 \).

(iii) From Lemma 4.2(ii), we know that \( \forall t \in [T_n, T_{n+1}) \), \( \|\tilde{x}^{(1)}(t)\| \leq K^{(1)}\|\tilde{x}^{(1)}(T_n)\| \). Since, \( \tilde{x}^{(1)}(t) \) is a linear interpolation of the iterates \( x^{(1)} \), therefore, \( \|x^{(1)}_n\| \leq \sup_{t \in [t(n), t(n+1)]} \|\tilde{x}^{(1)}(t)\| \). The claim therefore follows by choosing \( K^*_1 = C_1 K^{(1)} \).

\( \square \)

Next we consider the intermediate timescales of \( \{\alpha^{(l)}_n\}_{n \geq 1}, 2 \leq l \leq N - 1 \), and re-define the terms below. Note the abuse of notation here when defining terms such as \( T_n, r(n), m(n), M_{n+1}^{(l)} \), etc., below.

(E1) Define

\[ t^{(l)}(n) = \sum_{i=0}^{n-1} \alpha^{(l)}_i, n \geq 1 \] with \( t^{(l)}(0) = 0 \).

Recall that \( X_n = (x^{(1)}_n, x^{(2)}_n, \ldots, x^{(N)}_n), n \geq 0 \). For \( t \in [t^{(l)}(n), t^{(l)}(n+1)], l = 2, \ldots, N \), define

\[ \tilde{X}^{(l)}(t) = X_n + (X_{n+1} - X_n) \frac{t - t^{(l)}(n)}{t^{(l)}(n+1) - t^{(l)}(n)}. \]

(E2) Given \( t^{(l)}(n), n \geq 0 \) and a constant \( T > 0 \) define

\[ T_0 = 0, \]

\[ T_n = \min(t^{(l)}(m) : t^{(l)}(m) \geq T_{n-1} + T), n \geq 1 \]

One can find a subsequence \( \{m(n)\} \) such that \( T_n = t^{(l)}(m(n)) \) \( \forall n \), and \( m(n) \to \infty \) as \( n \to \infty \).

(E3) The scaling sequence is defined as:

\[ r(n) = \max(r(n-1), \|\tilde{X}^{(l)}(T_n)\|), n \geq 1 \]

(E4) The scaled iterates for \( m(n) \leq k \leq m(n+1) - 1 \) are defined by:

\[ \hat{x}^{(1)}_k = \hat{x}^{(1)}_k, \hat{x}^{(2)}_k = \hat{x}^{(2)}_k, \ldots, \hat{x}^{(N)}_k = \hat{x}^{(N)}_k. \]

Further,

\[ \hat{x}^{(l)}_{k+1} = \hat{x}^{(l)}_k + \alpha^{(l)}_k \left( \frac{h^{(l)}(c\tilde{x}^{(1)}_k, \ldots, c\tilde{x}^{(N)}_k)}{c} + \hat{M}^{(l)}_{k+1} \right), \]

where

\[ \hat{x}^{(l)}_{k+1} = \hat{x}^{(2)}_k + \alpha^{(2)}_k \left( \frac{h^{(2)}(c\tilde{x}^{(1)}_k, \ldots, c\tilde{x}^{(N)}_k)}{c} + \hat{M}^{(2)}_{k+1} \right), \]

\[ \vdots \]

\[ \hat{x}^{(l)}_{k+1} = \hat{x}^{(l)}_k + \alpha^{(l)}_k \left( \frac{h^{(l)}(c\tilde{x}^{(1)}_k, \ldots, c\tilde{x}^{(N)}_k)}{c} + \hat{M}^{(l)}_{k+1} \right). \]
The proof of the above two Lemmas follows in a similar manner as the proof of Lemmas 5 and 9, respectively, of [22].

**Lemma 4.4.** Assume Lemma 4.5 with \( \epsilon > 0 \), \( s \geq h_{\alpha_k} \), \( r(n) \), and let \( \forall \leq j \leq N - 1 \),

\[
\dot{x}^{(N)}_{k+1} = \dot{x}^{(N)}_k + \alpha_k \left( \epsilon^{(N)}_k + \dot{M}_{k+1}^{(N)} \right)
\]

where, \( c = r(n) \), and \( \forall \leq j \leq N - 1 \),

\[
\dot{c}^{(j)}_k = \frac{n^{(j)}_k h^{(j)}(c^{(j)}_k, \ldots, c^{(j)}_k)}{c},
\]

\[
\dot{M}_{k+1}^{(j)} = \frac{n^{(j)}_k M_{k+1}^{(j)}}{r(n)}, \quad \text{for } 1 \leq j \leq l,
\]

\[
\dot{M}_{k+1}^{(j)} = \frac{n^{(j)}_k M_{k+1}^{(j)}}{r(n)}, \quad \text{for } l + 1 \leq j \leq N - 1.
\]

(E5) Next, we define the linearly interpolated trajectory for the scaled iterates \( \dot{X}_n^{(l)} = (x^{(1)}_n, x^{(2)}_n, \ldots, x^{(N)}_n) \) for \( t \in [t^{(l)}_n, t^{(l)}_n + 1] \) as follows:

\[
\dot{X}_n^{(l)}(t) = \dot{X}_n^{(l)} + (\dot{X}_n^{(l)} - \dot{X}_n^{(l)})(t - t^{(l)}_n).
\]

(E6) Let \( X_n^{(l)}(t) = (x^{(1)}_n(t), \ldots, x^{(N)}_n(t)), t \in [T_n, T_{n+1}] \) denote the trajectory of the ODE:

\[
\dot{x}^{(1)}(t) = h^{(1)}(x^{(1)}(t), \ldots, x^{(N)}(t)),
\]

\[
\dot{x}^{(2)}(t) = h^{(2)}(x^{(2)}(t), \ldots, x^{(N)}(t)),
\]

\[
\vdots
\]

\[
\dot{x}^{(l)}(t) = h^{(l)}(x^{(l)}(t), \ldots, x^{(N)}(t)),
\]

\[
\dot{x}^{(l+1)}(t) = 0
\]

\[
\vdots
\]

\[
\dot{x}^{(N)}(t) = 0
\]

with \( x^{(j)}(T_n) = \dot{x}^{(j)}(T_n), \forall 1 \leq j \leq N \). We refer the reader to Assumption (B.N.i) for the definition of \( h^{(l)}(x^{(1)}(t), \ldots, x^{(N)}(t)), i = 1, \ldots, N \).

**Lemma 4.4.** Let \( x^{(l)} \in B^1(0, 1) \subset \mathbb{R}^d, x^{(j)} \in W^{(j)} \subset \mathbb{R}^d, l + 1 \leq j \leq N \) and let (B.N.i) hold. Then given \( \epsilon > 0, \exists c_\epsilon > 1, r_\epsilon > 0 \) and \( T_\epsilon > 0 \) such that for any external inputs satisfying \( \dot{x}^{(j)}(s) \in B^1(x^{(j)}, r_\epsilon), l + 1 \leq j \leq N - 1, \forall s \in [0, T], \forall c > c_\epsilon, t \geq T_\epsilon \),

\[
\|x^{(l)}_{c_\epsilon}(t), x^{(l+1)}_{c_\epsilon}(t), \ldots, x^{(N)}_{c_\epsilon}(t) - x^{(l)}(t), x^{(l+1)}(t), \ldots, x^{(N)}(t)\| \leq 2\epsilon,
\]

with \( x^{(l)}_{c_\epsilon}(t), x^{(l+1)}_{c_\epsilon}(t), \ldots, x^{(N)}_{c_\epsilon}(t), (t, x^{(l)}) \) defined analogously as in Lemma 4.1.

**Lemma 4.5.** Under (A:1)-(A:3),

(i) For \( 0 \leq k \leq m(n + 1) - m(n), \) and \( 2 \leq l \leq N - 1 \), \( \|\dot{X}_n^{(l)}(t(m(n) + k))\| \leq K^{(l)} \) a.s. for some constant \( K^{(l)} > 0 \).

(ii) We have

\[
\lim_{n \to \infty} \sup_{t \in [T_n, T_{n+1}]} \|\dot{X}_n^{(l)}(t) - X_n^{(l)}(t)\| = 0 \text{ a.s.}
\]

The proof of the above two Lemmas follows in a similar manner as the proof of Lemmas 5 and 9, respectively, of [22].

**Lemma 4.6.** Assume (A:1)-(A:3), (B.N.i)\( 1 \leq i \leq N \) and (C.N.i)\( 1 \leq i \leq N \) hold. Then,
(i) For \( n \) large enough, there exists \( T > 0 \), such that if \( \|x^{(l)}(T_n)\| > C_l(\|x^{(l+1)}(T_n)\| + \cdots + \|x^{(N)}(T_n)\|) \), for some \( C_l > 0 \), then \( \|\dot{x}^{(l)}(T_{n+1})\| < \frac{\epsilon}{4} \|\dot{x}^{(l)}(T_n)\| \).

(ii) \( \|x^{(l)}(T_n)\| \leq C_l^* (\|x^{(l+1)}(T_n)\| + \cdots + \|x^{(N)}(T_n)\|) \), for some \( C_l^* > 1 \).

(iii) \( \|x^{(l)}(T_n)\| \leq K_l^*(\|x^{(l+1)}(T_n)\| + \cdots + \|x^{(N)}(T_n)\|) \), for some \( K_l^* > 0 \).

