Proceedings of Machine Learning Research vol 75:1–35, 2018 31st Annual Conference on Learning Theory

Finite Sample Analysis of Two-Timescale Stochastic Approximation with Applications to Reinforcement Learning

Gal Dalal* GALD@CAMPUS.TECHNION.AC.IL *Technion, Israel*
Balázs Szörényi* SZORENYI.BALAZS@GMAIL.COM *Yahoo Research, NYC*
Gugan Thoppe* GUGAN.THOPPE@GMAIL.COM *Duke University, USA*
Shie Mannor SHIE@EE.TECHNION.AC.IL *Technion, Israel*

Editors: Sébastien Bubeck, Vianney Perchet and Philippe Rigollet

Abstract

Two-timescale Stochastic Approximation (SA) algorithms are widely used in Reinforcement Learning (RL). Their iterates have two parts that are updated using distinct stepsizes. In this work, we develop a novel recipe for their finite sample analysis. Using this, we provide a concentration bound, which is the first such result for a two-timescale SA. The type of bound we obtain is known as "lock-in probability". We also introduce a new projection scheme, in which the time between successive projections increases exponentially. This scheme allows one to elegantly transform a lock-in probability into a convergence rate result for projected two-timescale SA. From this latter result, we then extract key insights on stepsize selection. As an application, we finally obtain convergence rates for the projected two-timescale RL algorithms GTD(0), GTD2, and TDC.

1. Introduction

Stochastic Approximation (SA) is the subject of a vast literature, both theoretical and applied (Kushner and Yin, 1997). It is used for finding optimal points or zeros of a function for which only noisy access is available. Consequently, SA lies at the core of machine learning; in particular, it is widely used in Reinforcement Learning (RL) and, more so, when function approximation is used.

A powerful, commonly used analysis tool for SA algorithms is the Ordinary Differential Equation (ODE) method (Borkar and Meyn, 2000). Its underlying idea is that, under the right conditions, the noise effects eventually average out and the SA iterates then closely track the trajectory of the so-called "limiting ODE". The ODE method is classically used as a convenient recipe for showing asymptotic SA convergence. The RL literature, therefore, has several results of such type, especially when the state-space is large and function approximation is used (Sutton et al., 2009a,b, 2015; Bhatnagar et al., 2009b). Contrarily, finite sample analyses for SA are scarce; in fact, they are nonexistent in the case of two-timescale SA. This provides the motivation for our work.

1.1. Related Work

A broad, rigorous study of SA is given in (Borkar, 2008); in particular, it contains concentration bounds for single-timescale methods. A more recent work (Thoppe and Borkar, 2015) obtains tighter concentration bounds under weaker assumptions for single-timescale SA using a variational

* Equal contribution.

© 2018 G. Dalal, B. Szörényi, G. Thoppe & S. Mannor.

methodology called Alekseev's Formula. In the context of single-timescale RL, Konda (2002); Korda and Prashanth (2015); Dalal et al. (2018) discuss convergence rates for TD(0).

Convergence rate results for two-timescale SA are, on the other hand, relatively scarce. Asymptotic convergence rates appear in (Spall, 1992; Gerencsér, 1997; Konda and Tsitsiklis, 2004; Mokkadem and Pelletier, 2006); these are of different nature than the finite-time analysis conducted in our work. In the case of two-timescale RL methods, relevant literature can be partitioned into two principal classes: actor-critic and gradient Temporal Difference (TD). In an actor-critic setting, a policy is being evaluated by the critic in the fast timescale, and improved by the actor in the slow timescale; two asymptotic convergence guarantees appear in (Peters and Schaal, 2008; Bhatnagar et al., 2009b). The second class, gradient TD methods, was introduced in (Sutton et al., 2009a). This work presented the GTD(0) algorithm, which is a gradient descent variant of TD(0); being applicable to the so-called off-policy setting, it has a clear advantage over TD(0). Later variants, GTD2 and TDC, were reported to be faster than GTD(0) while enjoying its benefits. These three methods were shown to asymptotically converge in the case of linear and non-linear function approximation (Sutton et al., 2009a,b; Bhatnagar et al., 2009a). Separately, there also exists a convergence rate result for altered versions of the GTD family (Liu et al., 2015). There, projections are used and the learning rates are set to a fixed ratio. The latter makes the altered algorithms single-timescale variants of the original ones.

1.2. Our Contributions

Our main contributions are the following:

- Inspired by (Borkar, 2008), we develop a novel recipe for finite sample analysis of linear two-timescale SA. An initial key step here is a transformation of the iterates (see Remark 3), which we believe can be elevated to general (non-linear) two-timescale settings. Then, by employing the Variation of Parameters method, we obtain a tighter bound on the distance between the SA trajectories and suitable limiting ODE solutions than the one handled in (Borkar, 2008).

- Using the above recipe, we obtain a concentration bound for linear two-timescale SA (see Theorem 4); this is the first such result for two-timescale SA of any kind. In literature, such concentration bounds are also known as "lock-in probability".

- Additionally, we introduce a novel projection scheme, in which the time between successive projections progressively doubles; we refer to it as "sparse projection". This scheme enables one to elegantly transform a concentration bound, of the type we obtain, into a convergence rate for projected two-timescale SA (see Theorem 6). We stress the strength of this tool in bridging the gap between two research communities: those who are interested in lock-in probabilities/concentration bounds, and those who care about convergence rates.

- As an application, we obtain convergence rates for the sparsely projected variants of two-timescale RL algorithms: GTD(0), GTD2, and TDC. This is the first finite time result for the above algorithms in their true two-timescale form (see Remark 1).

- Finally, we do away with the usual square summability assumption on stepsizes (see Remark 2). Therefore, our tool is relevant for a broader family of stepsizes. An example of its usefulness is Polyak-Ruppert-averaging with constant stepsizes (Défossez and Bach, 2014;

Lakshminarayanan and Szepesvari, 2018), whose behavior, we believe, is similar to two-timescale algorithms with slowly-decaying non-square-summable stepsizes (e.g., $n^{-\alpha}$ with α close to 0).

2. Preliminaries

Here we present the linear two-timescale SA paradigm, state our goal, and list our assumptions.

A generic linear two-timescale SA is

$$\theta_{n+1} = \theta_n + \alpha_n[h_1(\theta_n, w_n) + M^{(1)}_{n+1}] \ , \tag{1}$$

$$w_{n+1} = w_n + \beta_n[h_2(\theta_n, w_n) + M^{(2)}_{n+1}] \ , \tag{2}$$

where $\alpha_n, \beta_n \in \mathbb{R}$ are stepsizes, $M^{(i)}_n \in \mathbb{R}^d$ denotes noise, and $h_i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ has the form

$$h_i(\theta, w) = v_i - \Gamma_i \theta - W_i w \tag{3}$$

for a vector $v_i \in \mathbb{R}^d$ and matrices $\Gamma_i, W_i \in \mathbb{R}^{d\times d}$.

Remark 1 *In this work, we are interested in the analysis of a "true two-timescale process". By this, we mean that Γ_2 ought to be invertible and that $\alpha_n/\beta_n \to 0$. The first condition couples the two iterates together; nevertheless, all the results in this work hold even without this restriction. The second condition is indeed assumed throughout (see $\mathcal{A}_2$ below); we do not allow α_n/β_n to converge to a positive constant, as that would then turn* (1) *and* (2) *into a single-timescale SA.*

Our aim is to finite time behaviour of (1) and (2) under the following assumptions.

$\mathcal{A}_1$. W_2 and $X_1 := \Gamma_1 - W_1 W_2^{-1}\Gamma_2$ are positive definite (not necessarily symmetric).

$\mathcal{A}_2$. Stepsize sequences $\{\alpha_n\}$, $\{\beta_n\}$, and $\{\eta_n := \alpha_n/\beta_n\}$ satisfy

$$\sum_{n=0}^{\infty}\alpha_n = \sum_{n=0}^{\infty}\beta_n = \infty, \ \ \alpha_n, \beta_n, \eta_n \le 1, \text{ and } \lim_{n\to\infty}\alpha_n = \lim_{n\to\infty}\beta_n = \lim_{n\to\infty}\eta_n = 0. \tag{4}$$

$\mathcal{A}_3$. $\{M^{(1)}_n\}, \{M^{(2)}_n\}$ are martingale difference sequences w.r.t. the family of $\sigma-$fields $\{\mathcal{F}_n\}$, where $\mathcal{F}_n = \sigma(\theta_0, w_0, M^{(1)}_1, M^{(2)}_1, \ldots, M^{(1)}_n, M^{(2)}_n)$. There exist constants $m_1, m_2 > 0$ so that $\left\|M^{(1)}_{n+1}\right\| \le m_1(1 + \|\theta_n\| + \|w_n\|)$ and $\left\|M^{(2)}_{n+1}\right\| \le m_2(1 + \|\theta_n\| + \|w_n\|)$ for all $n \ge 0$.

Remark 2 *Notice that, unlike most works, $\sum_{n\ge 0}\alpha_n^2$ or $\sum_{n\ge 0}\beta_n^2$ need not be finite. Thus, our analysis is applicable for a wider class of stepsizes; e.g., $1/n^\kappa$ with $\kappa \in (0, 1/2]$. In (Borkar, 2008), on which much of the existing RL literature is based on, the square summability assumption is due to the Gronwall inequality based approach. In contrast, for the specific setting here, we do a tighter analysis using the variation of parameters formula (Lakshmikantham and Deo, 1998).*

We now briefly outline the ODE method from (Borkar, 2008, pp. 64-65) for the analysis of (1) and (2), and also describe how our approach builds upon it. Since $\eta_n \to 0$, $\{w_n\}$ is the fast transient and $\{\theta_n\}$ is the slow component. Therefore, the ODE that (2) might be expected to track is

$$\dot{w}(t) = v_2 - \Gamma_2\theta - W_2 w(t) \tag{5}$$

for some fixed θ, and the ODE that (1) might be expected to track is

$$\dot{\theta}(t) = h_1(\theta(t), \lambda(\theta(t))) = b_1 - X_1\theta(t), \tag{6}$$

where $b_1 := v_1 - W_1 W_2^{-1} v_2$ and $\lambda(\theta) := W_2^{-1}[v_2 - \Gamma_2\theta]$. Due to $\boldsymbol{\mathcal{A}_1}$, the function $\lambda(\cdot)$ and b_1 are well defined. Moreover, $\lambda(\theta)$ and $\theta^* := X_1^{-1} b_1$ are unique globally asymptotically stable equilibrium points of (5) and (6), respectively.

Lemma 1, (Borkar, 2008, p. 66), applied to (1) and (2) gives $\lim_{n\to\infty} \|w_n - \lambda(\theta_n)\| = 0$ under suitable assumptions. Inspired by this, we work with $\{z_n\}$ here instead of $\{w_n\}$ directly, where

$$z_n := w_n - \lambda(\theta_n) \ . \tag{7}$$

Due to (2), $\{z_n\}$ satisfies the update rule

$$z_{n+1} = z_n - \beta_n W_2 z_n + \beta_n M_{n+1}^{(2)} + \lambda(\theta_n) - \lambda(\theta_{n+1}) \ . \tag{8}$$

Hence, and as $\{\theta_n\}$ is the slow component, the limiting ODE that (8) might be expected to track is

$$\dot{z}(s) = -W_2 z(s) \ . \tag{9}$$

As W_2 is positive definite (see $\boldsymbol{\mathcal{A}_1}$), $z^* = 0$ is the globally asymptotically stable equilibrium of (9).

Remark 3 *Using $\{z_n\}$ instead of $\{w_n\}$ is the main reason why our approach works. Observe that the limiting ODE in* (5) *varies as θ_n evolves; in contrast,* (9) *remains unchanged. Hence, comparing* (8) *with* (9) *is easier than comparing* (2) *with* (5)*. While this idea is indeed inspired by (Borkar, 2008, Lemma 1, p. 66), there* (8) *and* (9) *are not required to be explicitly dealt with.*

3. Main Results

In this section, we give our two main results on two-timescale stochastic approximation and also introduce our projection scheme. The first result is a general concentration bound for any stepsizes satisfying $\boldsymbol{\mathcal{A}_2}$. This result concerns the behavior of a two-timescale SA from some time index n_0 onwards and requires that the iterates be bounded at n_0. This is in the spirit of most existing concentration bounds/lock-in probability results for single-timescale methods (Borkar, 2008; Thoppe and Borkar, 2015). By projecting the iterates of a two-timescale SA via our novel projection scheme, we then transform our above concentration bound into a convergence rate result. This latter result applies for all time indices and the boundedness assumption holds here due to projections.

3.1. A General Concentration Bound

Let $q_1, q_2 > 0$ be lower bounds on the real part of the eigenvalues of matrices X_1 and W_2, respectively. For $n \geq 0$, let $a_n := \sum_{k=0}^{n-1} \alpha_k^2 e^{-2q_1 \sum_{i=k+1}^{n-1} \alpha_i}$ and $b_n := \sum_{k=0}^{n-1} \beta_k^2 e^{-2q_2 \sum_{i=k+1}^{n-1} \beta_i}$. These sums are obtained from the Azuma-Hoeffding concentration bound that we use later. Also, let

$$s_n := \sum_{k=0}^{n-1} \beta_k, \quad \text{and} \quad t_n := \sum_{k=0}^{n-1} \alpha_k \ . \tag{10}$$

Theorem 4 gives our concentration bound; the additional terms in it are defined in Tables 1 and 2.