**Proof.** We use an inductive argument to show that the lemma holds. The base case \( l = 1 \) holds from Lemma 4.3. Assume the claim holds for \( l - 1 \). We now show that it holds for \( l \). As in Lemma 4.3, we can show \( \|\dot{x}^{(j)}(T_n)\|_p < \frac{\epsilon}{4} \), \( l + 1 \leq j \leq N \). Next we show that \( \|\dot{x}^{(l)}(T_n)\| > \frac{1}{((l-1)(C_{l,max})^{l-1} + 1)(1 + \frac{1}{C_l^*})} \), where

\[ C_{l,max} = \max_{1 \leq j \leq l-1} (C_j^*). \]

Here again we are considering the case when the iterates are blowing up. Therefore let \( r(n) = \|\dot{X}^{(l)}(T_n)\|. \) Now, from the inductive step, we know that \( \|\dot{x}^{(j)}(T_n)\| \leq C_j^* (\|\dot{x}^{(j+1)}(T_n)\| + \cdots + \|\dot{x}^{(N)}(T_n)\|), l \leq j \leq N \) and therefore,

\[
r(n) \leq \sum_{j=l}^{l-1} C_j^* (\|\dot{x}^{(j+1)}(T_n)\| + \cdots + \|\dot{x}^{(N)}(T_n)\|)
+ \|\dot{x}^{(l)}(T_n)\| + \cdots + \|\dot{x}^{(N)}(T_n)\|
= \left( \sum_{j=l}^{l-1} \prod_{v=1}^{j} C_v^* + 1 \right) \left( \|\dot{x}^{(l)}(T_n)\| + \cdots + \|\dot{x}^{(N)}(T_n)\| \right)
\leq \left( (l-1)(C_{l,max})^{l-1} + 1 \right) \left( \|\dot{x}^{(l)}(T_n)\| + \cdots + \|\dot{x}^{(N)}(T_n)\| \right).
\]

We thus have, \( \|\dot{x}^{(l)}(T_n)\|_p \)

\[
\geq \frac{\|\dot{x}^{(l)}(T_n)\|}{((l-1)(C_{l,max})^{l-1} + 1)(\|\dot{x}^{(l)}(T_n)\| + \cdots + \|\dot{x}^{(N)}(T_n)\|)}
\geq \frac{1}{((l-1)(C_{l,max})^{l-1} + 1)(1 + \frac{1}{C_l^*})}
= \frac{\epsilon}{1 + \frac{\epsilon}{C_l^*}}.
\]

where \( \epsilon = \frac{1}{((l-1)(C_{l,max})^{l-1} + 1)} \). Let \( \dot{x}^{(j)}(t - T_n) = x^{(j)}_n(t), 2 \leq j \leq N, \forall t \in [T_n, T_{n+1}] \). From Lemma 4.4, \( \exists r_{\epsilon/4}, c_{\epsilon/4}, T_{\epsilon/4} > 0 \) such that

\[ \|x^{(l)}_c, \dot{x}^{(l+1)}(t), \ldots, \dot{x}^{(N)}(t), t, \dot{x}^{(l)}(T_n)\| \leq \epsilon, \]

\[ \forall t \geq T_{\epsilon/4}, \forall c \geq c_{\epsilon/4}, \text{ whenever } \dot{x}^{(j)}(t) \in B^j(0, r_{\epsilon/4}), l + 1 \leq j \leq N. \] Choose \( T = T_{\epsilon/4} \). Since \( \dot{x}^{(j)}(t) = 0, l + 1 \leq j \leq N \) for the ODE defined in (E6) and \( \dot{x}^{(j)}(t - T_n) = x^{(j)}_n(t) = \dot{x}^{(j)}(T_n) \forall t \in [T_n, T_{n+1}] \) and we choose \( C_l > \max \left( c_{\epsilon/4}, \frac{2}{r_{\epsilon/4}} \right) \), from \( \|\dot{x}^{(j)}(T_n)\| < \frac{\epsilon}{C_l^*} \), it follows that \( \dot{x}^{(j)}(t) \in B^j(0, r_{\epsilon/4}), \forall s \in [0, T] \). Using Lemma 4.5(ii), \( \exists \Gamma_1 > 0 \) s.t. \( \|\dot{x}^{(j)}(T_{n+1}) - x^{(j)}(T_{n+1})\| < \frac{\epsilon}{2} \) for large enough \( n \) and \( r(n) > \Gamma_1 \). Choose \( C_l > \max \left( c_{\epsilon/4}, \frac{2}{r_{\epsilon/4}}, C_1 \right) \). Also observe that

\[
\|x^{(l)}_n(T_{n+1})\| = \|x^{(l)}_{r(n)}, \dot{x}^{(l+1)}(t), \ldots, \dot{x}^{(N)}(t), T_{n+1} - T_n, \dot{x}^{(l)}(T_n)\| 
\leq \frac{\epsilon}{4}.
\]

Using these, we have

\[
\|\dot{x}^{(l)}(T_{n+1})\| \leq \|\dot{x}^{(l)}(T_{n+1}) - x^{(l)}(T_{n+1})\| + \|x^{(l)}(T_{n+1})\| 
\leq \frac{\epsilon}{2}.
\]
Finally, since
\[ \frac{\|x^{(l)}(T_{n+1})\|}{\|x^{(l)}(T_n)\|} = \frac{\|\hat{x}^{(l)}(T_{n+1})\|}{\|\hat{x}^{(l)}(T_n)\|}, \]
we have
\[ \|\hat{x}^{(l)}(T_{n+1})\| = \frac{\|\hat{x}^{(l)}(T_{n+1})\|}{\|\hat{x}^{(l)}(T_n)\|} \|x^{(l)}(T_n)\| \]
\[ < \frac{\frac{1}{2}}{(t^{-1}(1+1/C))} \|x(T_n)\| \]
\[ < 1 + \frac{1}{C_1}. \]

Choosing \( C_1 > 2 \) proves the claim.

(ii) As before, we will prove the claim by contradiction. Suppose there exists a monotonically increasing sequence \( \{n_k\} \) such that \( C_{n_k} \uparrow \infty \) as \( k \to \infty \) and \( \|\hat{x}^{(l)}(T_{n_k})\| \geq C_{n_k}(\|\hat{x}^{(l+1)}(T_{n_k})\| + \cdots + \|\hat{x}^{(N)}(T_{n_k})\|) \), on a set of positive probability. From Lemma 4.5 (i), we know that if \( \|\hat{x}^{(l)}(T_{n_k})\| > C_l(\|\hat{x}^{(l+1)}(T_{n_k})\| + \cdots + \|\hat{x}^{(N)}(T_{n_k})\|) \), then \( \|\hat{x}^{(l)}(T_{n_k})\| \) falls at an exponential rate into a ball of radius \( C_l(\|\hat{x}^{(l+1)}(T_{n_k})\| + \cdots + \|\hat{x}^{(N)}(T_{n_k})\|) \) for \( k \geq n \). Therefore, corresponding to the sequence \( \{n_k\} \), there exists another sequence \( \{n_k'\} \) such that \( \|\hat{x}^{(l)}(T_{n_k'})\| \leq C_l(\|\hat{x}^{(l+1)}(T_{n_k'})\| + \cdots + \|\hat{x}^{(N)}(T_{n_k'})\|) \) for \( n_k - 1 \leq n_k' \leq n_k \), but \( \|\hat{x}^{(l)}(T_{n_k'})\| > C_{n_k}(\|\hat{x}^{(l+1)}(T_{n_k'})\| + \cdots + \|\hat{x}^{(N)}(T_{n_k'})\|) \). From Lemma 4.5 (ii), however, we know that the iterates \( \hat{x}^{(l)} \) can only grow by a factor of \( K_l^* \) between \( m(n_k' - 1) \) and \( m(n_k') \), leading to a contradiction. Therefore, \( \|\hat{x}^{(l)}(T_{n_k'})\| \leq C_l(\|\hat{x}^{(l+1)}(T_{n_k'})\| + \cdots + \|\hat{x}^{(N)}(T_{n_k'})\|) \) a.s. for some \( C_l^* \geq 1 \).

(iii) From Lemma 4.2 (ii), we know that \( \forall t \in [T_n, T_{n+1}), \|\hat{x}^{(l)}(t)\| \leq K_l^* \|\hat{x}^{(l)}(T_{n+1})\| \). Since, \( \hat{x}^{(l)}(t) \) is a linear interpolation of the iterates \( x^{(l)}_n \), therefore, \( \|x^{(l)}(t)\| \leq \sup_{t \in [T_n, T_{n+1})} \|\hat{x}^{(l)}(t)\| \). The claim therefore follows by choosing \( C_l^* = C_l^+ K_l^* \).

\[ \square \]

Finally, we consider the slowest timescale recursion corresponding to the step-size \( \alpha_{(n)} \) \( n \geq 0 \), and redefine the terms used. Note (again) the abuse of notation in definitions of terms such as \( T_n, m(n), r(n), M_{n+1}^2 \), etc., below.

(F1) Define \( t^{(N)}(n) = \sum_{i=0}^{n-1} \alpha_{i}^{(N)}, n \geq 1 \) with \( t(0) = 0 \). Let \( X^{(N)}_k = (x^{(1)}_k, x^{(2)}_k, \ldots, x^{(N)}_k), k \geq 0, \) and for \( t \in [t^{(N)}(n), t^{(N)}(n + 1)], \)
\[ \hat{X}^{(N)}(t) = X^{(N)}_n + \left( X^{(N)}_{n+1} - X^{(N)}_n \right) \frac{t - t^{(N)}(n)}{t^{(N)}(n + 1) - t^{(N)}(n)}. \]

(F2) Given \( t^{(N)}(n), n \geq 0 \) and constant \( T > 0 \), define
\[ T_n = \min(t^{(N)}(m) : t^{(N)}(m) \geq T_{n-1} + T), n \geq 1, \]
with \( T_0 = 0 \). One can find a subsequence \( \{m(n)\} \) such that \( T_n = t^{(N)}(m(n)), \forall n, \) and \( m(n) \to \infty \) as \( n \to \infty \).

(F3) The scaling sequence is defined as \( r(n) = \max(r(n-1), \|\hat{X}^{(N)}(T_n)\|), n \geq 1 \).

(F4) The scaled iterates for \( m(n) \leq k \leq m(n + 1) - 1 \) are given by
\[ \hat{x}^{(1)}_{k} = \begin{cases} \frac{x^{(1)}_k}{r(n)}, & x^{(2)}_{k} = \frac{x^{(2)}_k}{r(n)}, \ldots, x^{(N)}_{k} = \frac{x^{(N)}_k}{r(n)}. \end{cases} \]
The corresponding updates are as follows:
\[ \hat{x}^{(1)}_{k+1} = \begin{cases} \hat{x}^{(1)}_{k} + \alpha_{1}^{(1)} \left( \frac{h^{(1)}(c \hat{x}^{(1)}_{k}, \ldots, c \hat{x}^{(N)}_{k})}{c} + \hat{M}_{k+1}^{(1)} \right), & \end{cases} \]
Lemma 4.9. Under assumptions of [22].

Proof.

Next, we define the linearly interpolated trajectory for the scaled iterates \( \hat{X}_n^{(N)} = (x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) \) as follows: For \( t \in [t^{(N)}(n), t^{(N)}(n+1)] \),

\[
\hat{X}_n^{(N)}(t) = \hat{X}_n^{(N)} + \left( \hat{X}_{n+1}^{(N)} - \hat{X}_n^{(N)} \right) \frac{t - t^{(N)}(n)}{t^{(N)}(n+1) - t^{(N)}(n)}.
\]

(F5) Let \( x_n^{(j)}(T_n) = \hat{x}^{(j)}(T_n), 1 \leq j \leq N \).

Lemma 4.7. Next, we define the linearly interpolated trajectory for the scaled iterates \( \hat{X}_n^{(N)} = (x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(N)}) \) as follows: For \( t \in [t^{(N)}(n), t^{(N)}(n+1)] \),

\[
\hat{X}_n^{(N)}(t) = \hat{X}_n^{(N)} + \left( \hat{X}_{n+1}^{(N)} - \hat{X}_n^{(N)} \right) \frac{t - t^{(N)}(n)}{t^{(N)}(n+1) - t^{(N)}(n)}.
\]

(F6) Let \( X_n^{(N)}(t) = (x_n^{(1)}(t), \ldots, x_n^{(N)}(t)), t \in [T_n, T_{n+1}] \), denote the trajectory of the ODE:

\[
\dot{x}^{(1)}(t) = h^{(1)}_{r^{(n)}}(x^{(1)}(t), \ldots, x^{(N)}(t)),
\]

\[
\dot{x}^{(2)}(t) = h^{(2)}_{r^{(n)}}(x^{(2)}(t), \ldots, x^{(N)}(t)),
\]

\[\vdots\]

\[
\dot{x}^{(N)}(t) = h^{(N)}_{r^{(n)}}(x^{(N)}(t)),
\]

with \( x_n^{(j)}(T_n) = \hat{x}^{(j)}(T_n), 1 \leq j \leq N \).

Lemma 4.8. Under \((A:1)-(A:3)\),

(i) For \( 0 \leq k \leq m(n + 1) - m(n) \), \( \| \hat{X}^{(N)}(t(m(n) + k)) \| \leq K^{(N)} \) a.s.

(ii) For sufficiently large \( n \), we have \( \sup_{[T_n, T_{n+1}]} \| \hat{X}^{(N)}(t) - x_n^{(N)}(t) \| \rightarrow 0 \) almost surely as \( c \rightarrow \infty \).

As before, the proof of the above two Lemmas follows in a similar manner as the proof of Lemmas 5 and 9, respectively, of [22].

Lemma 4.9. Under assumptions \((A:1)-(A:3), (B.N.i)_{1 \leq i \leq N}\) and \((C.N.i)_{1 \leq i \leq N}\), we have:

(i) For \( n \) large, \( \exists T \) such that \( \| \hat{x}^{(N)}(T_n) \| > C_N \) for some \( C_N > 0 \) then \( \| \hat{x}^{(N)}(T_{n+1}) \| < \frac{1}{2} \| \hat{x}^{(N)}(T_n) \| \).

(ii) \( \| \hat{x}^{(N)}(T_n) \| \leq C^*_N \) for some \( C^*_N > 0 \).

(iii) \( \sup_n \| x_n^{(N)} \| < \infty \) a.s.

Proof. (i) From Lemmas 4.3 and 4.6 we know that

\[
r(n) \leq \sum_{j=1}^{N-1} C^*_j (\| \hat{x}^{(j+1)}(T_n) \| + \cdots + \| \hat{x}^{(N)}(T_n) \|)
\]

\[
= \left( \sum_{j=1}^{N-1} \prod_{\nu=1}^{N-j} C^*_\nu + 1 \right) (\| \hat{x}^{(N)}(T_n) \|).
\]

Therefore,

\[
\| \hat{x}^{(N)}(T_n) \| = \frac{\| \hat{x}^{(N)}(T_n) \|}{r(n)} > \frac{1}{\sum_{j=1}^{N-1} \left( \prod_{\nu=1}^{N-j} C^*_\nu + 1 \right)} \frac{1}{((N-1/C^*_N) + 1)} \approx \tilde{c},
\]

\[\blacksquare\]
where, $C_{N, \text{max}} = \max_{1 \leq j \leq N-1} (C^*_j)$. Since, $0 \in \mathbb{R}^d_+$ is the unique globally asymptotically stable equilibrium, using Lemma 4.7, \( \exists \varepsilon_{\ell/4}, T_{\ell/4} > 0 \) such that \( \|x^{(N)}(t, x^{(N)})\| < \frac{\varepsilon}{4}, \forall \varepsilon \geq \varepsilon_{\ell/4}, t \geq T_{\ell/4} \). Also, for sufficiently large $n$, from Lemma 4.8 (ii), \( \exists \varepsilon > 0 \) such that \( \|x^{(N)}(T_{n+1}) - x^{(N)}(T_{n+1})\| < \frac{\varepsilon}{4} \) for \( r(n) > \Gamma_2 \). We pick $C_N = \max(c_{\ell/4}, \Gamma_2)$ and $T = T_{\ell/4}$. For $n$ large, it follows that
\[
\|x^{(N)}(T_{n+1})\| \leq \|x^{(N)}(T_{n+1}) - x^{(N)}(T_{n+1})\| + \|x^{(N)}(T_{n+1})\| \leq \frac{\varepsilon}{4}.
\]
Finally, since
\[
\frac{\|x^{(N)}(T_{n+1})\|}{\|x^{(N)}(T_{n})\|} = \frac{\|x^{(N)}(T_{n+1})\|}{\|x^{(N)}(T_{n})\|},
\]
and
\[
\|x^{(N)}(T_{n+1})\| \leq \frac{\varepsilon}{4},
\]
the claims in (ii) and (iii) now follow in a similar manner as Theorem 10 (iii)-(iv), respectively, of [22], and the arguments are repeated here for completeness.

(ii) We will prove the claim by contradiction. Suppose there exists a monotonically increasing sequence \{\( n_k \)\} such that \( C_{n_k} \uparrow \infty \) as \( k \to \infty \) and \( \|x^{(N)}(T_{n_k})\| \geq C_{n_k} \), on a set of positive probability. From Lemma 4.8 (i), we know that if \( \|x^{(N)}(T_{n_k})\| > C_N \), then \( \|x^{(N)}(T_{n_k})\| \) falls at an exponential rate into a ball of radius \( C_N \) for \( k \geq n \). Therefore, corresponding to the sequence \{\( n_k \)\}, there exists another sequence \{\( n'_k \)\} such that \( \|x^{(N)}(T_{n'_k})\| \leq C_N \) for \( n_k - 1 \leq n'_k \leq n_k \), but \( \|x^{(N)}(T_{n'_k})\| > C_{n_k} \). From Lemma 4.8 (i), however, we know that the iterates \( x^{(N)} \) can only grow by a factor of \( K(N) \) between \( m(n'_k - 1) \) and \( m(n'_k) \), leading to a contradiction. Therefore, \( \|x^{(N)}(T_{n_k})\| \leq C_N \) a.s. for some \( C_N \geq 1 \).

(iii) From the previous part we have \( \|x^{(N)}\| \leq C_N^* \), and from Lemma 4.8 (i), we have \( \|x^{(N)}(t)\| \leq K(N)\|x^{(N)}(T_{n_k})\| \), \( \forall t \in [T_{n_k}, T_{n_k+1}) \). Therefore, \( \|x^{(N)}\| \leq K(N)C_N^* \) almost surely.

\[\]

The fact that \( \sup_n \|x^{(N)}\| < \infty \) a.s. \( \forall t \in \{1, \ldots, N\} \) follows by combining Lemma 4.9 (iii) and Lemma 4.6 (iii).

## 5 Application: Gradient TD with Momentum

In this section we use our results on \( N \)-timescale recursions to show stability and convergence of Gradient Temporal Difference (Gradient TD), a class of reinforcement learning (RL) algorithms, with momentum. We briefly discuss these settings here. For a detailed description of the Gradient TD algorithms, see [23, 30, 31] and for Gradient TD methods with momentum, see [16].

In the standard RL framework, an agent interacts with a stochastic and dynamic environment. At each discrete time step \( t \), the agent is in state \( s_t \in S \), picks an action \( a_t \in \mathcal{A} \), receives a reward \( r_{t+1} = r(s_t, a_t, s_{t+1}) \in R \) and probabilistically moves to another state \( s_{t+1} \in S \). The tuple \((S, \mathcal{A}, \mathbb{P}, R, \gamma)\) constitutes a Markov Decision Process (MDP). Here \( S \) and \( \mathcal{A} \) are assumed finite. Also, \( \gamma \in (0, 1) \) is the discount factor. A policy \( \pi : S \times \mathcal{A} \to [0, 1] \) is a mapping that defines the probability of picking an action in a state. We let \( P^\pi(s'|s) \) denote the probability of transition to state \( s' \) from state \( s \) when an action is chosen as per \( \pi \). We let \( \{d^\pi(s)\}_{s \in S} \) denote the steady-state distribution for the Markov chain induced by \( \pi \). The matrix \( D \) is a \( n \times n \) diagonal matrix with elements \( d^\pi(s) \) on its diagonals with \( n \) being the number of states. The value function corresponding to state \( s \) under policy \( \pi \) for state \( s \) is defined by:
\[
V^\pi(s) = \mathbb{E}_\pi \left[ \sum_{t=0}^{\infty} \gamma^t R_{t+1} | s_0 = s \right].
\]

With linear function approximation for policy evaluation (i.e., for a fixed \( \pi \)), the goal is to estimate \( V^\pi(s) \) from samples of the form \((s_t, r_{t+1}, s_{t+1})\) through a linear model \( V_\theta(s) = \theta^T \phi(s) \). Here \( \phi(s) \equiv \phi_s \) is a feature vector associated with the state \( s \) and \( \theta \) is the associated parameter vector. The TD-error is defined by \( \delta_t = r_{t+1} + \gamma \theta^T \phi_{s_{t+1}} - \theta^T \phi_s \). The feature matrix \( \Phi \) is an \( n \times d \) matrix where the \( s^{th} \) row is \( \phi(s)^T \). In the following, we consider the i.i.d setting, where the tuple \((\phi_s, \phi'_s)\) (with \( \phi'_s \equiv \phi_{s+1} \)) is drawn independently from the stationary distribution \( \{d_s\} \). Let \( \Delta \triangleq \mathbb{E}[\phi_s (\gamma \phi'_s - \phi_s)^T] \) and \( \delta \triangleq \mathbb{E}[r_{t+1} \phi_s] \), where the expectations are w.r.t. the stationary distribution of the induced Markov chain. The matrix \( \hat{A} \) is known to be negative definite (see [23, 32]). In the off-policy setting, the behaviour policy \( \pi \) is used to sample trajectories from the MDP while the target policy \( \mu \) is the one whose associated value function needs to be approximated. Let \( \rho_t = \frac{\pi(a_t | s_t)}{\mu(a_t | s_t)} \) denote the importance sampling ratio. Gradient TD algorithms
are a class of TD algorithms that are convergent even in the off-policy setting. We first present the iterates associated with the algorithms GTD2 and TDC, see [31].

1. **GTD2:**

\[
\begin{align*}
\theta_{t+1} &= \theta_t + \alpha_t (\phi_t - \gamma \phi_t') \phi_t^T u_t, \\
u_{t+1} &= u_t + \beta_t (\delta_t - \phi_t^T u_t) \phi_t.
\end{align*}
\]

2. **TDC:**

\[
\begin{align*}
\theta_{t+1} &= \theta_t + \alpha_t \delta_t \phi_t - \alpha_t \gamma \phi_t' (\phi_t^T u_t), \\
u_{t+1} &= u_t + \beta_t (\delta_t - \phi_t^T u_t) \phi_t.
\end{align*}
\]

In the GTD algorithm, the objective function considered is the Norm of Expected Error defined as \(\text{NEU}(\theta) = \mathbb{E}[\delta \phi]\) and the algorithm is derived by expressing the gradient direction as \(-\frac{1}{2} \nabla \text{NEU}(\theta) = \mathbb{E}[(\phi - \gamma \phi') \phi^T] \mathbb{E}[\delta(\theta)]\phi\). Here \(\phi' \equiv \phi(s')\). Since the expectation becomes biased by the correlation of the two terms if both the terms are sampled separately, an estimate of the second expectation is maintained as a long-term quasi-stationary estimate while samples for the first expectation are used. For GTD2 and TDC, a similar approach is used on the objective function Mean Square Projected Bellman Error defined as \(\text{MSPBE}(\theta) = \|V_0 - \Pi^\gamma V_0\|_D\), where for any \(x \in \mathbb{R}^n, \|x\|_D = \sqrt{x^T D x}\). Here, \(\Pi\) is the projection operator that projects vectors to the subspace \(\{\phi(\theta) \in \mathbb{R}^d\}\) and \(T^\gamma\) is the Bellman operator defined as \(T^\gamma V = R^\gamma + \gamma P^\gamma V\). It was shown in all the three cases that \(\theta_n \rightarrow \theta^* = -A^{-1}b\).