Constant	Source	Constant	Source
K_1, K_2	(68), (70)	$L_{1b}^{\text{te}} = K_1 \\|W_1\\| \\|W_2\\| R_2^{\text{in}}/q_1$	Lemma 20
q_1, q_2	above (10)	$L_{1c}^{\text{te}} = K_1 \\|W_1\\| /q_1$	Lemma 20
$q_{\min} = \min\{q_1, q_2\}$	(71)	$L_1^{\text{md}} = K_1 m_1[1 + R^* + R_1^{\text{out}} + R_2^w]$	Lemma 20
$q \in (0, q_{\min})$	(71)	$L_1^{\text{de}} = \frac{K_1 \\|X_1\\| J^\theta}{q_1}$	Lemma 20
$R_1^{\text{out}} = R_1^{\text{in}} + \frac{4K_1\\|W_1\\|K_2 R_2^{\text{in}}}{(q_{\min}-q)e}$	(56)	$L_a^\theta = L_{1a}^{\text{te}}$	Lemma 21
$R^* = \\|X_1^{-1}\\| \\|b_1\\|$	(63)	$L_c^\theta = L_{1c}^{\text{te}}$	Lemma 21
$R_2^w = R_2^{\text{out}} + \\|W_2^{-1}\\|$	(65)	$L_b^\theta = L_1^{\text{de}} + L_1^{\text{md}} + \\|X_1\\| R_1^{\text{in}} + L_{1b}^{\text{te}}$	Lemma 21
$\times [\\|v_2\\| + \\|\Gamma_2\\| [R^* + R_1^{\text{out}}]]$		$L_2^{\text{md}} = K_2 m_2[1 + R^* + R_1^{\text{out}} + R_2^w]$	Lemma 18
$R_1^{\text{gap}} = R_1^{\text{out}} - R_1^{\text{in}}$, $R_2^{\text{gap}} = R_2^{\text{out}} - R_2^{\text{in}}$	(57)	$L_2^{\text{sd}} = K_2 \frac{\\|W_2^{-1}\\| \\|\Gamma_2\\| J^\theta}{q_2}$	Lemma 18
$J^\theta = \\|\Gamma_1\\| [R^* + R_1^{\text{out}}] + \\|W_1\\| R_2^w$	Lemma 17	$L_2^{\text{de}} = K_2 \frac{\\|W_2\\| J^z}{q_2}$	Lemma 18
$+ \\|v_1\\| + m_1[1 + R^* + R_1^{\text{out}} + R_2^w]$		$L^z = \\|W_2\\| R_2^{\text{in}} + L_2^{\text{de}} + L_2^{\text{sd}} + L_2^{\text{md}}$	Lemma 19
$J^z = \\|W_2\\| R_2^{\text{out}} + \\|W_2^{-1}\\| \\|\Gamma_2\\| J^\theta$	Lemma 17	$c_1 = (16K_1^2 d^3 [L_1^{\text{md}}]^2)^{-1}$	Theorem 4
$+ m_2(1 + R^* + R_1^{\text{out}} + R_2^w)$		$c_2 = (9K_2^2 d^3 [L_2^{\text{md}}]^2)^{-1}$	Theorem 4
$L_{1a}^{\text{te}} = K_1 \\|W_1\\| K_2 R_2^{\text{in}} \frac{1}{(q_{\min}-q)e}$	Lemma 20	$c_3 = (64K_2^2 [L_c^\theta]^2 d^3 [L_2^{\text{md}}]^2)^{-1}$	Theorem 4

Table 1: A summary of constants and where they are defined. Here $m_1, m_2 > 0$ are as in $\mathcal{A}_3$, and $R_1^{\text{in}} > 0$ and $R_2^{\text{out}} > R_2^{\text{in}} > 0$ are constants chosen as in Theorems 4 and 6. Note that constants in the left column do not depend on constants in the right column. Similarly, no constant depends on constants below it in the same column, or on ϵ_1, ϵ_2, $\{\alpha_k\}$ or $\{\beta_k\}$.

Theorem 4 (Main Technical Result) *Fix some constants $R_1^{in}, R_2^{in} > 0$ and $R_2^{out} > R_2^{in}$. Pick $\epsilon_1 \in \left(0, \min\left\{R_1^{in}, 4L_a^\theta\right\}\right)$ and $\epsilon_2 \in (0, \min(R_2^{in}, R_2^{out} - R_2^{in}))$. Fix some $n_0 \geq N_0$ and $n_1 \geq N_1$, where $N_0 \equiv N_0(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$ and $N_1 \equiv N_1(n_0, \epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$ are as in Table 2. Consider the process defined by (1) and (2) for $n \geq n_0$, initialized at arbitrary $\theta_{n_0}, w_{n_0} \in \mathbb{R}^d$ such that*

$$\|\theta_{n_0} - \theta^*\| \leq R_1^{in} \text{ and } \|z_{n_0}\| \leq R_2^{in}, \tag{11}$$

where z_{n_0} is as in (7). Then,

$$\Pr\{\|\theta_n - \theta^*\| \leq \epsilon_1, \|z_n\| \leq \epsilon_2, \forall n \geq n_1\} \geq 1 - 2d^2 \sum_{n \geq n_0} \left[\exp\left[\frac{-c_1 \epsilon_1^2}{a_n}\right] + \exp\left[\frac{-c_2 \epsilon_1^2}{b_n}\right] + \exp\left[\frac{-c_3 \epsilon_2^2}{b_n}\right]\right]. \tag{12}$$

where $c_1 = (16K_1^2 d^3 [L_1^{md}]^2)^{-1}$, $c_2 = (9K_2^2 d^3 [L_2^{md}]^2)^{-1}$, and $c_3 = (64K_2^2 [L_c^\theta]^2 d^3 [L_2^{md}]^2)^{-1}$ are constants independent of $\epsilon_1, \epsilon_2, \{\alpha_k\}$ and $\{\beta_k\}$.

Proof See Section 4 for the outline of the proof, and Appendix D for the detailed proof. ■

Section 2 already discusses the close relation between the SA iterates $\{\theta_n\}$ and $\{z_n\}$ and the corresponding ODE trajectories, which suggests that the analysis of the former should be based on the latter. However, the sole fact that the ODE trajectories approach their respective solutions does not guarantee the same for the SA trajectories. The latter may drift away due to several factors (e.g., martingale noise), as discussed in Subsection 4.2. However, Theorem 4 makes it clear that, w.h.p., this does not happen. These subtleties are discussed in more details in the following remark.

Term	Definition
$N_{0,a} \equiv N_{0,a}(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$	$\min\left\{N : \max\left\{\sup_{k\geq N} \beta_k,\ \sup_{k\geq N} \eta_k\right\} \leq \dfrac{\min\{\epsilon_1/8, \epsilon_2/3\}}{L^z \max\{L_c^\theta, 1\}}\right\}$
$N_{0,b} \equiv N_{0,b}(\epsilon_1, \{\beta_k\})$	$\min\left\{N : \sup_{k\geq N} \beta_k \leq \epsilon_1/(4L_b^\theta)\right\}$
$N_0 \equiv N_0(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$	$\max\{N_{0,a}, N_{0,b}\}$
$N_{1,a} \equiv N_{1,a}(n_0, \epsilon_1, \{\alpha_k\})$	$\min\{j \geq n_0 : [K_1 R_1^{\text{in}} + L_a^\theta] e^{-q(t_j - t_{n_0})} \leq \epsilon_1/4\}$
$N_{1,b} \equiv N_{1,b}(n_0, \epsilon_2, \{\beta_k\})$	$\min\{j \geq n_0 : K_2 R_2^{\text{in}} e^{-q_2(s_j - s_{n_0})} \leq \epsilon_2/3\}$
$N_1 \equiv N_1(n_0, \epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$	$\max\{N_{1,a}, N_{1,b}\}$

Table 2: A summary of terms depending on ϵ_1, ϵ_2, and the stepsize sequences, as appearing in the main theorems. These terms are formally introduced in Lemmas 22 and 23.

Remark 5 *Theorem 4 involves two key notions introduced in Table 2:* N_0 *and* N_1.

1. *A large* N_0 *ensures the stepsizes are small enough to mitigate the factors that may cause the SA trajectories to drift. In the case of martingale difference noise, this can be directly seen from the terms* $\alpha_n M_{n+1}^{(1)}$ *and* $\beta_n M_{n+1}^{(2)}$ *in* (1) *and* (8).
2. *The term* N_1 *is an intrinsic property of the limiting ODEs. It quantifies the number of iterations required by the two ODE trajectories to hit the* ϵ*-neighborhoods of their respective solutions (and stay there) when started in* R_1^{in} *and* R_2^{in} *radii balls. As shown in Theorem 4,* N_1 *depends on* n_0. *A larger* n_0 *means smaller stepsizes, which implies that a longer time is required for the trajectories to hit the* ϵ*-neighbourhoods, in turn making* N_1 *larger.*

3.2. A Bound for Sub-exponential Series

In order to make Theorem 4 more applicable, we derive closed form expressions for the r.h.s. of (12) for the case of inverse polynomial stepsizes; see Appendix C. In particular, we obtain a bound on the generic expression $\sum_{n=n_0}^{\infty} \exp[-Bn^p]$, where $B \geq 0$ and $p \in (0,1)$. Such expressions are common in SA analyses. Thus, this result can be useful on its own.

3.3. Convergence Rate of Sparsely Projected Iterates

Here we first describe our projection scheme, following which we give our convergence rate result in Theorem 6. In this latter result, we work with a specific family of stepsizes to obtain concrete closed-form expressions for the rate of convergence.

For n that is a power of 2, let $\Pi_{n,R}$ denote the projection into the R-ball; for every other n, let $\Pi_{n,R}$ denote the identity, where $R > 0$ is some arbitrary constant. We call this sparse projection as we project only on indices which are powers of 2. With $\theta_0', w_0' \in \mathbb{R}^d$, let

$$\theta'_{n+1} = \Pi_{n+1, R_1^{\text{in}}/2}\left(\theta'_n + \alpha_n[h_1(\theta'_n, w'_n) + M_{n+1}^{(1')}]\right) \ , \tag{13}$$

$$w'_{n+1} = \Pi_{n+1, R_2^{\text{in}}/2}\left(w'_n + \beta_n[h_2(\theta'_n, w'_n) + M_{n+1}^{(2')}]\right) \tag{14}$$

denote the sparsely projected variant of (1) and (2) , where $\{M_n^{(1')}\}$ and $\{M_n^{(2')}\}$ are martingale difference sequences satisfying assumption $\mathcal{A}_3$, just like $\{M_n^{(1)}\}$ and $\{M_n^{(2)}\}$. The idea of projection is indeed common (Borkar, 2008; Kushner, 1980); but, the novelty here is in doing only exponentially infrequent projections. As seen in the proof of Theorem 6, this significantly simplifies our analysis.

We now introduce a carefully chosen instantiation of N_0 (where N_0 is as in Table 2) for the stepsize choice in Theorem 6 below. This choice also regulates the N_1 term in an appropriate way, as we show later in the theorem's analysis. For some $1 > \alpha > \beta > 0$ and $\epsilon \in (0,1)$, let

$$N_0'(\epsilon,\alpha,\beta) := \max\Bigg\{ \left[\tfrac{8L^z}{\epsilon}\max\left\{L_c^\theta, 1\right\}\right]^{\frac{1}{\min\{\beta,\alpha-\beta\}}}, \left[\tfrac{4L_b^\theta}{\epsilon}\right]^{1/\beta}, \left[\tfrac{1-\alpha}{((1.5)^{1-\alpha}-1)q}\ln\tfrac{4[K_1R_1^{\text{in}}+L_a^\theta]}{\epsilon}\right]^{\frac{1}{1-\alpha}}, \left[\tfrac{1-\beta}{((1.5)^{1-\beta}-1)q_2}\ln\tfrac{3K_2R_2^{\text{in}}}{\epsilon}\right]^{\frac{1}{1-\beta}}, 3\Bigg\} \ . \tag{15}$$

Theorem 6 (Finite Time Behavior of Sparsely Projected Iterates) *Fix $R_1^{in}, R_2^{in} > 0$. Suppose*

$$\|\theta^*\| \le R_1^{in}/4 \ \text{, and} \tag{16}$$

$$\{\theta\in\mathbb{R}^d : \|\theta\| \le R_1^{in}/2\} \subseteq \{\theta\in\mathbb{R}^d : \|\lambda(\theta)\| \le R_2^{in}/4\} \ . \tag{17}$$

Let $\alpha_n = (n+1)^{-\alpha}$ and $\beta_n = (n+1)^{-\beta}$ with $1 > \alpha > \beta > 0$. Then the following hold.

1. *For any $R_2^{out} > R_2^{in}$, $\epsilon \in (0, \min\{R_1^{in}/4, R_2^{in}/4, 4L_a^\theta, R_2^{out} - R_2^{in}\})$, and $n_0' \ge N_0'(\epsilon,\alpha,\beta)$, such that n_0' is a power of 2 and $N_0'(\epsilon,\alpha,\beta) = O\left(\epsilon^{-\frac{1}{\min\{\beta,\alpha-\beta\}}}\right)$ is as in (15), we have*

$$\begin{aligned}&\Pr\{\|\theta_n' - \theta^*\| \le \epsilon, \|z_n'\| \le \epsilon, \forall n \ge 2n_0'\} \\ &\ge 1 - \frac{2d^2c_{7a}}{\epsilon^{2/\alpha}}\exp\left[c_{5a}\epsilon^2 - c_{6a}\epsilon^2(n_0')^\alpha\right] - \frac{4d^2c_{7b}}{\epsilon^{2/\beta}}\exp\left[c_{5b}\epsilon^2 - c_{6b}\,\epsilon^2(n_0')^\beta\right] \ ,\end{aligned} \tag{18}$$

   where $c_4 := \min\{c_2, c_3\}$, $c_{5a} = c_5(c_1,\kappa,\alpha,q_1)$, $c_{5b} = c_5(c_4,\kappa,\beta,q_2)$, and so on for c_{6a}, c_{7a}, c_{6b}, and c_{7b}. The terms c_5, c_6, and c_7 are as defined in Lemma 13. [1]

2. *There is some constant $C > 0$ such that, for all $n > 3$ and $\delta \in (0,1)$, it holds that* [2]

$$\Pr\left\{\max\{\|\theta_n' - \theta^*\|, \|z_n'\|\} \le C\max\left[n^{-\beta/2}\sqrt{\ln(n/\delta)}\ ,\ n^{-(\alpha-\beta)}\right]\right\} \ge 1-\delta \ . \tag{19}$$

Proof See Appendix E. ■

Remark 7 *To the best of our knowledge, the only other work that provides a high probability convergence rate for projected SA algorithm is (Liu et al., 2015), and it also assumes (16). Without this assumption, one would be required to study the convergence to the closest point to θ^* within $R_1^{in}/4$. Assumption (17) can be easily seen to hold if R_2^{in} is set to $4\|W_2^{-1}v_2\| + 2R_1^{in}\|\Gamma_2\|$ or greater.*

Remark 8 *In continuation of Remark 2, notice α and β are not constrained to lie in $(1/2,1)$. This is, to the best of our knowledge, in contrast with any other two-timescale analysis in the literature.*

1. Consult Table 1 for the rest of the constants, such as c_1, c_2 and c_3.
2. An explicit expression for C can be derived from the proof of Theorem 6 which, for brevity, we haven't introduced.

Remark 9 *Clearly, the tightest possible upper bound in* (19) *approaches* $O(n^{-1/3})$ *as* α *and* β *simultaneously approach* 1 *and* $2/3$*, respectively. We now briefly discuss the origin of the two limiting terms there. The* $n^{-\beta/2}$ *term stems from the convergence of* (14)*, and corresponds to the known* $n^{-1/2}$ *rate limit of any single-timescale SA. It is also in line with Theorem 3.1 in (Dalal et al., 2018) for generic* β*. We note that a similar* $n^{-\alpha/2}$ *rate, stemming from the convergence of* (13)*, originally appears in the proof of Theorem 6; however, we drop it from the statement since* $\alpha > \beta$*. Separately, the* $n^{-(\alpha-\beta)}$ *term stems from the interaction between the* θ *and* z *iterates, originating in the last two terms in* (8)*. As discussed above Remark 3, the slow component* $\{\theta_n\}$ *evolves in the* α_n *timescale; yet, it is part of* z_n *update rule, which evolves in the* β_n *timescale. Hence, the slow drift error (see Subsection 4.2) is governed by the stepsize ratio* α_n/β_n *(yielding the* $\alpha - \beta$ *here).*

In transforming Theorem 4 to Theorem 6, N_0' (see (15)) inherits the properties of both N_0 and N_1 from Theorem 4, whose roles we have portrayed in Remark 5. Theorem 6, along with (15), relates the above roles to the choice of α and β; it suggests several valuable tradeoffs between speeding up convergence of the noiseless ODE, and mitigating the martingale noise and other drift factors (see Subsection 4.2) to aid the SA to follow this process. Namely, N_0' explodes:

1. As α or β approach 0 (stepsizes approach constants); this stems from N_0 blowing up. This occurs since the stepsizes' slow decay rate impairs i) their ability to mitigate the martingale noise and other drift factors; and hence ii) the ability of the SA to track the ODE trajectories.