### 5.1 Three Timescale Gradient TD Algorithms with Momentum

We consider the Gradient TD algorithms with an added heavy ball term to the first iterate.

1. **GTD2 with momentum (GTD2-M-3TS):**

\[
\begin{align*}
\theta_{t+1} &= \theta_t + \alpha_t (\phi_t - \gamma \phi_t') \phi_t^T u_t + \eta_t (\theta_t - \theta_{t-1}), \\
u_{t+1} &= u_t + \beta_t (\delta_t - \phi_t^T u_t) \phi_t.
\end{align*}
\]

2. **TDC with momentum (TDC-M-3TS):**

\[
\begin{align*}
\theta_{t+1} &= \theta_t + \alpha_t (\delta_t \phi_t - \gamma \phi_t' (\phi_t^T u_t)) + \eta_t (\theta_t - \theta_{t-1}), \\
u_{t+1} &= u_t + \beta_t (\delta_t - \phi_t^T u_t) \phi_t.
\end{align*}
\]

The momentum parameter \(\eta_t\) is chosen as in [2] as follows:

\[
\eta_t = \frac{\bar{\eta}_t - w \alpha_t}{\bar{\eta}_t - 1},
\]

where \(\{\bar{\eta}_t\}\) is a positive sequence and \(w \in \mathbb{R}\) is a constant. We let \(\frac{\bar{\eta}_t - \bar{\eta}_t}{\bar{\eta}_t} = \frac{\bar{\eta}_t}{\bar{\eta}_t} - \bar{\eta}_t\), and \(\varepsilon_t = \bar{\eta}_t - \bar{\eta}_t - \bar{\eta}_t\). Then the iterates for **GTD2-M-3TS** can then be decomposed into the three recursions as below:

\[
\begin{align*}
v_{t+1} &= v_t + \xi_t \left((\phi_t - \gamma \phi_t') \phi_t^T u_t - w v_t\right), \\
u_{t+1} &= u_t + \beta_t (\delta_t \phi_t - \phi_t^T u_t), \\
\theta_{t+1} &= \theta_t + \bar{\eta}_t (v_t + \varepsilon_t).
\end{align*}
\]

Similarly, the iterates for **TDC-M-3TS** can be decomposed as:

\[
\begin{align*}
v_{t+1} &= v_t + \xi_t \left(\delta_t \phi_t - \gamma \phi_t' \phi_t^T u_t - w v_t\right), \\
u_{t+1} &= u_t + \beta_t (\delta_t \phi_t - \phi_t^T u_t), \\
\theta_{t+1} &= \theta_t + \bar{\eta}_t (v_t + \varepsilon_t).
\end{align*}
\]

Consider the following assumptions:

**Assumption 1.** All rewards \(r(s, s')\) and features \(\phi(s)\) are bounded, i.e., \(r(s, s') \leq 1\) and \(\|\phi(s)\| \leq 1\) \(\forall s, s' \in \mathcal{S}\). Also, the matrix \(\Phi\) has full rank, where \(\Phi\) is an \(n \times d\) matrix where the \(s^{th}\) row is \(\phi(s)^T\).
We choose the step-size sequences such that they satisfy
\[ \sum_t \xi_t = \sum_t \beta_t = \sum_t \theta_t = \infty, \quad \sum_t (\xi_t^2 + \beta_t^2 + \theta_t^2) < \infty, \]
\[ \frac{\beta_t}{\xi_t} \to 0, \quad \frac{\theta_t}{\beta_t} \to 0 \text{ as } t \to \infty, \]
and the momentum parameter satisfies: \( \eta_t = \frac{\theta_t - \omega \alpha_t}{\theta_{t-1}}. \)

**Assumption 3.** The samples \((\phi_t, \phi'_t)\) are drawn i.i.d from the stationary distribution of the Markov chain induced by the target policy \(\pi.\)

**Remark 1.** Assumptions 1, 2, and 3 are standard requirements. Assumption 3, on the other hand, is a restrictive requirement though often used in the literature, see for instance, \([23, 37, 38] [23, 72]\), where this assumption has been made. This can however be relaxed if our analysis of N-timescale SA algorithms is extended to the case when the noise can have a general iterate-dependent Markovian structure that can appear in each of the \(N\) recursions. This setting is popularly referred to as the Markov noise setting. There is no prior work on stability of multi-scale algorithms with Markov noise even though \([27]\) does present stability conditions for one-timescale algorithms with Markov noise.

**Theorem 5.1.** Suppose Assumptions 1, 2, and 3 hold and let \(w > 0.\) Then, the GTD2-M-3TS iterates given by \((22)-(24)\) satisfy \(\theta_t \to \theta^* = -\tilde{A}^{-1} \tilde{b}\) a.s. as \(t \to \infty.\)

**Proof.** We transform the iterates given by \((22)-(24)\) into the standard SA form. Let \(F_t = \sigma(u_0, v_0, \theta_0, r_{j+1}, \phi_j, \phi'_j : j < t).\) Let, \(A_t = \phi_t(\gamma \phi'_t - \phi_t)^T\) and \(b_t = r_{t+1} \phi_t.\) Then, \((22)\) can be re-written as:
\[ v_{t+1} = v_t + \xi_t \left( h^{(1)}(v_t, u_t, \theta_t) + M^{(1)}_{t+1} \right) \]
\[ u_{t+1} = u_t + \beta_t \left( h^{(2)}(v_t, u_t, \theta_t) + M^{(2)}_{t+1} \right) \]
\[ \frac{\beta_t}{\xi_t} \to 0, \quad \frac{\theta_t}{\beta_t} \to 0 \text{ as } t \to \infty, \]
and the momentum parameter satisfies: \( \eta_t = \frac{\theta_t - \omega \alpha_t}{\theta_{t-1}}. \)

The functions \(h^{(1)}, h^{(2)}, h^{(3)}\) are linear in \(v, u, \theta\) and hence Lipschitz continuous, thereby satisfying \((A:1).\) One popular choice is
\[ \xi_t = \frac{1}{(t + 1)\gamma}, \quad \beta_t = \frac{1}{(t + 1)\sigma}, \quad \theta_t = \frac{1}{(t + 1)\rho}, \quad \text{and any } 0 < \gamma < \beta < \rho \leq 1. \]

Since \(M^{(1)}_{t+1}, M^{(2)}_{t+1} \) and \(M^{(3)}_{t+1} \) are martingale difference sequences w.r.t \(F_t\) by construction. Further,
\[ \mathbb{E}[||M^{(1)}_{t+1}||^2 | F_t] \leq ||(\tilde{A}^T - A_t^T)||^2 ||u_t||^2, \]
\[ \mathbb{E}[||M^{(2)}_{t+1}||^2 | F_t] \leq 3( ||(A_t - \tilde{A})||^2 ||\theta_t||^2 + ||(b_t - \tilde{b})||^2 + ||(\tilde{C} - C_t)||^2 ). \]

Note that \((A:3)\) is satisfied with \(K_1 = ||(\tilde{A}^T - A_t^T)||^2, K_2 = 3 \max( ||A_t - \tilde{A}||^2, ||b_t - \tilde{b}||^2, ||(\tilde{C} - C_t)||^2 ),\) and any \(K_3 > 0.\) The fact that \(K_1, K_2 < \infty\) follows from Assumption 1. For a fixed \(u, \theta \in \mathbb{R}^d,\) consider the ODE
\[ \dot{v}(t) = -\tilde{A}^T u - w(t). \]
For $w > 0$, $\lambda^{(1)}(u, \theta) = -\frac{\dot{A}^T w}{w}$ is the unique globally asymptotically stable equilibrium (g.a.s.e), is linear and therefore Lipschitz continuous. This satisfies (C.3.1). Next, for a fixed $\theta \in \mathbb{R}^d$,
\[
\dot{u}(t) = \dot{A} \theta + \dot{b} - \dot{C} u(t),
\]
has $\lambda^{(2)}(\theta) = \dot{C}^{-1}(\dot{A} \theta + \dot{b})$ as its unique g.a.s.e because $-\dot{C}^{-1}$ is negative definite. Also $\lambda^{(2)}(\theta)$ is linear in $\theta$ and therefore Lipschitz. This satisfies (C.3.2). Finally, to satisfy (C.3.3), consider,
\[
\dot{\theta}(t) = \frac{-\dot{A}^T \dot{C}^{-1} \dot{A} \theta(t) - \dot{A}^T \dot{C}^{-1} \dot{b}}{w}.
\]
Since $\dot{A}$ is negative definite and $\dot{C}$ is positive definite, $-\dot{A}^T \dot{C}^{-1} \dot{A}$ is negative definite as well. Therefore, $\theta^* = -\dot{A}^{-1} \dot{b}$ is the unique g.a.s.e for the above ODE.

Next, we show that the sufficient conditions for stability of the three iterates are satisfied. The function,
\[
h_c^{(1)}(v, u, \theta) = \frac{-c \dot{A}^T u - w v}{w} = -\dot{A}^T u - w v \rightarrow h_c^{(1)}(v, u, \theta) = -\dot{A}^T u - w v \text{ uniformly on compacts as } c \rightarrow \infty.
\]
This limiting ODE:
\[
\dot{v}(t) = -\dot{A}^T u - w v(t)
\]
has $\lambda^{(1)}(u, \theta) = -\frac{\dot{A}^T u}{w}$ as its unique g.a.s.e. $\lambda^{(1)}$ is Lipschitz with $\lambda^{(1)}(0, 0) = 0$, thus satisfying assumption (B.3.1). The function,
\[
h_c^{(2)}(u, \theta) = \frac{c \dot{A} \dot{\theta} + \dot{b} - c \dot{C} u}{c} = \dot{A} \theta - C u + \frac{\dot{b}}{c} \rightarrow h_c^{(2)}(u, \theta) = \dot{A} \theta - C u \text{ uniformly on compacts as } c \rightarrow \infty.
\]
This limiting ODE:
\[
\dot{u}(t) = \dot{A} \theta - C u(t)
\]
has $\lambda^{(2)}(\theta) = \dot{C}^{-1} \dot{A} \theta$ as its unique g.a.s.e. since $-\dot{C}$ is negative definite. $\lambda^{(2)}$ is Lipschitz with $\lambda^{(2)}(0) = 0$. Thus assumption (B.3.2) is satisfied.

Finally, $h_c^{(3)}(\theta) = \frac{-c \dot{A}^T \dot{C}^{-1} \dot{A} \theta}{w} \rightarrow h_c^{(3)}(\theta) = -\frac{-\dot{A}^T \dot{C}^{-1} \dot{A} \theta(t)}{w}$ uniformly on compacts as $c \rightarrow \infty$ and the ODE:
\[
\dot{\theta}(t) = -\frac{-\dot{A}^T \dot{C}^{-1} \dot{A} \theta(t)}{w}
\]
has the origin in $\mathbb{R}^d$ as its unique g.a.s.e. This ensures the final condition (B.3.3). Further, observe that $\|\varepsilon_t^{(3)}\| = \xi_t \|((\phi_t - \gamma \phi_t')\phi_t^T u_t - w u_t)\| \rightarrow 0$ since $\xi_t \rightarrow 0$ as $t \rightarrow \infty$. By Theorem 3.4,
\[
\begin{pmatrix}
v_t \\
u_t \\
\theta_t
\end{pmatrix} \rightarrow \begin{pmatrix}
\lambda(\Gamma(-\dot{A}^{-1} \dot{b}), -\dot{A}^{-1} \dot{b}) \\
\Gamma(-\dot{A}^{-1} \dot{b}) \\
-\dot{A}^{-1} \dot{b}
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
\dot{A}^{-1} \dot{b}
\end{pmatrix}
\]
Specifically, $\theta_t \rightarrow -\dot{A}^{-1} \dot{b}$.

\begin{proof}
As before, we transform the iterates given by (25), (26) and (27) into the standard SA form. Let $F_t = \sigma(u_0, v_0, \theta_0, r_{j+1}, \phi_{j}, \phi_{j}^T : j < t)$. Let $A_t = \phi_t \phi_t^T - \phi_t \phi_t^T$ and $b_t = r_{t+1} \phi_t$. Then, (25) can be re-written as:
\[
v_{t+1} = v_t + \xi_t \left( h^{(1)}(v_t, u_t, \theta_t) + M_{t+1}^{(1)} \right),
\]
where $h^{(1)}(v_t, u_t, \theta_t) = \mathbb{E} \{ \delta_t \phi_t - \gamma \phi_t \phi_t^T u_t - w u_t | F_t \} = \dot{A} \theta_t + \dot{b} - \gamma \mathbb{E} \{ \phi_t \phi_t^T u_t - w u_t \}$ and $M_{t+1}^{(1)} = \delta_t \phi_t - \gamma \phi_t \phi_t^T u_t - w u_t - h(v_t, u_t, \theta_t) = (A_t - \bar{A}) \theta_t + (b_t - \bar{b}) + \gamma (\mathbb{E} \phi_t \phi_t^T u_t - \phi_t \phi_t^T u_t)$. Next, (25) can be re-written as:
\[
u_{t+1} = v_t + \beta_t \left( h^{(2)}(v_t, u_t, \theta_t) + M_{t+1}^{(2)} \right),
\]
where $h^{(2)}(v_t, u_t, \theta_t) = \mathbb{E} \{ \delta_t \phi_t - \gamma \phi_t \phi_t^T u_t | F_t \} = \dot{A} \theta_t + \dot{b} - \dot{C} u_t$ and $M_{t+1}^{(2)} = A_t \theta_t + b_t - C_t u_t - h^{(2)}(v_t, u_t, \theta_t) = (A_t - \bar{A}) \theta_t + (b_t - \bar{b}) + (\bar{C} - C_t) u_t$. Here, $C_t = \phi_t \phi_t^T$ and $\bar{C} = \mathbb{E} \phi_t \phi_t^T$. Finally, (27) can be re-written as:
\[
\theta_{t+1} = \theta_t + \tilde{\theta}_t \left( h^{(3)}(v_t, u_t, \theta_t) + \varepsilon_t + M_{t+1}^{(3)} \right),
\]
where $h^{(3)}(v_t, u_t, \theta_t) = v_t$ and $M_{t+1}^{(3)} = 0$. As before we show that the conditions (A:1)-(A:3), (B:3.1)-(B:3.3) and (C:3.1)-(C:3.3) hold. The functions $h^{(1)}, h^{(2)}, h^{(3)}$ are linear in $v, u, \theta$ and hence Lipschitz continuous, therefore satisfying (A:1). We choose the step-size sequences such that they satisfy (A:2). One popular choice is (25).
\end{proof}
Observe now that $M_{t+1}^{(1)}$, $M_{t+1}^{(2)}$ and $M_{t+1}^{(3)}$ $t \geq 0$, are martingale difference sequences w.r.t $\mathcal{F}_t$ by construction. Next,
\[
\begin{align*}
E[||M_{t+1}^{(1)}||^2 |\mathcal{F}_t] &\leq 3(||(A_t - \bar{A})||^2||\theta_t||^2 + ||(b_t - \bar{b})||^2 \\
+ \gamma(\|E[\phi_t' \phi_t^T] - \phi_t' \phi_t^T\|_2)^2 ||u_t||^2), \\
E[||M_{t+1}^{(2)}||^2 |\mathcal{F}_t] &\leq 3(||(A_t - \bar{A})||^2||\theta_t||^2 + ||(b_t - \bar{b})||^2 \\
+ ||(C - C_t)||^2 ||u_t||^2). 
\end{align*}
\]
The first part of (A:3) is satisfied with $K_1 = 3 \max(||A_t - \bar{A}||^2, ||b_t - \bar{b}||^2, \gamma(||E[\phi_t' \phi_t^T] - \phi_t' \phi_t^T||^2)), K_2 = 3 \max(||A_t - \bar{A}||^2, ||b_t - \bar{b}||^2, ||(C - C_t)||^2)$ and any $K_3 > 0$. The fact that $K_1, K_2 < \infty$ follows from the bounded features and bounded rewards assumption in Assumption [1] For a fixed $u, \theta \in \mathbb{R}^d$, consider the ODE
\[
\dot{v}(t) = \bar{A}_\theta + \bar{b} - \gamma E[\phi_t' \phi_t^T] u - w(t).
\]
For $w > 0$, $\lambda^{(1)}(u, \theta) = \frac{\bar{A}_\theta + \bar{b} - \gamma E[\phi_t' \phi_t^T] u}{w}$ is the unique g.a.s.e, is linear and therefore Lipschitz continuous. This satisfies (C:3.1). Next, for a fixed $\theta \in \mathbb{R}^d$,
\[
\dot{u}(t) = \bar{A}_\theta + \bar{b} - \bar{C} u(t),
\]
has $\lambda^{(2)}(\theta) = \bar{C}^{-1}(\bar{A}_\theta + \bar{b})$ as its unique g.a.s.e because $-\bar{C}^{-1}$ is negative definite. Also $\lambda^{(2)}(\theta)$ is linear in $\theta$ and therefore Lipschitz. This satisfies (C:3.2). Finally, to satisfy (C:3.3), consider,
\[
\dot{\theta}(t) = \frac{(I - \gamma E[\phi_t' \phi_t^T] \bar{C}^{-1})(\bar{A}_\theta(t) + \bar{b})}{w}.
\]
Now, $(I - \gamma E[\phi_t' \phi_t^T] \bar{C}^{-1}) \bar{A} = (E[\phi_t' \phi_t^T] - \gamma E[\phi_t' \phi_t^T]) \bar{C}^{-1} \bar{A} = E[(\phi_t - \gamma \phi_t') \phi_t^T] \bar{C}^{-1} \bar{A} = -\bar{A}^T \bar{C}^{-1} \bar{A}$. Since, $\bar{A}$ is negative definite and $\bar{C}$ is positive definite, therefore $-\bar{A}^T \bar{C}^{-1} \bar{A}$ is negative definite and hence the above ODE has solution $\theta^* = -\bar{A}^{-1} \bar{b}$ as its unique g.a.s.e.

Next, we show that the sufficient conditions for stability of the three iterates are satisfied. The function, $h_{\infty}^{(1)}(v, u, \theta) = \frac{c \bar{A}_\theta + \bar{b} - \gamma E[\phi_t' \phi_t^T] u}{w} = \bar{A}_\theta + \bar{b} - \gamma E[\phi_t' \phi_t^T] u - wv \rightarrow h_{\infty}^{(1)}(v, u, \theta) = \bar{A}_\theta - \gamma E[\phi_t' \phi_t^T] u - wv$ uniformly on compacts as $c \rightarrow \infty$. The limiting ODE:
\[
\dot{v}(t) = \bar{A}_\theta - \gamma E[\phi_t' \phi_t^T] u - wv(t)
\]
has $\lambda_{\infty}^{(1)}(u, \theta) = \frac{\bar{A}_\theta - \gamma E[\phi_t' \phi_t^T] u}{w}$ as its unique g.a.s.e. $\lambda_{\infty}^{(1)}$ is Lipschitz with $\lambda_{\infty}^{(1)}(0, 0) = 0$, thus satisfying assumption (B:3.1).

The function $h_{\infty}^{(2)}(u, \theta) = \frac{c \bar{A}_\theta + \bar{b} - c \bar{C} u}{w} = \bar{A}_\theta - \bar{C} u + \frac{\bar{b}}{c} \rightarrow h_{\infty}^{(2)}(u, \theta) = \bar{A}_\theta - \bar{C} u$ uniformly on compacts as $c \rightarrow \infty$. The limiting ODE
\[
\dot{u}(t) = \bar{A}_\theta - \bar{C} u(t)
\]
has $\lambda_{\infty}^{(2)}(\theta) = \bar{C}^{-1} \bar{A}_\theta$ as its unique g.a.s.e. since $-\bar{C}$ is negative definite. $\lambda_{\infty}^{(2)}$ is Lipschitz with $\lambda_{\infty}^{(2)}(0) = 0$. Thus assumption (B:3.2) is satisfied.