2. As α and β get close to each other; this is due to N_0 blowing up. This occurs as the true two-timescale nature is then nullified (see Remark 1). Our analysis suggests that convergence of z_n to z^* must be faster than that of θ_n to θ^*, and a decaying stepsize ratio η_n ensures this.

3. As α or β approach 1, the largest value for which (4) holds; this stems from N_1 blowing up. This occurs as the stepsizes then decay too fast, impairing the speed of the ODE convergence; more accurately, N_1 (see Table 2) moves away from exponential nature to inverse polynomial.

4. Proof Outline of Theorem 4

In this section, we bring the essence of the proof of Theorem 4. For intermediate results and the complete proof, see Appendix D. Naturally, throughout this section, we assume (11). Also, as mentioned in Section 2, we work here with the iterates $\{z_n\}$ defined using (8) instead of $\{w_n\}$ directly. As stated in Remark 3, our analysis follows through thanks to this choice.

The proof has two key steps. First, in Subsection 4.2, we use the Variation of Parameters (VoP) formula (Lakshmikantham and Deo, 1998) to quantify the distance between the SA trajectories generated with (1) and (8) and suitable solutions of their respective limiting ODEs (6) and (9).

As for the second step, note that the choice of N_0 in Theorem 4 ensures that $\{\beta_k\}_{k\geq N_0}$ and $\{\eta_k\}_{k\geq N_0}$ are sufficiently small—i.e., of order $O(\max(\epsilon_1, \epsilon_2))$. Exploiting this fact and using the Azuma-Hoeffding inequality, in Subsection 4.3 we show that the bounds on the distances obtained in the first step are small with very high probability. More explicitly, when the ODE solutions are sufficiently close to θ^* and z^* respectively, we show that the same also holds for the sequences $\{\theta_n\}$ and $\{z_n\}$ with high probability. A visualization of the process is given in Fig. 1.

Before discussing these two steps, we now introduce some notations and terminology.

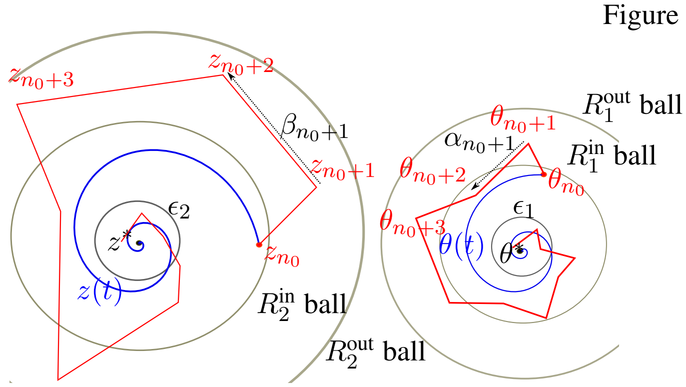

Figure 1: Visualization of the proof methodology. The red SA trajectories $\{\theta_n\}$ and $\{z_n\}$ are compared to their blue respective limiting ODE trajectories $\theta(t)$ and $z(s)$. The three balls on each side of the figure (from small to large), are respectively the solution's ϵ-neighborhood; the R^{in} ball in which the SA trajectory and ODE trajectory are initialized; and the R^{out} ball in which the SA trajectory is ensured to reside.

4.1. Analysis Preliminaries

To begin with, we define the linearly interpolated trajectories of the iterates $\{\theta_n\}$ and $\{z_n\}$. Having a continuous version of the discrete SA algorithm enables our analysis. Keeping (10) in mind, let $\bar\theta(\cdot)$ be the linear interpolation of $\{\theta_n\}$ on $\{t_n\}$; i.e., let $\bar\theta(t_n) = \theta_n$ and, for $\tau \in (t_n, t_{n+1})$, let

$$\bar\theta(\tau) = \bar\theta(t_n) + \tfrac{(\tau - t_n)}{\alpha_n}[\bar\theta(t_{n+1}) - \bar\theta(t_n)]. \tag{20}$$

Similarly, let $\bar z(\cdot)$ be the linear interpolation of $\{z_n\}$, but on $\{s_n\}$. For $\tau \in [t_n, t_{n+1})$, let

$$\xi(\tau) := s_n + \tfrac{\beta_n}{\alpha_n}(\tau - t_n). \tag{21}$$

The mapping $\xi(\cdot)$ linearly interpolates $\{s_n\}$ on $\{t_n\}$.

With the first parameter being time, the second being starting time, and the third being initial point, let $\theta(t, t_{n_0}, \theta_{n_0})$, $t \ge t_{n_0}$, be the solution to (6) satisfying $\theta(t_{n_0}, t_{n_0}, \theta_{n_0}) = \theta_{n_0}$. Similarly, define $z(s, s_{n_0}, z_{n_0})$. From (6) and standard ODE results (see (Hirsch et al., 2012, p. 129)),

$$\theta(t) \equiv \theta(t, t_{n_0}, \theta_{n_0}) = \theta^* + e^{-X_1(t-t_{n_0})}(\theta_{n_0} - \theta^*), \quad \forall t \ge t_{n_0}. \tag{22}$$

In the same way, it follows from (9) that

$$z(s) \equiv z(s, s_{n_0}, z_{n_0}) = e^{-W_2(s-s_{n_0})} z_{n_0}, \quad \forall s \ge s_{n_0}. \tag{23}$$

Remark 10 *Since X_1 is positive definite due to $\mathcal{A}_1$, (22) implies that $\lim_{t\to\infty}\theta(t, t_{n_0}, \theta_{n_0}) = \theta^*$. Further, $\frac{d}{dt}\|\theta(t) - \theta^*\|^2 = -2(\theta(t) - \theta^*)^\top X_1(\theta(t) - \theta^*) < 0$; hence, assuming (11) holds, $\|\theta(t, t_{n_0}, \theta_{n_0}) - \theta^*\| \le R_1^{in}$, $\forall t \ge t_{n_0}$. Likewise, we have $\lim_{s\to\infty} z(s, s_{n_0}, z_{n_0}) = z^*$ and $\|z(s, s_{n_0}, z_{n_0})\| \le R_2^{in}$ for all $s \ge s_{n_0}$.*

4.2. Comparing the SA and corresponding Limiting ODE Trajectories

Our aim here is to use the VoP formula to bound $\|\bar z(s) - z(s)\|$ and $\left\|\bar\theta(t) - \theta(t)\right\|$. Note that both the SA trajectory $\bar\theta(t)$ and the corresponding limiting ODE trajectory $\theta(t)$ equal θ_{n_0} at time $t = t_{n_0}$. Similarly, $\bar z(s_{n_0}) = z(s_{n_0}) = z_{n_0}$.

Using (7), (1) translates to $\theta_{n+1} = \theta_n + \alpha_n[b_1 - X_1\theta_n] + \alpha_n[-W_1 z_n] + \alpha_n M^{(1)}_{n+1}$. Iteratively using the above update rule, we then have

$$\theta_{n+1} = \theta_{n_0} + \sum_{k=n_0}^{n} \alpha_k[b_1 - X_1\theta_k - W_1 z_k + M^{(1)}_{k+1}] \ .$$

From this and the definition of $\bar{\theta}(\cdot)$, it consequently follows that

$$\bar{\theta}(t) = \theta_{n_0} + \int_{t_{n_0}}^{t} [b_1 - X_1\bar{\theta}(\tau)]\mathrm{d}\tau + \int_{t_{n_0}}^{t} \zeta(\tau)\mathrm{d}\tau, \ \ \forall t \geq t_{n_0} \ , \tag{24}$$

with $\zeta(\tau) := \zeta^{\mathrm{de}}(\tau)+\zeta^{\mathrm{te}}(\tau)+\zeta^{\mathrm{md}}(\tau)$, where, for $\tau \in [t_k, t_{k+1})$, $\zeta^{\mathrm{de}}(\tau) := X_1[\bar{\theta}(\tau)-\theta_k]$, $\zeta^{\mathrm{te}}(\tau) := -W_1 z_k$, and $\zeta^{\mathrm{md}}(\tau) := M^{(1)}_{k+1}$. Let $E^{\mathrm{de}}_1(t) = \int_{t_{n_0}}^{t} e^{-X_1(t-\tau)}\zeta^{\mathrm{de}}(\tau)\mathrm{d}\tau$. Define $E^{\mathrm{te}}_1(t)$ and $E^{\mathrm{md}}_1(t)$ in the same spirit. We refer to these three terms as the discretization error, tracking error, and martingale difference noise, respectively. The tracking error is called so, because it depends on $z_k = w_k - \lambda(\theta_k)$ which, by (8), tells how close w_k is to its ODE solution $\lambda(\theta_k)$. From (6), we have $\theta(t) = \theta_{n_0} + \int_{t_{n_0}}^{t}[b_1 - X_1\theta(\tau)]\mathrm{d}\tau$, and thus (24) can be viewed as a perturbation of $\theta(t)$. Defining then $E_1(t) := E^{\mathrm{de}}_1(t) + E^{\mathrm{te}}_1(t) + E^{\mathrm{md}}_1(t)$, and applying the VoP formula (see Appendix D.1), it follows easily that

$$\bar{\theta}(t) = \theta(t, t_{n_0}, \theta_{n_0}) + E_1(t) \ . \tag{25}$$

Using (8), it is easy to see in the same way as above that

$$\bar{z}(s) = z_{n_0} + \int_{s_{n_0}}^{s} [-W_2]\bar{z}(\mu)\mathrm{d}\mu + \int_{s_{n_0}}^{s} \chi(\mu)\mathrm{d}\mu, \ \ \forall s \geq s_{n_0} \ , \tag{26}$$

with $\chi(\mu) := \chi^{\mathrm{de}}(\mu) + \chi^{\mathrm{sd}}(\mu) + \chi^{\mathrm{md}}(\mu)$, where, for $\mu \in [s_k, s_{k+1})$,

$$\chi^{\mathrm{de}}(\mu) := W_2[\bar{z}(\mu) - z_k], \quad \chi^{\mathrm{sd}}(\mu) := \frac{\lambda(\theta_k) - \lambda(\theta_{k+1})}{\beta_k}, \quad \chi^{\mathrm{md}}(\mu) := M^{(2)}_{k+1} \ . \tag{27}$$

Let $E^{\mathrm{de}}_2(s) = \int_{s_{n_0}}^{s} e^{-W_2(s-\mu)}\chi^{\mathrm{de}}(\mu)\mathrm{d}\mu$. Define $E^{\mathrm{sd}}_2(t)$ and $E^{\mathrm{md}}_2(t)$ in the same spirit. We refer to these three terms as discretization error, slow drift in the equilibrium of (5), and martingale difference noise. We refer to $\chi^{\mathrm{sd}}(\mu)$ as the slow drift error because as $\{\theta_n\}$ evolve, the ODE solution $\{\lambda(\theta_n)\}$ drift, and it is slow since $\{\theta_n\}$ is updated on the slow time scale $\{t_n\}$ (recall that $\eta_n \to 0$). Finally, defining $E_2(t) := E^{\mathrm{de}}_2(t) + E^{\mathrm{sd}}_2(t) + E^{\mathrm{md}}_2(t)$, it follows simlarly as above that

$$\bar{z}(s) = z(s, s_{n_0}, z_{n_0}) + E_2(s) \ . \tag{28}$$

The below result is now trivial to see.

Lemma 11 *The following two statements hold:*

1. *For* $t \geq t_{n_0}$, $\|\bar{\theta}(t) - \theta(t, t_{n_0}, \theta_{n_0})\| \leq \|E^{de}_1(t)\| + \|E^{te}_1(t)\| + \|E^{md}_1(t)\|$.
2. *For* $s \geq s_{n_0}$, $\|\bar{z}(s) - z(s, s_{n_0}, z_{n_0})\| \leq \|E^{de}_2(s)\| + \|E^{sd}_2(s)\| + \|E^{md}_2(s)\|$.

To stress the tightness of the above analysis, we compare it with that in (Borkar, 2008, p. 14). There, the distance between the SA and ODE trajectories is bounded by the tail sum of the squared stepsizes; this necessitates the usual square summability assumption. We do not require it here thanks to the additional exponentials, $e^{-X_1(t-\tau)}$ and $e^{-W_2(s-\mu)}$, in the error terms $E^{\mathrm{de}}_1(t)$, $E^{\mathrm{de}}_2(s)$, etc., which is a consequence of the VoP formula.

4.3. Concentration Bounds for Two-Timescale SA

Next, with Lemma 11 bounding the distance of $\bar{\theta}(t)$ and $\bar{z}(s)$ from their respective ODE trajectories for all t and s, we consequently bound the distance of $\bar{\theta}(t)$ and $\bar{z}(s)$ from the solutions θ^* and z^*. To do so, we break the convergence event into an incremental union using a novel inductive technique (see Appendix D.2, Lemma 14). Each event in the union has the following structure: "good" up to time n (ensured by an event G_n, where the iterates remain bounded in certain regions) and "bad" in the subsequent interval ($\bar{\theta}(t_{n+1})$ and $\bar{z}(s_{n+1})$ leave the bounded regions). By conditioning on G_n, and using (11) with Lemma 11, we bound $\left\|\bar{\theta}(t) - \theta^*\right\|$ and $\|\bar{z}(s) - z^*\|$. Each of the resulting bounds consists of three kinds of terms (see Appendix D.4, Lemmas 19 and 21): i) sum of martingale differences (originating in E_i^{md}), ii) stepsize based term (originating in $E_i^{\text{de}},\ E_1^{\text{te}},\ E_2^{\text{sd}}$), and iii) exponentially decaying term (originating in the ODE trajectory convergence). Type i) terms are small w.h.p. due to the Azuma-Hoeffding inequality; these terms give the r.h.s. in (12). Type ii) terms are small since N_0 is chosen sufficiently large (consult Table 1 for the definition of N_0). Type iii) terms are small for n sufficiently larger than N_0 (in particular, for $n > N_1$—consult Table 1 for the definition of N_1). This summarizes the proof of Theorem 4, which is described in Appendix D.5.

5. Applications to Two-timescale Reinforcement Learning

Here we show how our Theorem 6 implies convergence rates of linear two-timescale methods for policy evaluation in Markov Decision Processes (MDP). An MDP is defined by the 5-tuple $(\mathcal{S}, \mathcal{A}, P, R, \gamma)$ (Sutton, 1988), where these are respectively the state and action spaces, transition kernel, reward function, and discount factor. Let policy $\pi : \mathcal{S} \to \mathcal{A}$ be a stationary mapping from states to actions and $V^\pi(s) = \mathbb{E}^\pi[\sum_{n=0}^\infty \gamma^n r_n | s_0 = s]$ be the value function at state s w.r.t π.

We consider the policy evaluation setting. In it, the goal is to estimate the value function $V^\pi(s)$ with respect to a given π using linear regression, i.e., $V^\pi(s) \approx \theta^\top \phi(s)$, where $\phi(s) \in \mathbb{R}^d$ is a feature vector at state s, and $\theta \in \mathbb{R}^d$ is a parameter vector. For brevity, we omit the notation π and denote $\phi(s_n),\ \phi(s'_n)$ by $\phi_n,\ \phi'_n$. Finally, let $\delta_n = r_n + \gamma \theta_n^\top \phi'_n - \theta_n^\top \phi_n$, $A = \mathbb{E}[\phi(\phi - \gamma\phi')^\top]$, $C = \mathbb{E}[\phi\phi^\top]$, and $b = \mathbb{E}[r\phi]$, where the expectations are w.r.t. the stationary distribution of the induced chain [3].

We assume all rewards $r(s)$ and feature vectors $\phi(s)$ are bounded: $|r(s)| \le 1, \|\phi(s)\| \le 1\ \forall s \in S$. Also, it is assumed that the feature matrix Φ is full rank, so A and C are full rank. This assumption is standard (Maei et al., 2010; Sutton et al., 2009a). Therefore, due to its structure, A is also positive definite (Bertsekas, 2012). Moreover, by construction, C is positive semi-definite; thus, by the full-rank assumption, it is actually positive definite.

5.1. The GTD(0) Algorithm

We now present the GTD(0) algorithm (Sutton et al., 2009a), verify its required assumptions, and obtain the necessary constants to apply Theorem 6 for it. GTD(0) is designed to minimize the objective function $J^{\text{NEU}}(\theta) = \frac{1}{2}(b - A\theta)^\top (b - A\theta)$. Its update rule is

$$\theta_{n+1} = \theta_n + \alpha_n \left(\phi_n - \gamma\phi'_n\right) \phi_n^\top w_n, \qquad w_{n+1} = w_n + \beta_n r_n \phi_n + \phi_n[\gamma\phi'_n - \phi_n]^\top \theta_n.$$

3. The samples $\{(\phi_n, \phi'_n)\}$ are generated iid. This assumption is standard when dealing with convergence bounds in reinforcement learning (Liu et al., 2015; Sutton et al., 2009a,b). In the few papers where this assumption is not made, it is replaced with an exponentially-fast mixing time assumption (Korda and Prashanth, 2015; Tsitsiklis et al., 1997).

Method	X_1	W_2	m_1	m_2
GTD(0)	$A^\top A$	$\mathbb{I}$	$(1+\gamma+\|A\|)$	$1+\max(\|b\|,\gamma+\|A\|)$
GTD2	$A^\top C^{-1}A$	C	$(1+\gamma+\|A\|)$	$1+\max(\|b\|,\gamma+\|A\|,\|C\|)$
TDC	$A^\top C^{-1}A$	C	$(2+\gamma+\|A\|+\|C\|)$	$(2+\gamma+\|A\|+\|C\|)$

Table 3: Translation of notations for relevant matrices and constants in the case of the GTD family of algorithms. The parameters $X_1,\ W_2,\ m_1,\ m_2$ are defined in Section 2.

It thus takes the form of (1) and (2) with $h_1(\theta,w) = A^\top w$, $h_2(\theta,w) = b - A\theta - w$, $M^{(1)}_{n+1} = (\phi_n - \gamma\phi'_n)\,\phi_n^\top w_n - A^\top w_n$, $M^{(2)}_{n+1} = r_n\phi_n + \phi_n[\gamma\phi'_n - \phi_n]^\top\theta_n - (b - A\theta_n)$. That is, in case of GTD(0), the relevant matrices in the update rules take the form $\Gamma_1 = 0$, $W_1 = -A^\top$, $v_1 = 0$, and $\Gamma_2 = A$, $W_2 = \mathbb{I}$, $v_2 = b$. Additionally, $X_1 = \Gamma_1 - W_1 W_2^{-1}\Gamma_2 = A^\top A$. By our assumption above, both W_2 and X_1 are symmetric positive definite matrices, and thus the real parts of their eigenvalues are also positive. Also, $\|M^{(1)}_{n+1}\| \le (1+\gamma+\|A\|)\|w_n\|$, $\|M^{(2)}_{n+1}\| \le 1 + \|b\| + (1+\gamma+\|A\|)\|\theta_n\|$. Hence, $\mathcal{A}_3$ is satisfied with constants $m_1 = (1+\gamma+\|A\|)$ and $m_2 = 1+\max(\|b\|,\gamma+\|A\|)$.

We now can apply Theorem 6 for a specific stepsize choice to obtain the following simplified result. A more detailed statement with all relevant constants can be directly derived from Theorem 6.

Corollary 12 (Convergence Rate for Sparsely Projected GTD(0)) *Consider the Sparsely Projected variant of GTD(0) as in (13) and (14). Set some $\kappa \in (0,1)$. Then for $\alpha_n = 1/n^{1-\kappa}$, $\beta_n = 1/n^{(2/3)(1-\kappa)}$, the algorithm converges at a rate of $O(n^{-1/3+\kappa/3})$ w.h.p.*

For GTD2 and TDC (Sutton et al., 2009b), the above result can be similarly be reproduced; see Table 3 for the relevant parameters. The detailed derivation is provided in Appendix F.

A reviewer has pointed us to the fact that, unlike in the GTD(0) and GTD2 convergence results, there exists a special condition on the stepsize ratio for TDC (Maei, 2011, Theorem 3). However, we find that this condition to be unnecessary because A and C are positive definite.

6. Discussion

In this work, we conduct the first finite sample analysis for two-timescale SA algorithms. We provide it as a general methodology that applies to all linear two-timescale SA algorithms.

A natural extension to our methodology is considering the non-linear function-approximation case, in a similar fashion to (Thoppe and Borkar, 2015). Such a result can be of high interest due to the recently growing attractiveness of neural networks in the RL community. An additional direction for future research is to extend our results to actor-critic RL algorithms. Moreover, off-policy extensions can be made for the results here; see Appendix A. Lastly, for a discussion on the tightness of the results here and comparison to known asymptotic rates see Appendix B.

Acknowledgments

This research was supported by ERC grant 306638 (SUPREL). GT was initially supported by ERC grant 320422 (at Technion) and now by NSF grants DMS-1613261, DMS-1713012, IIS-1546413.

References

D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Vol II. Athena Scientific, fourth edition, 2012.

Shalabh Bhatnagar, Doina Precup, David Silver, Richard S Sutton, Hamid R Maei, and Csaba Szepesvári. Convergent temporal-difference learning with arbitrary smooth function approximation. In *Advances in Neural Information Processing Systems*, pages 1204–1212, 2009a.

Shalabh Bhatnagar, Richard Sutton, Mohammad Ghavamzadeh, and Mark Lee. Natural actor-critic algorithms. *Automatica*, 45(11), 2009b.

Vivek S Borkar. *Stochastic approximation: a dynamical systems viewpoint*. 2008.

Vivek S Borkar and Sean P Meyn. The ode method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000.

Gal Dalal, Balazs Szorenyi, Gugan Thoppe, and Shie Mannor. Finite sample analyses for td(0) with function approximation. In *AAAI*, 2018.

Alexandre Défossez and Francis Bach. Constant step size least-mean-square: Bias-variance trade-offs and optimal sampling distributions. *arXiv preprint arXiv:1412.0156*, 2014.

László Gerencsér. Rate of convergence of moments of spall's spsa method. In *Control Conference (ECC), 1997 European*, pages 2192–2197. IEEE, 1997.

Morris W Hirsch, Stephen Smale, and Robert L Devaney. *Differential equations, dynamical systems, and an introduction to chaos*. Academic press, 2012.

J Zico Kolter. The fixed points of off-policy td. In *Advances in Neural Information Processing Systems*, pages 2169–2177, 2011.

Vijay R Konda and John N Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *Annals of applied probability*, pages 796–819, 2004.

Vijaymohan Konda. *Actor-Critic Algorithms*. PhD thesis, Department of Electrical Engineering and Computer Science, MIT, 6 2002.

Nathaniel Korda and LA Prashanth. On td (0) with function approximation: Concentration bounds and a centered variant with exponential convergence. In *ICML*, pages 626–634, 2015.

H Kushner. A projected stochastic approximation method for adaptive filters and identifiers. *IEEE Transactions on Automatic Control*, 25(4):836–838, 1980.

Harold J. Kushner and G. George Yin. *Stochastic Approximation Algorithms and Applications*. 1997.

Vangipuram Lakshmikantham and Sadashiv G Deo. *Method of variation of parameters for dynamic systems*. CRC Press, 1998.

Chandrashekar Lakshminarayanan and Shalabh Bhatnagar. A stability criterion for two timescale stochastic approximation schemes. *Automatica*, 79:108–114, 2017.

Chandrashekar Lakshminarayanan and Csaba Szepesvari. Linear stochastic approximation: How far does constant step-size and iterate averaging go? In *International Conference on Artificial Intelligence and Statistics*, pages 1347–1355, 2018.

Bo Liu, Ji Liu, Mohammad Ghavamzadeh, Sridhar Mahadevan, and Marek Petrik. Finite-sample analysis of proximal gradient td algorithms. In *UAI*, pages 504–513. Citeseer, 2015.

Hamid Reza Maei. Gradient temporal-difference learning algorithms. 2011.

Hamid Reza Maei, Csaba Szepesvári, Shalabh Bhatnagar, and Richard S Sutton. Toward off-policy learning control with function approximation. In *ICML*, pages 719–726, 2010.

Abdelkader Mokkadem and Mariane Pelletier. Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *The Annals of Applied Probability*, 16(3):1671–1702, 2006.

Jan Peters and Stefan Schaal. Natural actor-critic. *Neurocomputing*, 71(7):1180–1190, 2008.

James C Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control*, 37(3):332–341, 1992.

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

Richard S Sutton, Hamid R Maei, and Csaba Szepesvári. A convergent o(n) temporal-difference algorithm for off-policy learning with linear function approximation. In *Advances in neural information processing systems*, pages 1609–1616, 2009a.

Richard S Sutton, Hamid Reza Maei, Doina Precup, Shalabh Bhatnagar, David Silver, Csaba Szepesvári, and Eric Wiewiora. Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 993–1000. ACM, 2009b.

Richard S Sutton, A Rupam Mahmood, and Martha White. An emphatic approach to the problem of off-policy temporal-difference learning. *The Journal of Machine Learning Research*, 17:1–29, 2015.

Gerald Teschl. Ordinary differential equations and dynamical systems. 2004.

Gugan Thoppe and Vivek S Borkar. A concentration bound for stochastic approximation via alekseev's formula. *arXiv:1506.08657*, 2015.

John N Tsitsiklis, Benjamin Van Roy, et al. An analysis of temporal-difference learning with function approximation. *IEEE transactions on automatic control*, 42(5):674–690, 1997.

Appendix A. Off-Policy Extensions

Off-policy results play a central role in reinforcement learning; however, we were focusing here exclusively on the on-policy setting. Nonetheless, our results can be similarly extended as in (Liu et al., 2015). Namely, we can repeat the elegant reduction conducted there, where the bound on $\|\theta_n - \theta^*\|$ is transformed into one on the approximation error $\|V - \bar{v}_n\| = \|V - \Phi\bar{\theta}_n\|$. More precisely, we can bound the second term on the RHS in (Liu et al., 2015, Appendix B, (42)) using (Kolter, 2011, Theorem 2), and apply our result to bound the first one. Except for a slightly different rescaling of the matrices (since we use L2 norm as opposed to ξ-weighted L2), we would then obtain an off-policy result as in Proposition 5 there. Two benefits would then be: a result directly consisting of θ_n (instead of its average), and a generic stepsize family $n^{-\alpha}$ (instead of $C/\sqrt{n}$, where $C = f(\|A\| + \|b\|)$, as depicted above (Liu et al., 2015, Appendix B, (40)). Notice, also, that transforming one type of bound to the other, as explained above, is a trick by (Liu et al., 2015) that can be applied in general and not only in our case.

Appendix B. Tightness

Here, we compare our convergence rates with other existing works. To the best of our knowledge, no other finite time results exist for two-timescale SA algorithms. However, there are a few relevant works that deal with this question in an asymptotic sense. Before discussing them, we highlight some key differences between finite-time rates and aysmptotic ones. In the latter, the constants hidden in the order notations are often sample-path dependent and hence are less attractive to practitioners. Contrarily, explicit constants in finite-time rates, as ours, often reveal intriguing dependencies amongst system and stepsize parameters that crucially affect convergence rates (e.g., $1/q_i$ in Table 1; see also (Dalal et al., 2018, Section 6)). Moreover, trajectory-independent constants help in obtaining stopping time theorems.

Following Remark 9, the best convergence rate possible according to our results is $n^{-1/3}$. This contrasts the single time-scale case, where the optimal rate is known to be $n^{-1/2}$ under various settings. In the context of asymptotic rates, there exist two works that deal with two-timescale SA which achieve the optimal rate of $n^{-\alpha/2}$ for the slow-timescale iterate and $n^{-\beta/2}$ for the fast-timescale iterate (Konda and Tsitsiklis, 2004; Mokkadem and Pelletier, 2006). In (Konda and Tsitsiklis, 2004), according to Assumption 2.1, the noise sequence is assumed to be independent of itself, and their variance-covariance matrices are constant w.r.t. iteration index n. In our case, in contrast, the noise depends on (θ_n, w_n), making the variance-covariance matrices of the noise sequence explicitly depend on n. These differences make their results inapplicable for the RL algorithms we consider in our paper; see Section 5.1. A later work (Mokkadem and Pelletier, 2006) improved upon (Konda and Tsitsiklis, 2004) by removing the above assumption. There, in (A1), convergence was posed as an assumption on its own. Such an assumption is not straighforward to verify in general; it was only recently established for square-summable stepsizes (Lakshminarayanan and Bhatnagar, 2017). However, in the case of non-square-summable stepsizes (which is not covered in (Mokkadem and Pelletier, 2006)) this Assumption (A1) has not been shown to hold in general, since converegence is an open question for such stepsizes.

Lastly, while we do not show our bound to be tight, we stress that our result coincides with known results on a particular SA method of two-timescale nature, called Spall's method (Spall, 1992, Proposition 2) and (Gerencsér, 1997, Theorem 5.1). Specifically, it was shown for the iterate θ_n there that $n^{-\kappa}\theta_n$ converges in distribution to some normal distribution for various parameter settings that restrict κ to be at least $1/3$. This raises the intriguing question whether the rates achieved in our work and in (Spall, 1992; Gerencsér, 1997) are sub-optimal and stem from loose analyses, or whether it is the problem setup itself that intrinsically limits the rate to $n^{-1/3}$.