Finally,
\[
\begin{align*}
h_{\infty}^{(3)}(\theta) &= \frac{\bar{A}_\theta - c \gamma E[\phi_t' \phi_t^T] \bar{C}^{-1} \bar{A}_\theta}{cw} \\
\rightarrow h_{\infty}^{(3)} &= \frac{(I - \gamma E[\phi_t' \phi_t^T] \bar{C}^{-1}) \bar{A}_\theta}{w}
\end{align*}
\]
uniformly on compacts as $c \rightarrow \infty$. Consider the ODE:
\[
\dot{\theta}(t) = \frac{(I - \gamma E[\phi_t' \phi_t^T] \bar{C}^{-1}) \bar{A}_\theta(t)}{w}.
\]
Now $(I - \gamma E[\phi_t' \phi_t^T] \bar{C}^{-1}) \bar{A} = (E[\phi_t' \phi_t^T] - \gamma E[\phi_t' \phi_t^T]) \bar{C}^{-1} \bar{A} = E[(\phi_t - \gamma \phi_t') \phi_t^T] \bar{C}^{-1} \bar{A} = -\bar{A}^T \bar{C}^{-1} \bar{A}$. Since $\bar{A}$ is negative definite and $\bar{C}$ is positive definite, $-\bar{A}^T \bar{C}^{-1} \bar{A}$ is negative definite and hence the above ODE has the origin as its unique g.a.s.e. This ensures the final condition (B:3.3). Next, observe that $\|\epsilon_{\infty}^{(3)}\| = \|\epsilon_t\| ((\phi_t - \gamma \phi_t') \phi_t^T u_t - wv_t) \| \rightarrow 0$ since $\xi_t \rightarrow 0$ as $t \rightarrow \infty$. By Theorem 3.4
\[
\begin{pmatrix}
\xi_t \\
u_t \\
\theta_t
\end{pmatrix} \rightarrow \begin{pmatrix}
\lambda(\Gamma(-\bar{A}^{-1} \bar{b}), -\bar{A}^{-1} \bar{b}) \\
\Gamma(-\bar{A}^{-1} \bar{b}) \\
-\bar{A}^{-1} \bar{b}
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
-\bar{A}^{-1} \bar{b}
\end{pmatrix}
\]
Specifically, $\theta_t \rightarrow -\bar{A}^{-1} \bar{b}$ almost surely. \qed
5.2 Four Timescale Gradient TD Algorithms with Momentum

Next, we consider the Gradient TD algorithms with an added heavy ball term to both the iterates.

1. \textbf{GTD2-M-4TS:}

\[
\theta_{t+1} = \theta_t + \alpha_t(\phi_t - \gamma \phi_t^T u_t + \eta_t^{(1)}(\theta_t - \theta_{t-1})),
\]

\[
u_{t+1} = u_t + \beta_t(\delta_t - \phi_t^T u_t)\phi_t + \eta_t^{(2)}(u_t - u_{t-1}).
\]

2. \textbf{TDC-M-4TS:}

\[
\theta_{t+1} = \theta_t + \alpha_t(\delta_t\phi_t - \gamma \phi_t^T u_t) + \eta_t^{(1)}(\theta_t - \theta_{t-1}),
\]

\[
u_{t+1} = u_t + \beta_t(\delta_t - \phi_t^T u_t)\phi_t + \eta_t^{(2)}(u_t - u_{t-1}).
\]

The momentum parameters are chosen as in [2] as follows:

\[
\eta_{t}^{(1)} = \frac{\eta_{t}^{(1)} - \mu \alpha_t}{\eta_{t-1}^{(1)}}, \quad \eta_{t}^{(2)} = \frac{\eta_{t}^{(2)} - \mu \beta_t}{\eta_{t-1}^{(2)}},
\]

where \(\{\eta_{t}^{(1)}\}\) and \(\{\eta_{t}^{(2)}\}\) are positive sequences and \(\mu \in \mathbb{R}\) is a constant. We let

\[
\frac{\theta_{t+1} - \theta_t}{\theta_{t+1}} = \nu_{t+1} = \frac{u_{t+1} - u_t}{\theta_{t+1}} = z_{t+1}, \eta_t^{(1)} = \frac{\alpha_t}{\alpha_t}, \eta_t^{(2)} = \frac{\beta_t}{\beta_t}, \phi_t = z_{t+1} - z_t \text{ and } \epsilon_t^{(4)} = v_{t+1} - v_t.
\]

The iterates for \textbf{GTD2-M-4TS} can then be decomposed into the four recursions as below:

\[
v_{t+1} = v_t + \xi_t^{(1)}((\phi_t - \gamma \phi_t^T u_t) - w v_t)
\]

\[
z_{t+1} = z_t + \xi_t^{(2)}((\delta_t - \phi_t^T u_t) - w z_t)
\]

\[
u_{t+1} = u_t + \delta_t^{(2)}(z_t + \epsilon_t^{(3)})
\]

\[
\theta_{t+1} = \theta_t + \delta_t^{(1)}(v_t + \epsilon_t^{(4)}).
\]

The iterates for \textbf{TDC-M-4TS} can be decomposed as:

\[
v_{t+1} = v_t + \xi_t^{(1)}((\phi_t - \gamma \phi_t^T u_t) - w v_t)
\]

\[
z_{t+1} = z_t + \xi_t^{(2)}((\delta_t - \phi_t^T u_t) - w z_t)
\]

\[
u_{t+1} = u_t + \delta_t^{(2)}(z_t + \epsilon_t^{(3)})
\]

\[
\theta_{t+1} = \theta_t + \delta_t^{(1)}(v_t + \epsilon_t^{(4)}).
\]

We make use of the following assumption here instead of Assumption 2.

\textbf{Assumption 4.} The step-sizes satisfy \(\xi_t^{(1)} > 0, \xi_t^{(2)} > 0, \xi_t^{(1)} > 0, \xi_t^{(2)} > 0 \forall t\),

\[
\sum_t \xi_t^{(1)} = \sum_t \xi_t^{(2)} = \sum_t \xi_t^{(1)} = \sum_t \xi_t^{(2)} = \infty,
\]

\[
\sum_t \left(\left(\xi_t^{(1)}\right)^2 + \left(\xi_t^{(2)}\right)^2 + \left(\xi_t^{(1)}\right)^2 + \left(\xi_t^{(2)}\right)^2\right) < \infty,
\]

\[
\frac{\xi_t^{(2)}}{\xi_t^{(1)}} \to 0, \frac{\xi_t^{(2)}}{\xi_t^{(1)}} \to 0, \frac{\xi_t^{(1)}}{\theta_t} \to 0, \text{ as } t \to \infty.
\]

\textbf{Theorem 5.3.} Suppose Assumptions 2 and 3 hold and let \(w > 0\). Then, the GTD-M-4TS iterates given by (29) and (30) satisfy \(\theta_t \to \theta^* = -A^{-1}b\) a.s. as \(t \to \infty\).

\textbf{Proof.} As in the three timescale case, we begin by transforming the iterates to a standard 4 timescale SA form. Let

\[
F_t = \sigma(u_t, v_t, \theta_t, r_{j+1}, \phi_t, \phi_t^T : j < t), \quad A_t = \phi_t(\gamma \phi_t^T - \phi_t)^T \text{ and } b_t = r_{t+1} \phi_t.
\]

Then, (34) can be re-written as:

\[
v_{t+1} = v_t + \xi_t^{(1)}\left(h_t^{(1)}(v_t, z_t, u_t, \theta_t) + M_t^{(1)}\right)
\]
where, $h^{(1)}(v_t, z_t, u_t, \theta_t) = \mathbb{E}[(\phi_t - \gamma_t \phi_t) \phi_t^T u_t - vw_t | \mathcal{F}_t] = -\bar{A}^T u_t - vw_t$ and $M^{(1)}_{t+1} = -\bar{A}^T u_t - vw_t - h(v_t, u_t, \theta_t) = (\bar{A}^T - \bar{A}^T) u_t$. Next, (35) can be re-written as:

$$z_{t+1} = z_t + \varepsilon_t^{(2)} \left( h^{(2)}(v_t, z_t, u_t, \theta_t) + M^{(2)}_{t+1} \right)$$

where, $h^{(2)}(v_t, z_t, u_t, \theta_t) = \mathbb{E}[\delta_t \phi_t - \phi_t \phi_t^T u_t - w z_t | \mathcal{F}_t] = \bar{A} \theta_t + \bar{b} - \bar{C} u_t - w z_t$ and $M^{(2)}_{t+1} = (A_t - \bar{A}) \theta_t + (b_t - \bar{b}) + (C - \bar{C}) u_t$. Here, $C_t = \phi_t \phi_t^T$ and $\bar{C} = \mathbb{E}[\phi_t \phi_t^T]$. Equation (35) can be re-written as:

$$u_{t+1} = u_t + \varepsilon_t^{(2)} \left( h^{(3)}(v_t, z_t, u_t, \theta_t) + M^{(3)}_{t+1} \right)$$

where, $h^{(3)}(v_t, z_t, u_t, \theta_t) = z_t, M^{(3)}_{t+1} = 0$. Finally, (36) can be re-written as:

$$\theta_{t+1} = \theta_t + \varepsilon_t^{(1)} \left( h^{(4)}(v_t, z_t, u_t, \theta_t) + M^{(4)}_{t+1} \right)$$

where, $h^{(4)}(v_t, z_t, u_t, \theta_t) = v_t, M^{(4)}_{t+1} = 0$. The functions $h^{(1)}, h^{(2)}, h^{(3)}$ and $h^{(4)}$ are linear in $v, u, \theta$ and hence Lipschitz continuous, thereby satisfying (A:1). To satisfy (A:2), we choose the step-size sequence as $\xi_t^{(1)} = \frac{1}{(t+1)c^{(1)}}$, $\xi_t^{(2)} = \frac{1}{(t+1)c^{(2)}}$, $\xi_t^{(3)} = \frac{1}{(t+1)c^{(3)}}$, with $\frac{1}{\xi^{(1)}} < \xi^{(2)} < \xi^{(3)} < \frac{1}{\xi^{(1)}} \leq 1$. (A:3) can be shown to hold as in the proof of Theorem 5.1 For a fixed $z, u, \theta \in \mathbb{R}^d$, consider the ODE

$$\dot{v}(t) = -\bar{A}^T u - vw(t).$$

For $w > 0$, $\lambda^{(1)}(z, u, \theta) = -\frac{\bar{A}^T u}{w}$ is the unique g.a.s.e for the above ODE. Further, it is linear and therefore Lipschitz continuous. This satisfies (C.4.1). For a fixed $u, \theta \in \mathbb{R}^d$,

$$\dot{z}(t) = A \theta + \bar{b} - \bar{C} u - w z(t),$$

has $\lambda^{(2)}(u, \theta) = \frac{(A \theta + \bar{b} - \bar{C} u)}{w}$ as its unique g.a.s.e as $w > 0$. Also $\lambda^{(2)}(u, \theta)$ is linear in $u, \theta$ and therefore Lipschitz. This satisfies (C.4.2). Next, for a fixed $\theta \in \mathbb{R}^d$,

$$\dot{u}(t) = \frac{\bar{A} \theta + \bar{b} - \bar{C} u(t)}{w},$$

has $\lambda^{(3)}(\theta) = \bar{C}^{-1}(\bar{A} \theta + \bar{b})$ as its unique g.a.s.e because $-\bar{C}^{-1}$ is negative definite. $\lambda^{(3)}(\theta)$ is linear in $\theta$ and therefore Lipschitz thus satisfying (C.4.3). Finally, to satisfy (C.4.4), consider,

$$\dot{\theta}(t) = -\bar{A}^T \bar{C}^{-1} \bar{A} \theta(t) - \bar{A}^T \bar{C}^{-1} \bar{b},$$

Since $\bar{A}$ is negative definite and $\bar{C}$ is positive definite, $-\bar{A}^T \bar{C}^{-1} \bar{A}$ is negative definite. Therefore, $\theta^{*} = -\bar{A}^{-1} \bar{b}$ is the unique g.a.s.e of the above ODE.