Appendix C. A Bound for Sub-exponential Series

Let $p \in (0,1)$ and $\hat{q} > 0$. Let $i_1 \equiv i_1(p, \hat{q}) \geq 1$ be such that $e^{-\hat{q}\sum_{k=1}^{n-1}(k+1)^{-p}} \leq n^{-p}$ for all $n \geq i_1$; such an i_1 exists as the l.h.s. is exponentially decaying. Let

$$K_g \equiv K_g(p, \hat{q}) := \max_{1 \leq i \leq i_1} i^p e^{-\hat{q}\sum_{k=1}^{i-1}(k+1)^{-p}}. \tag{29}$$

Lemma 13 (Closed-form sub-exponential bounds) *Let $n_0 \geq 1$, $B > 0$, and $p \in (0,1)$. Then, for every $\kappa \in (0,1)$,*

$$\sum_{n=n_0}^{\infty} \exp[-Bn^p] \leq \frac{2}{B(1-\kappa)p}\left[\frac{(1-p)}{B\kappa p}\right]^{\frac{1-p}{p}} \exp\left[B(2-\kappa) - \frac{(1-p)}{p} - B(1-\kappa){n_0}^p\right] . \tag{30}$$

Further, for any $c, \hat{q} > 0$, and $n_0 \geq 1$, with $c_n := \sum_{k=0}^{n-1}[k+1]^{-2p}e^{-2\hat{q}\sum_{i=k+1}^{n-1}[i+1]^{-p}}$, we have

$$\sum_{n \geq n_0} \exp\left[\frac{-c\epsilon^2}{c_n}\right] \leq \frac{c_7}{\epsilon^{2/p}} e^{c_5\epsilon^2} e^{-c_6\epsilon^2 n_0^p} \quad , \tag{31}$$

where $c_7 \equiv c_7(c, \kappa, p, \hat{q}) = 2\left[\frac{K_g(p,\hat{q})e^{\hat{q}}}{c\hat{q}}\right]^{1/p}\frac{1}{(1-\kappa)p^{1/p}}\left[\frac{1-p}{e\kappa}\right]^{\frac{1-p}{p}}$, $c_5 \equiv c_5(c, \kappa, p, \hat{q}) = \frac{c\hat{q}(2-\kappa)}{K_g(p,\hat{q})e^{\hat{q}}}$, and $c_6 \equiv c_6(c, \kappa, p, \hat{q}) = \frac{c\hat{q}(1-\kappa)}{K_g(p,\hat{q})e^{\hat{q}}}$.

Proof Let $\lfloor \cdot \rfloor$ denote the floor operation. Then, for $p \in (0,1)$ and integers $n, i \geq 0$, we have

$$\begin{aligned} & |\{n : \lfloor n^p \rfloor = i\}| && (32)\\ =& |\{n : i \leq n^p < i+1\}| \\ =& |\{n : i^{1/p} \leq n < (i+1)^{1/p}\}| \\ \leq& (i+1)^{\frac{1}{p}} - i^{\frac{1}{p}} + 1 \\ \leq& 2\left[(i+1)^{\frac{1}{p}} - i^{\frac{1}{p}}\right], && (33) \end{aligned}$$

where the last inequality follows since $(i+1)^{\frac{1}{p}} - i^{\frac{1}{p}} \geq 1$.

From the concavity of x^p, $x^p \leq x_0^p + \frac{d}{dx}(x^p)\big|_{x=x_0}(x - x_0)$ for all $x, x_0 \in \mathbb{R}_+$. Equivalently,

$$x_0 - x \leq (x_0^p - x^p)\left[\frac{d}{dx}(x^p)\big|_{x=x_0}\right]^{-1}.$$

Setting $x_0 = (i+1)^{\frac{1}{p}}$ and $x = i^{\frac{1}{p}}$, it follows from (33) that

$$\begin{aligned}
&|\{n : \lfloor n^p \rfloor = i\}| \\
\le& 2\left[(i+1)^{\frac{1}{p}} - i^{\frac{1}{p}}\right] \\
\le& 2\left[\left((i+1)^{\frac{1}{p}}\right)^p - \left(i^{\frac{1}{p}}\right)^p\right]\left[px^{p-1}\Big|_{x=(i+1)^{\frac{1}{p}}}\right]^{-1} \\
=& \frac{2}{p}(i+1)^{\frac{1-p}{p}} \quad . \qquad (34)
\end{aligned}$$

For any $\kappa \in (0,1)$, observe that $e^{-xB\kappa}(x+1)^{\frac{1-p}{p}}$, restricted to $x \ge 0$, has a global maximum at $x = \frac{(1-p)}{B\kappa p} - 1$, and so

$$\max_{i\ge 0} e^{-iB\kappa}(i+1)^{\frac{1-p}{p}} \le \left[\frac{1-p}{B\kappa p}\right]^{\frac{1-p}{p}} e^{[\kappa B - \frac{(1-p)}{p}]} \quad . \qquad (35)$$

Fix an arbitrary $\kappa \in (0,1)$. From the above observations, we get

$$\begin{aligned}
&\sum_{n=n_0}^{\infty} \exp[-Bn^p] \\
\le& \sum_{i=\lfloor n_0^p \rfloor}^{\infty} e^{-iB}|\{n : \lfloor n^p \rfloor = i\}| \\
\le& \frac{2}{p}\sum_{i=\lfloor n_0^p \rfloor}^{\infty} e^{-iB}\,(i+1)^{\frac{1-p}{p}} \qquad (36) \\
=& \frac{2}{p}\sum_{i=\lfloor n_0^p \rfloor}^{\infty} e^{-iB(1-\kappa)}\, e^{-iB\kappa}\,(i+1)^{\frac{1-p}{p}} \\
\le& \frac{2}{p}\left[\frac{1-p}{B\kappa p}\right]^{\frac{1-p}{p}} e^{[\kappa B - \frac{(1-p)}{p}]}\sum_{i=\lfloor n_0^p \rfloor}^{\infty} e^{-iB(1-\kappa)} \qquad (37) \\
\le& \frac{2}{p}\left[\frac{1-p}{B\kappa p}\right]^{\frac{1-p}{p}} e^{[\kappa B - \frac{(1-p)}{p}]}\int_{\lfloor n_0^p \rfloor - 1}^{\infty} e^{-xB(1-\kappa)}\mathrm{d}x \qquad (38) \\
\le& \frac{2}{B(1-\kappa)p}\left[\frac{1-p}{B\kappa p}\right]^{\frac{1-p}{p}} e^{B(2-\kappa) - \frac{(1-p)}{p}} e^{-B(1-\kappa)n_0{}^p} \quad ,
\end{aligned}$$

where (36) follows from (34), (37) holds due to (35), and (38) is obtained by treating the sum as a right Riemann sum and using $\lfloor n_0^p \rfloor > n_0^p - 1$. This completes the proof of (30).

We now prove (31). Let $f(x) := (x+1)^p \log[(x+1)/x]$. Notice that $\lim_{x\to\infty} f(x) = 0$ for $x \ge 1,\ p \in (0,1)$ because $f(x)$ is positive for $x > 0$ and

$$(x+1)^p \log[(x+1)/x] \le (x+1)^p/x,$$

which goes to zero. Therefore, there is a $i_0 \equiv i_0(p, \hat{q}) \geq 0$ such that

$$(i+2)^p \log\left[\frac{i+2}{i+1}\right] \geq \frac{\hat{q}}{p} \ , \quad \text{if } 0 \leq i < i_0 \ ,$$
$$(i+2)^p \log\left[\frac{i+2}{i+1}\right] \leq \frac{\hat{q}}{p} \ , \quad \quad \text{if } i \geq i_0 \ .$$

This is equivalent to saying that, for every $n \geq i+2$, if $0 \leq i < i_0$, then

$$(i+1)^{-p} e^{-\hat{q}\sum_{k=i+1}^{n-1}(k+1)^{-p}} \geq (i+2)^{-p} e^{-\hat{q}\sum_{k=i+2}^{n-1}(k+1)^{-p}} \ ,$$

and if $i_0 \leq i \leq n-2$, then

$$(i+1)^{-p} e^{-\hat{q}\sum_{k=i+1}^{n-1}(k+1)^{-p}} \leq (i+2)^{-p} e^{-\hat{q}\sum_{k=i+2}^{n-1}(k+1)^{-p}} \ .$$

Therefore, the maximum of $(i+1)^{-p} e^{-\hat{q}\sum_{k=i+1}^{n-1}(k+1)^{-p}}$ is achieved in one of the terminal values, i.e., at $i = 0$ or $i = n-1$. Thus,

$$\begin{aligned} &\max_{0 \leq i \leq n-1} (i+1)^{-p} e^{-\hat{q}\sum_{k=i+1}^{n-1}(k+1)^{-p}} \\ \leq &\max\{e^{-\hat{q}\sum_{k=1}^{n-1}(k+1)^{-p}}, n^{-p}\} \quad (39) \\ \leq &K_g n^{-p} \ , \quad (40) \end{aligned}$$

where $K_g \geq 1$ (by its definition) is as defined in (29). The transition from (39) to (40) can be seen as follows. First, consider the case $n \geq i_1$, where i_1 is defined above (29). In this case, by the definition of i_1, the maximum in (39) is n^{-p}, which is bounded by $K_g n^{-p}$. If $n < i_1, \max\{n^{-p}\left(n^p e^{-\hat{q}\sum_{k=1}^{n-1}(k+1)^{-p}}\right), n^{-p}\} \leq K_g n^{-p}$ by the definition of K_g.

Now let $u_n := \sum_{k=0}^{n-1}[k+1]^{-p}$. For $n \geq 1$, we then have

$$\begin{aligned} c_n = &\sum_{i=0}^{n-1}[i+1]^{-2p} e^{-2\hat{q}\sum_{k=i+1}^{n-1}[k+1]^{-p}} \\ \leq &K_g n^{-p} \sum_{i=0}^{n-1}[i+1]^{-p} e^{-\hat{q}\sum_{k=i+1}^{n-1}[k+1]^{-p}} \quad (41) \\ = &K_g n^{-p} \sum_{i=0}^{n-1}[u_{i+1} - u_i] e^{-\hat{q}[u_n - u_{i+1}]} \quad (42) \\ \leq &K_g e^{\hat{q}} n^{-p} e^{-\hat{q}u_n} \sum_{i=0}^{n-1}[u_{i+1} - u_i] e^{\hat{q}u_i} \quad (43) \\ \leq &K_g e^{\hat{q}} n^{-p} e^{-\hat{q}u_n} \int_{u_0}^{u_n} e^{\hat{q}s} \mathrm{d}s \quad (44) \\ = &K_g e^{\hat{q}} n^{-p} e^{-\hat{q}u_n} \frac{e^{\hat{q}u_n} - e^{\hat{q}u_0}}{\hat{q}} \\ \leq &\frac{K_g e^{\hat{q}}}{\hat{q}} n^{-p} \ , \quad (45) \end{aligned}$$

where (41) follows from (40), (42) follows using the definition of u_n, (43) holds since $u_{i+1} = u_i + (i+1)^{-p} \leq u_i + 1$ for $i \geq 0$, (44) follows by treating the sum above as a left Riemann sum, and, lastly, (45) holds as $u_0 = 0$ and $e^{-\hat{q}u_n} \leq 1$.

Consequently, for any $c > 0$ and $n_0 \geq 1$,

$$\sum_{n \geq n_0} \exp\left[\tfrac{-c\epsilon^2}{c_n}\right] \leq \sum_{n \geq n_0} \exp\left[-\frac{c\hat{q}\epsilon^2}{K_g e^{\hat{q}}} n^p\right] \quad .$$

The desired result now follows from (30). This completes the proof of the lemma. ∎

Appendix D. Proof of Theorem 4

As the analysis in Section 4 is under assumption (11), the results in the corresponding Subsections D.2, D.4 and D.5 here are under this assumption as well.

D.1. Application of VoP Formula in Subsection 4.2

Recall the definitions given below (24). On the interval $[t_k, t_{k+1})$, the functions $\zeta^{\text{te}}(\cdot)$ and $\zeta^{\text{md}}(\cdot)$ are constant, while $\zeta^{\text{de}}(\cdot)$ is linear. Therefore, the function $\zeta(t)$, $t \geq t_{n_0}$, is piecewise continuous; specifically, it is continuous on the interval $[t_k, t_{k+1})$, for every $k \geq n_0$. Separately, owing to the fact that it is a linear interpolation, the function $\bar{\theta}(t)$, $t \geq t_{n_0}$, is continuous everywhere.

The evolution described in (24) can be viewed as a differential equation in integral form; further, it can be looked at as a perturbation of the ODE in (6). It is then not difficult to see from (Lakshmikantham and Deo, 1998, Theorem 1.1.2) that (25) holds for any $t \in [t_{n_0}, t_{n_0+1})$. Now, from the continuity of $\bar{\theta}(t)$, it follows that (25) holds even for $t = t_{n_0+1}$, i.e.,

$$\bar{\theta}(t_{n_0+1}) = \theta(t_{n_0+1}, t_{n_0}, \theta_{n_0}) + E_1(t_{n_0+1}). \tag{46}$$

Arguing in the same way as above, for any $t \in [t_{n_0+1}, t_{n_0+2})$, it is easy to see that

$$\bar{\theta}(t) = \theta(t, t_{n_0+1}, \bar{\theta}(t_{n_0+1})) + \int_{t_{n_0+1}}^{t} e^{-X_1(t-\tau)}\zeta(\tau)\mathrm{d}\tau \quad . \tag{47}$$

Moreover, observe that

$$\begin{aligned}
& \theta(t, t_{n_0+1}, \bar{\theta}(t_{n_0+1})) && (48)\\
= \;& \theta\left(t, t_{n_0+1}, \theta(t_{n_0+1}, t_{n_0}, \theta_{n_0}) + E_1(t_{n_0+1})\right) && (49)\\
= \;& \theta^* + e^{-X_1(t-t_{n_0+1})}(\theta(t_{n_0+1}, t_{n_0}, \theta_{n_0}) + E_1(t_{n_0+1}) - \theta^*) && (50)\\
= \;& \theta^* + e^{-X_1(t-t_{n_0+1})}(\theta(t_{n_0+1}, t_{n_0}, \theta_{n_0}) - \theta^*) + e^{-X_1(t-t_{n_0+1})}E_1(t_{n_0+1}) &&\\
= \;& \theta^* + e^{-X_1(t-t_{n_0+1})}(\theta(t_{n_0+1}, t_{n_0}, \theta_{n_0}) - \theta^*) + \int_{t_{n_0}}^{t_{n_0+1}} e^{-X_1(t-\tau)}\zeta(\tau)\mathrm{d}\tau && (51)\\
= \;& \theta(t, t_{n_0+1}, \theta(t_{n_0+1}, t_{n_0}, \theta_{n_0})) + \int_{t_{n_0}}^{t_{n_0+1}} e^{-X_1(t-\tau)}\zeta(\tau)\mathrm{d}\tau && (52)\\
= \;& \theta(t, t_{n_0}, \theta_{n_0}) + \int_{t_{n_0}}^{t_{n_0+1}} e^{-X_1(t-\tau)}\zeta(\tau)\mathrm{d}\tau \quad , && (53)
\end{aligned}$$

where (49) follows from (46); (50) and (52) hold as in (22); (51) follows from the definition of E_1 given below (24); and, finally, (53) is true because of the uniqueness and existence of ODE solutions (see Picard-Lindelöf theorem).

Substituting (53) in (47), it is easy to see that (25) holds for all $t \in [t_{n_0+1}, t_{n_0+2})$. Inductively arguing this way, it follows that (25) holds for all $t \geq t_{n_0}$.

D.2. A Useful Decomposition of the Event of Interest

For any event $\mathcal{E}$, let $\mathcal{E}^c$ be its complement. For all $n_0,\ T > 0$, define the event

$$\mathcal{E}(n_0, T) := \{\|\bar{\theta}(t) - \theta^*\| \leq \epsilon_1\ \forall t \geq t_{n_0} + T + 1\} \cap \{\|\bar{z}(s)\| \leq \epsilon_2\ \forall s \geq s_{n_0} + \xi(T) + 1\}\ , \quad (54)$$

where ϵ_1, ϵ_2 are as in the statement of Theorem 4. Eventually, we shall use a bound on $\Pr\{\mathcal{E}^c(n_0, T)\}$ to prove Theorem 4. Towards obtaining this bound, the aim here is to construct a well-structured superset for $\mathcal{E}^c(n_0, T)$, assuming (11) holds, which is easier for analysis.

Fix some $T > 0$ so that

$$T \leq t_{n_1+1} - t_{n_0} = \sum_{k=n_0}^{n_1} \alpha_k \leq T + 1. \quad (55)$$

By Remark 10, $\theta(t, t_{n_0}, \theta_{n_0})$ stays in the $R_1^{\text{in}}-$radius ball around θ^* for all $t \geq t_{n_0}$, and $z(s, s_{n_0}, z_{n_0})$ stays in the $R_2^{\text{in}}-$radius ball around z^* for all $s \geq s_{n_0}$. But the same cannot be said for $\bar{\theta}(t)$ and $\bar{z}(s)$ due to the presence of noise. We show instead that these lie with high probability in bigger but fixed radii balls R_1^{out} and R_2^{out}, where $R_2^{\text{out}} > R_2^{\text{in}}$ is an arbitrary constant, and

$$R_1^{\text{out}} := R_1^{\text{in}} + \frac{4K_1 \|W_1\| K_2 R_2^{\text{in}}}{(q_{\min} - q)e}\ . \quad (56)$$

Note that, by the choice of ϵ_1 and ϵ_2

$$R_1^{\text{gap}} := R_1^{\text{out}} - R_1^{\text{in}} \geq \epsilon_1\ , \text{ and } R_2^{\text{gap}} := R_2^{\text{out}} - R_2^{\text{in}} \geq \epsilon_2\ . \quad (57)$$

For $n \geq n_0$, let

$$\rho_{n+1} := \sup_{\tau \in [t_n, t_{n+1}]} \|\bar{\theta}(\tau) - \theta(\tau, t_{n_0}, \theta_{n_0})\|\ , \quad \rho_{n+1}^* := \sup_{\tau \in [t_n, t_{n+1}]} \|\bar{\theta}(\tau) - \theta^*\|\ , \quad (58)$$

$$\nu_{n+1} := \sup_{\mu \in [s_n, s_{n+1}]} \|\bar{z}(\mu) - z(\mu, s_{n_0}, z_{n_0})\|\ , \quad \nu_{n+1}^* := \sup_{\mu \in [s_n, s_{n+1}]} \|\bar{z}(\mu)\|\ , \quad (59)$$

and define the (“good”) event

$$G_n := \{\|\bar{\theta}(\tau) - \theta^*\| \leq R_1^{\text{out}}\ \forall \tau \in [t_{n_0}, t_n]\} \cap \{\|\bar{z}(\mu)\| \leq R_2^{\text{out}}\ \forall \mu \in [s_{n_0}, s_n]\}. \quad (60)$$

Additionally, define the (“bad”) events $\mathcal{E}_{\text{after}} := \bigcup_{n > n_1} \left[\{\rho_{n+1}^* > \epsilon_1\} \cup \{\nu_{n+1}^* > \epsilon_2\}\right]$ and

$$\mathcal{E}_{\text{mid}} := \left\{\left[\sup_{n_0 \leq n \leq n_1} \rho_{n+1}\right] \geq R_1^{\text{gap}}\right\} \cup \left\{\left[\sup_{n_0 \leq n \leq n_1} \nu_{n+1}\right] \geq R_2^{\text{gap}}\right\}\ .$$

The desired superset stated at the beginning of this subsection is given below.

Lemma 14 (Decomposition of Event of Interest) *Consider* (54) *and suppose that* (11) *holds. Then*

$$\mathcal{E}^c(n_0, T) \subseteq \bigcup_{n=n_0}^{n_1} \{G_n \cap [\{\rho_{n+1} \geq R_1^{gap}\} \cup \{\nu_{n+1} \geq R_2^{gap}\}]\}$$
$$\cup \bigcup_{n>n_1} [G_n \cap [\{\rho_{n+1}^* \geq \epsilon_1\} \cup \{\nu_{n+1}^* \geq \epsilon_2\}]] .$$

Proof By (55), as $t_{n_1+1} \leq T+1$, $\mathcal{E}^c(T) \subseteq \mathcal{E}_{\text{after}}$. For any two events $\mathcal{E}_1$ and $\mathcal{E}_2$, as

$$\mathcal{E}_1 = [\mathcal{E}_2 \cap \mathcal{E}_1] \cup [\mathcal{E}_2^c \cap \mathcal{E}_1] \subseteq \mathcal{E}_2 \cup [\mathcal{E}_2^c \cap \mathcal{E}_1],$$

we have $\mathcal{E}_{\text{after}} \subseteq \mathcal{E}_{\text{mid}} \cup [\mathcal{E}_{\text{mid}}^c \cap \mathcal{E}_{\text{after}}]$. Using Remark 10 and since (11) holds,

$$\left\{\left[\sup_{n_0 \leq k < n} \rho_{k+1}\right] \leq R_1^{\text{gap}}\right\} \cap \left\{\left[\sup_{n_0 \leq k < n} \nu_{k+1}\right] \leq R_2^{\text{gap}}\right\} \subseteq G_n.$$

for all $n \geq n_0$. Hence by simple manipulations, we have

$$\mathcal{E}_{\text{mid}} \subseteq \bigcup_{n=n_0}^{n_1} \{G_n \cap [\{\rho_{n+1} \geq R_1^{\text{gap}}\} \cup \{\nu_{n+1} \geq R_2^{\text{gap}}\}]\}.$$

Arguing similarly, one can see that

$$\begin{aligned} & \mathcal{E}_{\text{mid}}^c \cap \mathcal{E}_{\text{after}} \\ \subseteq \quad & G_{n_1+1} \cap \mathcal{E}_{\text{after}} \\ \subseteq \quad & \bigcup_{n>n_1} [G_n \cap [\{\rho_{n+1}^* \geq \epsilon_1\} \cup \{\nu_{n+1}^* \geq \epsilon_2\}]] , \end{aligned}$$

where the last inequality follows as $\epsilon_1 \leq R_1^{\text{out}}$ and $\epsilon_2 \leq R_2^{\text{out}}$. The desired result now follows. ∎

D.3. Technical Lemmas for Subsection D.4

We now provide two technical lemmas that will be used in the proofs of Lemmas 18 and 20.

Lemma 15 *Let $0 < r_0 < r_1 < \cdots < r_\ell$, let $\gamma_i = r_{i+1} - r_i$ for $i = 0, \ldots, \ell-1$, let U be some $d \times d$ matrix, and let $\rho : \mathbb{R} \to \mathbb{R}$ be some mapping. Assume that for some constant J it holds that $\|\rho(\sigma)\| \leq \gamma_i J$ for any $\sigma \in [r_i, r_{i+1}]$ and $i = 0, \ldots, \ell-1$. Assume, furthermore that for some constants $K > 0$ and $q_0 > 0$ it holds that $\|e^{-U(r-r_0)}\| \leq K e^{-q_0(r-r_0)}$ for any $r > r_0$. Then*

$$\left\| \int_{r_0}^{r_\ell} e^{-U(r_\ell - \sigma)} \rho(\sigma) d\sigma \right\| \leq \frac{KJ}{q_0} \left[\sup_{i=0,\ldots,\ell-1} \gamma_i \right].$$

Proof The claim of the lemma follows easily as, due to the assumptions,

$$\begin{aligned}\left\|\int_{r_0}^{r_\ell} e^{-U(r_\ell-\sigma)}\rho(\sigma)d\sigma\right\| &\leq \sum_{i=0}^{\ell-1}\int_{r_i}^{r_{i+1}}\left\|e^{-U(r_\ell-\sigma)}\right\|\|\rho(\sigma)\|\,\mathrm{d}\sigma\\ &\leq KJ\sum_{i=0}^{\ell-1}\gamma_i\int_{r_i}^{r_{i+1}} e^{-q_0(r_\ell-\sigma)}\mathrm{d}\sigma\\ &\leq KJ\left[\sup_{i=0,\ldots,\ell-1}\gamma_i\right]\int_{r_0}^{r_\ell} e^{-q_0(r_\ell-\sigma)}\mathrm{d}\sigma\\ &\leq \frac{KJ}{q_0}\left[\sup_{i=0,\ldots,\ell-1}\gamma_i\right],\end{aligned}$$

where, to get the last relation, we have used the fact that $\int_{r_0}^{r_\ell} e^{-q_0(r_\ell-\sigma)}\mathrm{d}\sigma \leq 1$. ∎

Lemma 16 (Dominating Decay Rate Bound) *Fix $q \in (0, q_{\min})$ where $q_{\min} := \min\{q_1, q_2\}$. Then for $n \geq n_0$,*

$$\sum_{k=n_0}^{n-1}\int_{t_k}^{t_{k+1}} e^{-q_1(t_n-\tau)}e^{-q_2(\xi(\tau)-s_{n_0})}\mathrm{d}\tau \leq \frac{1}{(q_{\min}-q)e}e^{-q(t_n-t_{n_0})}.$$

Proof From (4), $\beta_k \geq \alpha_k$ $\forall k \geq 1$. Using this and (21), $\forall k \geq 1$ and $\tau \in [t_k, t_{k+1}]$, $\xi(\tau) - s_k \geq \tau - t_k$. Hence for any $\tau \in [t_{n_0}, t_n]$,

$$-q_1(t_n-\tau) - q_2(\xi(\tau)-s_{n_0}) \leq -q_{\min}(t_n-t_{n_0}).$$

Now, since $\frac{1}{\alpha e}$ is the maximum of $xe^{-\alpha x}$,

$$\begin{aligned}(t_n-t_{n_0})e^{-q_{\min}(t_n-t_{n_0})} &= (t_n-t_{n_0})e^{-(q_{\min}-q)(t_n-t_{n_0})}e^{-q(t_n-t_{n_0})}\\ &\leq \frac{1}{(q_{\min}-q)e}e^{-q(t_n-t_{n_0})}.\end{aligned}$$

The desired result now follows. ∎

D.4. Bounding the Error Terms Discussed in Subsection 4.3

For obtaining the bounds in this subsection, we first show worst-case bounds on the increments. For $k \geq n_0$, let

$$I^\theta(k) := \|\theta_{k+1}-\theta_k\|/\alpha_k \tag{61}$$

and

$$I^z(k) := \|z_{k+1}-z_k\|/\beta_k. \tag{62}$$

Also, let

$$R^* := \left\|X_1^{-1}\right\|\|b_1\| \tag{63}$$

so that

$$\|\theta^*\| \le R^*. \tag{64}$$

On G_n, for $k \in \{n_0, \ldots, n\}$,

$$\begin{aligned}\|w_k\| &\le \|z_k\| + \|\lambda(\theta^*)\| + \|\lambda(\theta_k) - \lambda(\theta^*)\| \\ &\le R_2^{\text{out}} + \left\|W_2^{-1}\right\| \left[\, \|v_2\| + \|\Gamma_2\| \left[R^* + R_1^{\text{out}}\right]\right] \\ &=: R_2^w. \end{aligned} \tag{65}$$

Lemma 17 (Bounded Differences) *Fix $n_0 \ge 0$ and $n \ge n_0$. Then on G_n, assuming (11),*

$$\sup_{n_0 \le k \le n} I^\theta(k) \le J^\theta, \qquad \sup_{n_0 \le k \le n} I^z(k) \le J^z. \tag{66}$$

where

$$J^\theta = \|v_1\| + \|\Gamma_1\| \left[R^* + R_1^{out}\right] + \|W_1\| \, R_2^w + m_1[1 + R^* + R_1^{out} + R_2^w]$$

and

$$J^z := \|W_2\| \, R_2^{out} + \left\|W_2^{-1}\right\| \|\Gamma_2\| \, J^\theta + m_2(1 + R^* + R_1^{out} + R_2^w).$$

Proof Fix $k \in \{n_0, \ldots, n\}$. On G_n, using (1), $\mathcal{A}_3$, (64), (60), and (65), in that order,

$$\begin{aligned} I^\theta(k) &\le \|v_1 - \Gamma_1\theta_k - W_1 w_k\| + \left\|M_{k+1}^{(1)}\right\| \\ &\le \|v_1\| + \|\Gamma_1\| \left(\|\theta^*\| + \|\theta_k - \theta^*\|\right) \\ &\quad + \|W_1\| \, \|w_k\| \\ &\quad + m_1[1 + \|\theta^*\| + \|\theta_k - \theta^*\| + \|w_k\|] \\ &\le J^\theta. \end{aligned} \tag{67}$$

Similarly, on G_n, using (8), $\mathcal{A}_3$, (60), (4) from $\mathcal{A}_2$, (67), (64), and (65), in that order,

$$\begin{aligned} I^z(k) &\le \|W_2\| \, \|z_k\| + \|\lambda(\theta_k) - \lambda(\theta_{k+1})\| / \beta_k \\ &\quad + \left\|M_{k+1}^{(2)}\right\| \\ &\le \|W_2\| \, \|z_k\| + \left\|[W_2]^{-1}\right\| \|\Gamma_2\| \, \eta_k I^\theta(k) \\ &\quad + m_2(1 + \|\theta^*\| + \|\theta_k - \theta^*\| + \|w_k\|) \\ &\le J^z. \end{aligned}$$

Since k was arbitrary the result follows. ■

Let $q^{(1)}(W_2), \ldots, q^{(d)}(W_2)$ be the eigenvalues of W_2. Fix $q_2 \in (0, q_2')$, where

$$q_2' := \min_i \{\text{real}(q^{(i)}(W_2))\}.$$

Then from Corollary 3.6 (Teschl, 2004), there exists $K_2 \ge 1$ so that

$$\left\| e^{-W_2(s-\mu)} \right\| \le K_2 e^{-q_2(s-\mu)}, \; \forall s \ge \mu. \tag{68}$$

For the rest of the results in this subsection, we consider intermediate intervals $[s_n, s_{n+1}]$. The next lemma gives bounds on the three error terms of the interpolated trajectory $\bar{z}(s)$ at the extremes $\{s_n, s_{n+1}\}$. This suffices for bounding the deviation of $\bar{z}(s)$ from $z(s)$ on the whole interval, as is shown in the subsequent lemma.