Now we show that the conditions (B.4.1)-(B.4.4) are also satisfied and therefore the iterates are stable. The function, $h^{(1)}_c(v, z, u, \theta) = -\frac{\bar{A}^T u - \bar{w} v}{c} = -\bar{A}^T u - \bar{w} v \rightarrow h^{(1)}_c(v, z, u, \theta) = -\bar{A}^T u - \bar{w} v$ uniformly on compacts as $c \rightarrow \infty$. The limiting ODE:

$$\dot{v}(t) = -\bar{A}^T u - \bar{w} v(t)$$

has $\lambda^{(1)}(z, u, \theta) = -\frac{\bar{A}^T u}{\bar{w}}$ as its unique g.a.s.e. $\lambda^{(1)}_c$ is Lipschitz with $\lambda^{(1)}_c(0, 0, 0) = 0$, thus satisfying assumption (B.4.1).

The function, $h^{(2)}_c(v, z, u, \theta) = \frac{\bar{A} \theta - \bar{C} u - \bar{w} v}{c} \rightarrow h^{(2)}(v, z, u, \theta) = \bar{A} \theta - \bar{C} u - w z$ uniformly on compacts as $c \rightarrow \infty$. The limiting ODE

$$\dot{z}(t) = \bar{A} \theta - \bar{C} u - w z(t)$$

has $\lambda^{(2)}_c(u, \theta) = \frac{\bar{A} \theta - \bar{C} u}{\bar{w}}$ as its unique g.a.s.e and $\lambda^{(2)}_c$ is Lipschitz continuous with $\lambda^{(2)}_c(0, 0) = 0$. Thus assumption (B.4.2) is satisfied. The function, $h^{(3)}_c(u, \theta) = \frac{\bar{A} \theta - \bar{C} u}{\bar{w}} \rightarrow h^{(3)}(u, \theta) = \frac{\bar{A} \theta - \bar{C} u}{\bar{w}}$ uniformly on compacts as $c \rightarrow \infty$. The limiting ODE

$$\dot{u}(t) = \frac{\bar{A} \theta - \bar{C} u(t)}{\bar{w}}$$

26
has $\lambda_\infty^{(2)}(\theta) = \bar{C}^{-1} \bar{A} \theta$ as its unique g.a.s.e. since $-\bar{C}$ is negative definite. $\lambda_\infty^{(2)}$ is Lipschitz with $\lambda_\infty^{(2)}(0) = 0$ and thus assumption (B.4.3) is satisfied as well.

Finally, $h_c^{(4)}(\theta) = \frac{c w}{c} \rightarrow h_c^{(4)}(\theta) = -\frac{\bar{A}^T \bar{C}^{-1} \bar{A} \theta}{w}$ uniformly on compacts as $c \rightarrow \infty$ and the ODE:

$$\dot{\theta}(t) = -\frac{\bar{A}^T \bar{C}^{-1} \bar{A} \theta(t)}{w}$$

has the origin in $\mathbb{R}^d$ as its unique g.a.s.e. This ensures the final condition (B.4.4). Further, as before, $\|\varepsilon_t^{(3)}\|, \|\varepsilon_t^{(4)}\| \rightarrow 0$ as $t \rightarrow \infty$. By Theorem 3.4

$$\begin{pmatrix} v_t \\ z_t \\ u_t \\ \theta_t \end{pmatrix} \rightarrow \begin{pmatrix} \lambda^{(1:4)}(-\bar{A}^{-1} \bar{b}) \\ \lambda^{(2:4)}(-\bar{A}^{-1} \bar{b}) \\ \lambda^{(3:4)}(-\bar{A}^{-1} \bar{b}) \\ -\bar{A}^{-1} \bar{b} \end{pmatrix}$$

Specifically, $\theta_t \rightarrow -\bar{A}^{-1} \bar{b}$.

**Theorem 5.4.** Suppose Assumptions 7, 8, and 9 hold and let $w > 0$. Then, the TDC-M-4TS iterates given by (31) and (32) satisfy $\theta_t \rightarrow \theta^* = -\bar{A}^{-1} \bar{b}$ a.s. as $t \rightarrow \infty$.

**Proof.** As before, we transform the iterates to a standard 4 timescale SA form. Let $F_t = \sigma(u_0, v_0, \theta_0, r_{j+1}, \phi_j, \phi'_j : j < t), A_t = \phi_t(\gamma \phi'_t - \phi_t)^T$ and $b_t = r_{t+1} \phi_t.$ Then, (38) can be re-written as:

$$v_{t+1} = v_t + \varepsilon_t^{(1)} \left( h_1^{(1)}(v_t, z_t, u_t, \theta_t) + M_{t+1}^{(1)} \right)$$

where, $h_1^{(1)}(v_t, z_t, u_t, \theta_t) = \mathbb{E}[\delta_t \phi_t - \gamma \phi'_t \phi^T_t u_t - w v_t | F_t] = \bar{A} \theta_t + \bar{b} - \gamma \mathbb{E}[\phi'_t \phi^T_t] u_t - w v_t$ and $M_{t+1}^{(1)} = \delta_t \phi_t - \gamma \phi'_t \phi^T_t u_t - w v_t - h(v_t, u_t, \theta_t) = (A_t - \bar{A}) \theta_t + (b_t - \bar{b}) + \gamma \mathbb{E}[\phi'_t \phi^T_t] - \phi'_t \phi^T_t) u_t$. Next,(39) can be re-written as:

$$z_{t+1} = z_t + \varepsilon_t^{(2)} \left( h_1^{(2)}(v_t, z_t, u_t, \theta_t) + M_{t+1}^{(2)} \right)$$

where,

$$h_1^{(2)}(v_t, z_t, u_t, \theta_t) = \mathbb{E}[\delta_t \phi_t - \phi_t \phi^T_t u_t - w z_t | F_t]$$

$$= \bar{A} \theta_t + \bar{b} - \bar{C} u_t - w z_t$$

$$M_{t+1}^{(2)} = (A_t - \bar{A}) \theta_t + (b_t - \bar{b}) + (\bar{C} - C_t) u_t.$$

Here, $C_t = \phi_t \phi^T_t$ and $C_t = \mathbb{E}[\phi'_t \phi^T_t]$. Equation (40) can be re-written as:

$$u_{t+1} = u_t + g_t^{(2)} \left( h_3^{(2)}(v_t, z_t, u_t, \theta_t) + M_{t+1}^{(3)} \right)$$

where, $h_3^{(2)}(v_t, z_t, u_t, \theta_t) = z_t, M_{t+1}^{(3)} = 0$. Finally, (41) can be re-written as:

$$\theta_{t+1} = \theta_t + \varepsilon_t^{(4)} \left( h_4^{(4)}(v_t, u_t, \theta_t) + E_t^{(4)} + M_{t+1}^{(4)} \right)$$

where, $h_4^{(4)}(v_t, u_t, \theta_t) = v_t, M_{t+1}^{(4)} = 0$. The functions $h_1^{(1)}, h_2^{(2)}, h_3^{(3)}$ and $h_4^{(4)}$ are linear in $v, u, \theta$ and hence Lipschitz continuous, thereby satisfying (A.1). To satisfy (A.2), we choose the step-size sequence as $\varepsilon_t^{(1)} = \frac{1}{(t+1)^{\xi_t^{(1)}}}, \varepsilon_t^{(2)} = \frac{1}{(t+1)^{\xi_t^{(2)}}}, \varepsilon_t^{(1)} = \frac{1}{(t+1)^{\xi_t^{(1)}}}, \varepsilon_t^{(1)} = \frac{1}{(t+1)^{\xi_t^{(1)}}}$, with $\frac{1}{2} < \xi^{(1)} < \xi^{(2)} < \phi^{(2)} < \phi^{(1)} \leq 1$. (A.3) can be shown to hold as in the proof of Theorem 5.2

For a fixed $z, u, \theta \in \mathbb{R}^d$, consider the ODE

$$\dot{v}(t) = \bar{A} \theta + \bar{b} - \gamma \mathbb{E}[\phi'_t \phi^T_t] u - w v(t),$$

For $w > 0, \lambda^{(1)}(z, u, \theta) = \frac{\bar{A} \theta + \bar{b} - \gamma \mathbb{E}[\phi'_t \phi^T_t] u - w v(t)}{w}$ is the unique g.a.s.e for the above ODE. Further, it is linear and therefore Lipschitz continuous. This satisfies (C.4.1). For a fixed $u, \theta \in \mathbb{R}^d$, \dot{z}(t) = \bar{A} \theta + \bar{b} - \bar{C} u - w z(t),
has \( \lambda^{(2)}(u, \theta) = \frac{(\tilde{A} \theta + \bar{b} - C u) \epsilon}{w} \) as its unique g.a.s.e as \( w > 0 \). Also \( \lambda^{(2)}(u, \theta) \) is linear in \( u, \theta \) and therefore Lipschitz. This satisfies (C.4.2). Next, for a fixed \( \theta \in \mathbb{R}^d \),
\[
\dot{u}(t) = \frac{\tilde{A} \theta + \bar{b} - C u(t)}{w},
\]
has \( \lambda^{(3)}(\theta) = \bar{C}^{-1}(\tilde{A} \theta + \bar{b}) \) as its unique g.a.s.e because \( -\bar{C}^{-1} \) is negative definite. \( \lambda^{(3)}(\theta) \) is linear in \( \theta \) and therefore Lipschitz thus satisfying (C.4.3). Finally, to satisfy (C.4.4), consider,
\[
\dot{\theta}(t) = \frac{(I - \gamma E[\phi'_t \phi'_T] \bar{C}^{-1})(\tilde{A} \theta + \bar{b})}{w},
\]
As in the proof of Theorem 5.2 \( \theta^* = \tilde{A}^{-1} \bar{b} \) is the unique g.a.s.e of the above ODE.

Now we show that the conditions (B.4.1)-(B.4.4) are also satisfied and therefore the iterates are stable. The function, \( h_c^{(1)}(v, z, u, \theta) = \frac{c\tilde{A} \theta + \bar{b} - c \bar{C} u - c w z}{c \epsilon} = \tilde{A} \theta + \bar{b} - \gamma E[\phi'_t \phi'_T] u - w v \rightarrow h_c^{(1)}(v, z, u, \theta) = \tilde{A} \theta - \gamma E[\phi'_t \phi'_T] u - w v \) uniformly on compacts as \( c \to \infty \). The limiting ODE:
\[
\dot{v}(t) = \tilde{A} \theta - \gamma E[\phi'_t \phi'_T] u(t) - w v(t)
\]
has \( \lambda^{(1)}_{\infty}(z, u, \theta) = \frac{\tilde{A} \theta - \gamma E[\phi'_t \phi'_T] u}{w} \) as its unique g.a.s.e. \( \lambda^{(1)}_{\infty} \) is Lipschitz with \( \lambda^{(1)}_{\infty}(0, 0, 0) = 0 \), thus satisfying assumption (B.3.1).