Lemma 18 (Perturbation Error Bounds for z_n) *Fix $n_0 \geq 0$ and $n \geq n_0$ Then on G_n, assuming* (11)*,*

$$\sup_{\ell \in \{n,n+1\}} \left\|E_2^{de}(s_\ell)\right\| \leq L_2^{de} \left[\sup_{k \geq 0} \beta_k\right],$$

$$\sup_{\ell \in \{n,n+1\}} \left\|E_2^{sd}(s_\ell)\right\| \leq L_2^{sd} \left[\sup_{k \geq 0} \eta_k\right],$$

$$\left\|E_2^{md}(s_{n+1})\right\| \leq K_2 \left\|E_2^{md}(s_n)\right\| + L_2^{md}\beta_n.$$

where $L_2^{de} := \frac{K_2 J^z \|W_2\|}{q_2}$, $L_2^{sd} := \frac{K_2\left\|W_2^{-1}\right\|\|\Gamma_2\| J^\theta}{q_2}$, $L_2^{md} := K_2 m_2[1 + R^* + R_1^{out} + R_2^w]$.

Proof Fix $\ell \in \{n, n+1\}$.

For the first claim note that, by Lemma 17, on G_n,

$$\left\|\chi^{\mathrm{de}}(\mu)\right\| \leq \|W_2\| (\mu - s_k) I^z(k) \leq \|W_2\| \beta_k J^z$$

for $\mu \in [s_k, s_{k+1})$, where $I^z(k)$ is as in (62). The claim then follows easily by recalling (68), and applying Lemma 15 with $r_i = s_i$, $\gamma_i = \beta_i$, $U = W_2$, $\rho = \chi^{\mathrm{de}}$, $K = K_2$, $q_0 = -q_2$ and $J = \|W_2\| J^z$.

For the second claim, let $k \in \{n_0, \ldots, \ell - 1\}$ and $\mu \in [s_k, s_{k+1})$. With $I^\theta(k)$ as in (61),

$$\left\|\chi^{\mathrm{sd}}(\mu)\right\| \leq \eta_k \left\|W_2^{-1}\right\| \|\Gamma_2\| I^\theta(k).$$

Hence by Lemma 17, on G_n,

$$\left\|\chi^{\mathrm{sd}}(\mu)\right\| \leq \eta_k \left\|W_2^{-1}\right\| \|\Gamma_2\| J^\theta.$$

The claim then follows again by (68) and Lemma 15.

For the third claim, by its definition and the triangle inequality,

$$\begin{aligned}
&\left\|E_2^{\mathrm{md}}(s_{n+1})\right\| \\
= \; & \left\|\int_{s_{n_0}}^{s_{n+1}} e^{-W_2(s_{n+1}-\mu)} \chi^{\mathrm{md}}(\mu) \mathrm{d}\mu\right\| \\
\leq \; & \left\|e^{-W_2\beta_n} \int_{s_{n_0}}^{s_n} e^{-W_2(s_n-\mu)} \chi^{\mathrm{md}}(\mu) \mathrm{d}\mu\right\| + \left\|\int_{s_n}^{s_{n+1}} e^{-W_2(s_{n+1}-\mu)} \chi^{\mathrm{md}}(\mu) \mathrm{d}\mu\right\|.
\end{aligned}$$

Applying (68) on both terms, we get that

$$\left\|E_2^{\mathrm{md}}(s_{n+1})\right\| \leq K_2 \left\|E_2^{\mathrm{md}}(s_n)\right\| + K_2\beta_n \left\|M_{n+1}^{(2)}\right\|.$$

On G_n, using $\boldsymbol{\mathcal{A}_3}$ with (60), (64), and (65), we have $K_2 \left\|M_{n+1}^{(2)}\right\| \leq L_2^{\mathrm{md}}$. The third claim is now easy to see. ■

The next lemma shows that for $\tau \in [s_n, s_{n+1}]$, $\bar{z}(\tau)$ cannot deviate much from the ODE trajectory $z(\tau)$ if the stepsizes are small enough. In particular, it bounds the distance with decaying terms using Lemma 18.

Lemma 19 (ODE-SA Distance Bound for z_n) *Fix $n_0 \geq 0$ and $n \geq n_0$. Then on G_n and since* (11) *holds,*

$$\nu_{n+1} \leq K_2 \left\|E_2^{md}(s_n)\right\| + L^z \max\left\{\sup_{k\geq n_0} \beta_k, \sup_{k\geq n_0} \eta_k\right\},$$

$$\nu^*_{n+1} \leq K_2 \left\|E_2^{md}(s_n)\right\| + K_2 R_2^{in} e^{-q_2(s_n - s_{n_0})} + L^z \max\left\{\sup_{k\geq n_0} \beta_k, \sup_{k\geq n_0} \eta_k\right\},$$

where $L^z = L_2^{de} + L_2^{md} + \|W_2\| R_2^{in} + L_2^{sd}$.

Proof Let $\mu \in [s_n, s_{n+1}]$. Then there exists $\kappa \in [0,1]$ so that

$$\bar{z}(\mu) = (1-\kappa)\bar{z}(s_n) + \kappa\bar{z}(s_{n+1}).$$

Hence

$$\|\bar{z}(\mu) - z(\mu, s_{n_0}, z_{n_0})\| \leq (1-\kappa)\|\bar{z}(s_n) - z(\mu, s_{n_0}, z_{n_0})\| + \kappa\|\bar{z}(s_{n+1}) - z(\mu, s_{n_0}, z_{n_0})\|.$$

Using (9),

$$z(\mu, s_{n_0}, z_{n_0}) = z(s_n, s_{n_0}, z_{n_0}) + \int_{s_n}^{\mu} [-W_2\, z(\mu_1, s_{n_0}, z_{n_0})]\mathrm{d}\mu_1,$$

and

$$z(s_{n+1}, s_{n_0}, z_{n_0}) = z(\mu, s_{n_0}, z_{n_0}) + \int_{\mu}^{s_{n+1}} [-W_2\, z(\mu_1, s_{n_0}, z_{n_0})]\mathrm{d}\mu_1.$$

Combining the above three relations, we have

$$\begin{aligned}\|\bar{z}(\mu) - z(\mu, s_{n_0}, z_{n_0})\| \leq{}& (1-\kappa)\|\bar{z}(s_n) - z(s_n, s_{n_0}, z_{n_0})\| \\ &+ \kappa\|\bar{z}(s_{n+1}) - z(s_{n+1}, s_{n_0}, z_{n_0})\| + \int_{s_n}^{s_{n+1}} \|W_2\|\,\|z(\mu_1, s_{n_0}, z_{n_0})\|\,\mathrm{d}\mu_1.\end{aligned}$$

Since (11) holds, as $\|z_{n_0}\| \leq R_2^{\text{in}}$, from Remark 10, $\|z(\mu, s_{n_0}, z_{n_0})\| \leq R_2^{\text{in}}$ for all $s \geq s_{n_0}$. Using this with (28), (68), the facts that $K_2 \geq 1$ and $\beta_n \leq [\sup_{k\geq n_0}\beta_k]$, and Lemma 18, the first claim follows:

$$\begin{aligned}\nu_{n+1} \leq{}& L_2^{\text{de}}\left[\sup_{k\geq n_0}\beta_k\right] + L_2^{\text{sd}}\left[\sup_{k\geq n_0}\eta_k\right] + \kappa L_2^{\text{md}}\beta_n \\ &+ ((1-\kappa) + \kappa K_2)\left\|E_2^{\text{md}}(s_n)\right\| + \|W_2\|\,\beta_n R_2^{\text{in}}. \\ \leq{}& K_2\left\|E_2^{\text{md}}(s_n)\right\| + L^z \max\left\{\sup_{k\geq n_0}\beta_k, \sup_{k\geq n_0}\eta_k\right\}.\end{aligned} \tag{69}$$

For the second claim observe that

$$\|\bar{z}(\mu)\| \leq \|\bar{z}(\mu) - z(\mu, s_{n_0}, z_{n_0})\| + \|z(\mu, s_{n_0}, z_{n_0})\|.$$

Hence

$$\nu^*_{n+1} \leq \nu_{n+1} + \sup_{\mu\in[s_n, s_{n+1}]} \|z(\mu, s_{n_0}, z_{n_0})\|.$$

Lastly, since (11) holds, $\|z_{n_0}\| \leq R_2^{\text{in}}$, and hence using (23) and (68),

$$\|z(\mu, s_{n_0}, z_{n_0})\| \leq K_2 R_2^{\text{in}} e^{-q_2(\mu - s_{n_0})}.$$

Combining the above two relations with (69), the desired result is now easy to see. ■

We now reproduce the results of Lemma 19, this time for $\{\theta_n\}$ instead of $\{z_n\}$, and obtain bounds on ρ_{n+1} and ρ_{n+1}^* on G_n, assuming (11). To do so, it suffices to bound $\left\|E_1^{\text{de}}(\cdot)\right\|$, $\left\|E_1^{\text{md}}(\cdot)\right\|$, and $\|E_1^{\text{te}}(\cdot)\|$ on the interval $[t_n, t_{n+1}]$.

Similarly as in (68), there exist q_1 and $K_1 \geq 1$ so that

$$\left\|e^{-X_1(t-\tau)}\right\| \leq K_1 e^{-q_1(t-\tau)}, \ \forall t \geq \tau. \tag{70}$$

Fix

$$q \in (0, q_{\min}), \qquad q_{\min} := \min\{q_1, q_2\}, \tag{71}$$

where q_2 is from (68). The next lemma gives bounds on the three components of $E_1(t)$.

Lemma 20 (Perturbation Error Bounds for θ_n) *Fix $n_0 \geq 0$ and $n \geq n_0$. Then on G_n, assuming (11),*

$$\sup_{\ell \in \{n, n+1\}} \left\|E_1^{de}(t_\ell)\right\| \leq L_1^{de}\left[\sup_{k \geq n_0} \alpha_k\right],$$

$$\sup_{\ell \in \{n, n+1\}} \left\|E_1^{te}(t_\ell)\right\| \leq L_{1a}^{te}\, e^{-q(t_n - t_{n_0})} + L_{1b}^{te}\left[\sup_{k \geq n_0} \beta_k\right] + L_{1c}^{te}\left[\sup_{n_0 \leq k \leq n} \nu_{k+1}\right],$$

$$\left\|E_1^{md}(t_{n+1})\right\| \leq K_1\left\|E_1^{md}(t_n)\right\| + L_1^{md}\alpha_n,$$

where $L_1^{de} := \frac{K_1 J^\theta \|X_1\|}{q_1}$, $L_{1a}^{te} := K_1\|W_1\| K_2 R_2^{in} \frac{1}{(q_{\min} - q)e}$, $L_{1b}^{te} := K_1\|W_1\|\|W_2\| R_2^{in}/q_1$, $L_{1c}^{te} := K_1\|W_1\|/q_1$, $L_1^{md} := K_1 m_1[1 + R^* + R_1^{out} + R_2^{w}]$.

Proof For the first claim of the lemma fix $\ell \in \{n, n+1\}$. Let $k \in \{n_0, \ldots, \ell - 1\}$ and $\tau \in [t_k, t_{k+1})$. With $I^\theta(k)$ as in (61),

$$\left\|\zeta^{\text{de}}(\tau)\right\| \leq \|X_1\|(\tau - t_k) I^\theta(k) \leq \alpha_k \|X_1\| I^\theta(k).$$

So by Lemma 17, on G_n, $\left\|\zeta^{\text{de}}(\tau)\right\| \leq \alpha_k \|X_1\| J^\theta$. The first claim now follows by (70) and Lemma 15.

For proving the second claim of the lemma let $\ell = n$. By triangle inequality,

$$\|E_1^{\text{te}}(t_n)\| \leq \sum_{k=n_0}^{n-1} \int_{t_k}^{t_{k+1}} \left\|e^{-X_1(t_n - \tau)}\right\| \left\|\zeta^{\text{te}}(\tau)\right\| \mathrm{d}\tau.$$

Using (70), it follows that

$$\|E_1^{\text{te}}(t_n)\| \leq K_1 \sum_{k=n_0}^{n-1} \int_{t_k}^{t_{k+1}} e^{-q_1(t_n - \tau)} \left\|\zeta^{\text{te}}(\tau)\right\| \mathrm{d}\tau.$$

Fix $k \in \{n_0, \ldots, n-1\}$ and $\tau \in [t_k, t_{k+1})$. Then

$$\left\|\zeta^{\text{te}}(\tau)\right\| \leq \|W_1\| \, \|z_k\| \, .$$

Using (21) and the triangle inequality,

$$\begin{aligned}\|z_k\| \leq \|z(\xi(\tau), s_{n_0}, z_{n_0})\| \\ + \|z(\xi(\tau), s_{n_0}, z_{n_0}) - z(\xi(t_k), s_{n_0}, z_{n_0})\| + \|z_k - z(\xi(t_k), s_{n_0}, z_{n_0})\| \, .\end{aligned}$$

Since (11) holds, $\|z_{n_0}\| \leq R_2^{\text{in}}$; thus by (23) and (68),

$$\|z(\xi(\tau), s_{n_0}, z_{n_0})\| \leq K_2 R_2^{\text{in}} e^{-q_2(\xi(\tau)-s_{n_0})}.$$

Remark 10 also implies that, as $\|z_{n_0}\| \leq R_2^{\text{in}}$, $\|z(s, s_{n_0}, z_{n_0})\| \leq R_2^{\text{in}}$ for all $s \geq s_{n_0}$. Hence by (9),

$$\begin{aligned}\|z(\xi(\tau), s_{n_0}, z_{n_0}) - z(\xi(t_k), s_{n_0}, z_{n_0})\| &\leq \left\| \int_{\xi(t_k)}^{\xi(\tau)} [-W_2] \; z(\mu, s_{n_0}, z_{n_0}) \mathrm{d}\mu \right\| \\ &\leq \|W_2\| \, R_2^{\text{in}} \beta_k,\end{aligned}$$

where the last relation holds as $[\xi(\tau) - \xi(t_k)] \leq [s_{k+1} - s_k]$. Also note that, by (59),

$$\|z_k - z(\xi(t_k), s_{n_0}, z_{n_0})\| \leq \nu_{k+1}.$$

Combining the above relations,

$$\begin{aligned}&\left\|\zeta^{\text{te}}(\tau)\right\| \\ \leq\;& \|W_1\| \left[K_2 R_2^{\text{in}} e^{-q_2(\xi(\tau)-s_{n_0})} + \|W_2\| \, R_2^{\text{in}} \beta_k + \nu_{k+1} \right] \\ \leq\;& \|W_1\| \left[K_2 R_2^{\text{in}} e^{-q_2(\xi(\tau)-s_{n_0})} + \|W_2\| \, R_2^{\text{in}} \left[\sup_{k \geq n_0} \beta_k \right] + \left[\sup_{n_0 \leq k \leq n-1} \nu_{k+1} \right] \right]\end{aligned}$$

By Lemma 16 and the fact that $\int_{t_{n_0}}^{t_n} e^{-q_1(t_n - \tau)} \mathrm{d}\tau \leq 1/q_1$,

$$\|E_1^{\text{te}}(t_n)\| \leq L_{1a}^{\text{te}} e^{-q(t_n - t_{n_0})} + L_{1b}^{\text{te}} \left[\sup_{k \geq n_0} \beta_k \right] + L_{1c}^{\text{te}} \left[\sup_{n_0 \leq k \leq n-1} \nu_{k+1} \right].$$

A similar bound holds for $\ell = n+1$. Since $e^{-q(t_{n+1}-t_{n_0})} \leq e^{-q(t_n - t_{n_0})}$, the second claim of the lemma follows.