The function, \( h_c^{(2)}(v, z, u, \theta) = \frac{c\tilde{A} \theta + \bar{b} - c \bar{C} u - c w z}{c \epsilon} \rightarrow h_c^{(2)}(v, z, u, \theta) = \tilde{A} \theta - \bar{C} u - w z \) uniformly on compacts as \( c \to \infty \). The limiting ODE
\[
\dot{z}(t) = \tilde{A} \theta - \bar{C} u - w z(t)
\]
has \( \lambda^{(2)}_{\infty}(u, \theta) = \frac{\tilde{A} \theta - \bar{C} u}{w} \) as its unique g.a.s.e and \( \lambda^{(2)}_{\infty} \) is Lipschitz continuous with \( \lambda^{(2)}_{\infty}(0, 0) = 0 \). Thus assumption (B.4.2) is satisfied.

The function, \( h_c^{(3)}(u, \theta) = \frac{c\tilde{A} \theta + \bar{b} - c \bar{C} u}{c \epsilon} \rightarrow h_c^{(3)}(u, \theta) = \frac{\tilde{A} \theta - \bar{C} u}{w} \) uniformly on compacts as \( c \to \infty \). The limiting ODE
\[
\dot{u}(t) = \frac{\tilde{A} \theta - \bar{C} u(t)}{w}
\]
has \( \lambda^{(3)}_{\infty}(\theta) = \bar{C}^{-1} \tilde{A} \theta \) as its unique g.a.s.e since \( -\bar{C} \) is negative definite. \( \lambda^{(3)}_{\infty} \) is Lipschitz with \( \lambda^{(3)}_{\infty}(0) = 0 \) and thus assumption (B.4.3) is satisfied as well. Finally,
\[
h_c^{(4)}(\theta) = \frac{c\tilde{A} \theta - c \gamma E[\phi'_t \phi'_T] \bar{C}^{-1} \tilde{A} \theta}{c \epsilon} \rightarrow h_c^{(4)}(\theta) = \frac{(I - \gamma E[\phi'_t \phi'_T] \bar{C}^{-1}) \tilde{A} \theta}{w}
\]
uniformly on compacts as \( c \to \infty \). Consider the ODE:
\[
\dot{\theta}(t) = \frac{(I - \gamma E[\phi'_t \phi'_T] \bar{C}^{-1}) \tilde{A} \theta(t)}{w}
\]
As in the proof of Theorem 5.2, the above ODE has the origin as its unique g.a.s.e. This ensures the final condition (B.3.4). Further, as before, \( \|e_{t+1}\|, \|e_{t}^{(4)}\| \to 0 \) as \( t \to \infty \). By Theorem 3.4,
\[
\begin{pmatrix}
\lambda^{(1:4)}_{\infty}(\theta)
\end{pmatrix}
\rightarrow
\begin{pmatrix}
\lambda^{(1:4)}(-\tilde{A}^{-1} \bar{b})
\end{pmatrix}
\]
Specifically, \( \theta_t \to -\tilde{A}^{-1} \bar{b} \). \hfill \Box

### 5.3 Experiments

The Gradient TD algorithms along with their momentum variants are evaluated on two standard MDPs: 5-State Random Walk [31] and Boyan Chain [11]. We consider decreasing step-sizes of the form: \( \alpha_t^{(1)} = \frac{1}{(t+1)^{0.75}}, \alpha_t^{(2)} = \frac{1}{(t+1)^{0.75}} \), \( \beta_t = \frac{1}{(t+1)^{0.75}} \), \( \alpha_t = \frac{1}{(t+1)^{0.75}} \), respectively, in all the examples. In the 3-Timescale case the conditions on step size turn
out to be $\alpha < \rho + \beta$, and $\beta < \rho$, while in the 4-Timescale case the conditions are $\alpha < \beta + \rho^{(1)} - \rho^{(2)}$, $\beta < 2\rho^{(2)}$ and $\rho^{(2)} < \rho^{(1)}$. Our analysis of convergence requires square-summability of step-size sequences. However, such a choice is seen to slow down the convergence of the algorithm. Recently, [13] provided convergence rate results for Gradient TD schemes with non-square-summable step-sizes (See Remark 2 of [13]). Motivated by this, we look at non-square summable step-sizes for our experiments, and observe that the iterates empirically converge in such cases as well.

For a detailed description of the MDPs, see Figure 1 and 2. See Figure 3 for the results on the 5-State Random Walk and Figure 4 for results on the Boyan Chain. The exact values of $\rho^{(1)}$, $\rho^{(2)}$, $\alpha$, and $\beta$ are provided in Table 1 and $w = 0.1$ for both 3-TS and 4-TS settings. Our results indicate that adding momentum terms to the algorithms clearly improves performance over their vanilla counterparts. Nonetheless, it is less clear from these experiments as to whether adding momentum term to both the iterates is better as opposed to doing the same for only one of the iterates. In particular, for GTD2, the 4-TS scheme appears to perform better while for TDC, the 3-TS version is seen to be better. Further analysis of the finite time behaviour of these algorithms needs to be carried out in the future to better assess the performance of these algorithms.

### 6 Conclusions

In this work we have provided an easily verifiable set of sufficient conditions for stability and convergence of general $N$-timescale stochastic recursions with a martingale difference noise sequence, along with characterizing the limit of all the $N$ recursions. We then used these results to show that stochastic approximation algorithms with an added heavy ball term in the context of Gradient TD methods can be shown to converge a.s. to the same TD solution asymptotically. There are several directions for further research. A natural direction to pursue, as mentioned in Remark 1, would be to come up with sufficient conditions for the stability and convergence of $N$-timescale stochastic recursions with Markov noise.
In [27], stability conditions for single timescale stochastic approximation recursions with Markov noise have been provided. The results in [27] also apply to the case where the underlying Markov process does not possess a unique stationary distribution and can in fact depend on not just the underlying parameters but also an additional control sequence. In [19], the convergence of two-timescale stochastic approximation with Markov noise is studied assuming that the iterates remain stable. Separately, in [28], the convergence of two-timescale recursions with set-valued maps is analysed (in the absence of Markov noise) but assuming stability. In [33], the convergence of more general two-timescale stochastic approximation with set-valued maps and general Markov noise is analysed again assuming iterate-stability. Analysis of algorithms with set-valued maps is important as it paves the way for analysis of RL algorithms under partial observations/information. Thus, a natural extension of our results will be towards deriving sufficient conditions for both stability and convergence of stochastic approximation with (a) general Markov noise and (b) set-valued maps instead of the usual point-to-point maps as are normally considered and analysed.

Finally, Section 5 analyzed the asymptotic behaviour of the momentum algorithms. Further analysis of their finite time behaviour is called for to quantify the benefits of using momentum schemes in stochastic approximation. Towards this extension of weak convergence rate analysis of [20, 25] in the 2-TS setting and recent convergence rate results in expectation and high probability of 2-TS methods in [14, 17, 18, 12] to the $N$-timescale case would also be interesting directions to explore further.

References

[1] E. Altman. *Constrained Markov decision processes*. Vol. 7. CRC Press, 1999.
[2] K. Avrachenkov, K. Patil, and G. Thoppe. “Online Algorithms for Estimating Change Rates of Web Pages”. In: arXiv 2009.08142 (2020).
[3] S. Bhatnagar. “An actor–critic algorithm with function approximation for discounted cost constrained Markov decision processes”. In: *Systems & Control Letters* 59.12 (2010), pp. 760–766.
[4] S. Bhatnagar and J. R. Panigrahi. “Actor-critic algorithms for hierarchical Markov decision processes”. In: *Automatica* 42.4 (2006), pp. 637–644.
[5] S. Bhatnagar, H.L. Prasad, and L.A. Prashanth. *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods*. Springer, 2013.
[6] S. Bhatnagar et al. “Natural actor–critic algorithms”. In: *Automatica* 45.11 (2009), pp. 2471–2482.
[7] V. S. Borkar. “Stochastic approximation with two time scales”. In: *Systems & Control Letters* 29.5 (1997), pp. 291–294.
[8] V. S. Borkar and S. P. Meyn. “The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning”. In: SIAM Journal on Control and Optimization 38.2 (2000), pp. 447–469.
[9] V. S. Borkar. “An actor-critic algorithm for constrained Markov decision processes”. In: Systems & control letters 54.3 (2005), pp. 207–213.
[10] V. S. Borkar. Stochastic Approximation: A Dynamical Systems Viewpoint. Cambridge University Press, 2008. ISBN: 9780521515924.
[11] J. Boyan. “Least-Squares Temporal Difference Learning”. In: ICML. 1999.
[12] G. Dalal, B. Szorenyi, and G. Thoppe. “A Tale of Two-Timescale Reinforcement Learning with the Tightest Finite-Time Bound”. In: Proceedings of the AAAI Conference on Artificial Intelligence 34.04 (2020), pp. 3701–3708.
[13] G. Dalal et al. Finite Sample Analysis of Two-Timescale Stochastic Approximation with Applications to Reinforcement Learning. 2018. arXiv:1703.05376 [cs.AI].
[14] G. Dalal et al. “Finite Sample Analysis of Two-Timescale Stochastic Approximation with Applications to Reinforcement Learning”. In: Proceedings of the 31st Conference On Learning Theory. Ed. by Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet. Vol. 75. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1199–1233.
[15] C. Dann, G. Neumann, and J. Peters. “Policy Evaluation with Temporal Differences: A Survey and Comparison”. In: Journal of Machine Learning Research 15.24 (2014), pp. 809–883.
[16] R. Deb and S. Bhatnagar. Gradient Temporal Difference with Momentum: Stability and Convergence. 2021. arXiv:2111.11004 [cs.LG].
[17] H. Gupta, R. Srikant, and Lei Ying. Finite-Time Performance Bounds and Adaptive Learning Rate Selection for Two Time-Scale Reinforcement Learning. 2019. arXiv:1907.06290 [cs.LG].
[18] M. Kaledin et al. “Finite Time Analysis of Linear Two-timescale Stochastic Approximation with Markovian Noise”. In: Conference on Learning Theory 125 (2019), pp. 2144–2203.
[19] P. Marbach and J.N. Tsitsiklis. “A Convergent O(n) Temporal-difference Algorithm for Off-policy Learning with Linear Function Approximation”. In: Advances in Neural Information Processing Systems. Ed. by D. Koller et al. Vol. 21. Curran Associates, Inc., 2009.
[20] M. L. Puterman. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, 2014.
[21] A. Mokkadem and M. Pelletier. “Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms”. In: The Annals of Applied Probability 16.3 (2006), pp. 1711–1702.
[22] J. Tsitsiklis. “A stochastic approximation method”. In: The annals of mathematical statistics (1951), pp. 400–407.
[33] V.G. Yaji and S. Bhatnagar. “Stochastic recursive inclusions in two timescales with nonadditive iterate-dependent Markov noise”. In: *Mathematics of Operations Research* 45.4 (2020), pp. 1405–1444.