The third claim of the lemma, bounding $\left\|E_2^{\text{md}}(s_{n+1})\right\|$, follows in a similar way to the third claim of Lemma 18. ■

Similarly to Lemma 19, the next lemma bounds ρ_{n+1} and ρ_{n+1}^* with decaying terms using Lemma 20.

Lemma 21 (ODE-SA Distance Bound for θ_n) *Fix $n_0 \geq 0$ and $n \geq n_0$. Then on G_n, assuming (11),*

$$\rho_{n+1} \leq K_1 \left\|E_1^{md}(t_n)\right\| + L_a^\theta\, e^{-q(t_n - t_{n_0})} + L_b^\theta \left[\sup_{k \geq n_0} \beta_k\right] + L_c^\theta \left[\sup_{n_0 \leq k \leq n} \nu_{k+1}\right],$$

$$\rho_{n+1}^* \leq K_1 \left\|E_1^{md}(t_n)\right\| + [K_1 R_1^{in} + L_a^\theta] e^{-q(t_n - t_{n_0})} + L_b^\theta \left[\sup_{k \geq n_0} \beta_k\right] + L_c^\theta \left[\sup_{n_0 \leq k \leq n} \nu_{k+1}\right],$$

where $L_a^\theta = L_{1a}^{te}, L_c^\theta = L_{1c}^{te}$ and $L_b^\theta := L_1^{de} + L_1^{md} + \|X_1\|\, R_1^{in} + L_{1b}^{te}$.

Proof Let $\tau \in [t_n, t_{n+1}]$. Then arguing as in the proof of Lemma 19 and using (6), there exists $\kappa \in [0, 1]$ such that

$$\begin{aligned}\left\|\bar{\theta}(\tau) - \theta(\tau, t_{n_0}, \theta_{n_0})\right\| \leq{}& (1-\kappa) \left\|\bar{\theta}(t_n) - \theta(t_n, t_{n_0}, \theta_{n_0})\right\| \\ &+ \kappa \left\|\bar{\theta}(t_{n+1}) - \theta(t_{n+1}, t_{n_0}, \theta_{n_0})\right\| + \int_{t_n}^{t_{n+1}} \|X_1\| \left\|\theta(\tau', t_{n_0}, \theta_{n_0}) - \theta^*\right\| \mathrm{d}\tau'.\end{aligned}$$

Due to (11), $\left\|\bar{\theta}(t_{n_0}) - \theta^*\right\| \leq R_1^{\text{in}}$; thus, from Remark 10, $\|\theta(\tau, t_{n_0}, \theta_{n_0}) - \theta^*\| \leq R_1^{\text{in}}$ for all $t \geq t_{n_0}$. Using this with (25) and (68), the facts that $K_1 \geq 1$,

$$\alpha_n \leq \left[\sup_{k \geq n_0} \alpha_k\right] \leq \left[\sup_{k \geq n_0} \beta_k\right],$$

and Lemma 20, the first claim of the lemma follows:

$$\begin{aligned}\rho_{n+1} \leq{}& L_1^{\text{de}} \left[\sup_{k \geq n_0} \beta_k\right] + L_{1a}^{\text{te}} e^{-q(t_n - t_{n_0})} + L_{1b}^{\text{te}} \left[\sup_{k \geq n_0} \beta_k\right] + L_{1c}^{\text{te}} \left[\sup_{n_0 \leq k \leq n} \nu_{k+1}\right] \\ &+ \kappa L_1^{\text{md}} \left[\sup_{k \geq n_0} \beta_k\right] + (\kappa + (1-\kappa) K_1) \left\|E_1^{\text{md}}(t_n)\right\| + \|X_1\|\, R_1^{\text{in}} \left[\sup_{k \geq n_0} \beta_k\right] \\ \leq{}& K_1 \left\|E_1^{\text{md}}(t_n)\right\| + L_a^\theta\, e^{-q(t_n - t_{n_0})} + L_b^\theta \left[\sup_{k \geq n_0} \beta_k\right] + L_c^\theta \left[\sup_{n_0 \leq k \leq n} \nu_{k+1}\right].\end{aligned} \tag{72}$$

For the second claim of the lemma, notice that

$$\left\|\bar{\theta}(\tau) - \theta^*\right\| \leq \left\|\bar{\theta}(\tau) - \theta(\tau, t_{n_0}, \theta_{n_0})\right\| + \|\theta(\tau, t_{n_0}, \theta_{n_0}) - \theta^*\|.$$

Thus, we have

$$\rho_{n+1}^* \leq \rho_{n+1} + \sup_{\tau \in [t_n, t_{n+1}]} \|\theta(\tau, t_{n_0}, \theta_{n_0}) - \theta^*\|.$$

Lastly, using (11), $\left\|\bar{\theta}(t_{n_0}) - \theta^*\right\| \leq R_1^{\text{in}}$; thus, from (22),

$$\|\theta(\tau, t_{n_0}, \theta_{n_0}) - \theta^*\| \leq K_1 R_1^{\text{in}} e^{-q_1(\tau - t_{n_0})}.$$

Combining the above two relations, using (72) and the fact that $q < q_1$, the second claim of the lemma follows. ∎

D.5. Completing the Proof of Theorem 4

We first prove Lemmas 22 and 23 for bounding the terms appearing in Lemma 14 using the results from the previous subsections. Then, we provide a bound on the martingale difference noise in Lemma 24. Finally, we combine these results to prove Theorem 4.

Lemma 22 *In accordance with Table 2, let $N_{0,a} \equiv N_{0,a}(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$ denote the smallest positive value satisfying*

$$\max\left\{\sup_{k\geq N_{0,a}} \beta_k, \sup_{k\geq N_{0,a}} \eta_k\right\} \leq \frac{\min\{\epsilon_1/8,\ \epsilon_2/3\}}{L^z \max\{L_c^\theta, 1\}}, \tag{73}$$

$N_{0,b} \equiv N_{0,b}(\epsilon_1, \{\beta_k\})$ the smallest positive value satisfying

$$\sup_{k\geq N_{0,b}} \beta_k \leq \frac{\epsilon_1}{4L_b^\theta}, \tag{74}$$

and $N_0 \equiv N_0(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\}) = \max\{N_{0,a}, N_{0,b}\}$. Then, for any $n_0 \geq N_0$ and $n \geq n_0$, assuming (11),

$$[G_n \cap \{\nu_{n+1} \geq R_2^{gap}\}] \subseteq \left[G_n \cap \left\{K_2 \left\|E_2^{md}(s_n)\right\| \geq \frac{\epsilon_2}{3}\right\}\right] \tag{75}$$

and

$$\begin{aligned}&[G_n \cap \{\rho_{n+1} \geq R_1^{gap}\}] \\ &\qquad \subseteq \left[G_n \cap \left\{K_1 \left\|E_1^{md}(t_n)\right\| \geq \frac{\epsilon_1}{4}\right\}\right] \cup \bigcup_{k=n_0}^{n} \left[G_k \cap \left\{L_c^\theta K_2 \left\|E_2^{md}(s_k)\right\| \geq \frac{\epsilon_1}{8}\right\}\right].\end{aligned} \tag{76}$$

Proof Equation (75) follows from Lemma 19, (57), and the fact that

$$2\epsilon_2/3 \geq \epsilon_2/3 \geq L^z \max\left\{\sup_{k\geq n_0} \beta_k, \sup_{k\geq n_0} \eta_k\right\}$$

for $n_0 \geq N_{0,a}$.

We now prove (76). Due to (56) and (57), $R_1^{\text{gap}} = 4L_a^\theta$ (see Table 2 for the definition of L_a^θ), and thus $L_a^\theta e^{-q(t_n - t_{n_0})} \leq R_1^{\text{gap}}/4$ for $n \geq n_0$. Additionally, as $n_0 \geq N_{0,b}$, $L_b^\theta \left[\sup_{k\geq n_0} \beta_k\right] \leq \epsilon_1/4$. Consequently, by Lemma 21, and as $R_1^{\text{gap}} \geq \epsilon_1$ due to (57),

$$\begin{aligned}&[G_n \cap \{\rho_{n+1} \geq R_1^{\text{gap}}\}] \\ &\qquad \subseteq \left[G_n \cap \left\{K_1 \left\|E_1^{\text{md}}(t_n)\right\| \geq \frac{\epsilon_1}{4}\right\}\right] \cup \left[G_n \cap \left\{L_c^\theta \left[\sup_{n_0\leq k\leq n} \nu_{k+1}\right] \geq \frac{\epsilon_1}{4}\right\}\right].\end{aligned}$$

Noting also that $G_n \subseteq G_k$ for all $n_0 \leq k \leq n$, the desired result now follows from Lemma 19, and the fact that $\epsilon_1/8 \geq L^z \max\left\{\sup_{k\geq n_0} \beta_k, \sup_{k\geq n_0} \eta_k\right\}$ for $n_0 \geq N_{0,a}$ by the definition of $N_{0,a}$. ■

Lemma 23 *Fix some $n_0 \geq N_0$ and $n_1 \geq N_1$, where, in accordance with Table 2, $N_0 \equiv N_0(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$ is defined as in Lemma 22, $N_1 \equiv N_1(n_0, \epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\}) = \max\{N_{1,a}, N_{1,b}\}$, $N_{1,a} \equiv N_{1,a}(n_0, \epsilon_1, \{\alpha_k\})$ denotes the smallest positive value satisfying*

$$[K_1 R_1^{in} + L_a^{\theta}]e^{-q(t_{N_{1,a}} - t_{n_0})} \leq \frac{\epsilon_1}{4}, \tag{77}$$

and $N_{1,b} \equiv N_{1,b}(n_0, \epsilon_2, \{\beta_k\})$ denotes the smallest positive value satisfying

$$K_2 R_2^{in} e^{-q_2(s_{N_{1,b}} - s_{n_0})} \leq \frac{\epsilon_2}{3}. \tag{78}$$

Then, assuming (11), for all $n \geq n_1$,

$$[G_n \cap \{\nu_{n+1}^* \geq \epsilon_2\}] \subseteq \left[G_n \cap \left\{K_2 \left\|E_2^{md}(s_n)\right\| \geq \frac{\epsilon_2}{3}\right\}\right] \tag{79}$$

and

$$\begin{aligned} &[G_n \cap \{\rho_{n+1}^* \geq \epsilon_1\}] \\ &\quad \subseteq \left[G_n \cap \left\{K_1 \left\|E_1^{md}(t_n)\right\| \geq \frac{\epsilon_1}{4}\right\}\right] \cup \bigcup_{k=n_0}^{n} \left[G_k \cap \left\{L_c^{\theta} K_2 \left\|E_2^{md}(s_k)\right\| \geq \frac{\epsilon_1}{8}\right\}\right]. \end{aligned} \tag{80}$$

Proof Note that $K_2 R_2^{\text{in}} e^{-q_2(s_n - s_{n_0})} \leq \epsilon_2/3$ for all $n \geq N_{1,b}$ due to $q \leq q_2$, and that

$$L^z \max\left\{\sup_{k \geq n} \beta_k, \sup_{k \geq n} \eta_k\right\} \leq \epsilon_1/3$$

for all $n \geq N_{0,a}$ (recall $N_{0,a}$ from Lemma 22) . Therefore, due to Lemma 19, (79) holds.

For proving (80), note first that, as $n \geq N_{1,a}$ and $q \leq q_2$, it holds that

$$[K_1 R_1^{\text{in}} + L_a^{\theta}]e^{-q_2(t_{N_{1,a}} - t_{n_0})} \leq \frac{\epsilon_1}{4}.$$

Additionally, as $n \geq N_{0,b}$, $L_b^{\theta}\left[\sup_{k \geq n_0} \beta_k\right] \leq \epsilon_1/4$ (recall $N_{0,b}$ from Lemma 22). Consequently, by Lemma 21,

$$\begin{aligned} &[G_n \cap \{\rho_{n+1}^* \geq \epsilon_1\}] \\ &\quad \subseteq \left[G_n \cap \left\{K_1 \left\|E_1^{\text{md}}(t_n)\right\| \geq \frac{\epsilon_1}{4}\right\}\right] \cup \left[G_n \cap \left\{L_c^{\theta}\left[\sup_{n_0 \leq k \leq n} \nu_{k+1}\right] \geq \frac{\epsilon_1}{4}\right\}\right]. \end{aligned} \tag{81}$$

To complete the proof, we argue as in the last part of the proof of Lemma 22: noting that $G_n \subseteq G_k$ for all $n_0 \leq k \leq n$, the desired result follows from (81) using Lemma 19, and the fact that $\epsilon_1/8 \geq L^z \max\left\{\sup_{k \geq n_0} \beta_k, \sup_{k \geq n_0} \eta_k\right\}$ for $n_0 \geq N_{0,a}$ (recall, again, $N_{0,a}$ from Lemma 22). ■

Lastly, to provide the proof of our main technical theorem, we give the following lemma. We remind the reader that $a_n = \sum_{k=0}^{n-1} \alpha_k^2 e^{-2q_1(t_n - t_{k+1})}$, and $b_n := \sum_{k=0}^{n-1} \beta_k^2 e^{-2q_2(s_n - s_{k+1})}$ for $n \geq 0$. Also recall that $E_1^{\text{md}}(t_n)$ and $E_2^{\text{md}}(t_n)$ depend on n_0, as can be seen from their definition in Subsection 4.2.

Lemma 24 (Azuma-Hoeffding for E_1^{md} and E_2^{md}) *Fix $n_0 \geq 0$, $\delta > 0$. Then for any $n \geq n_0$,*

$$\Pr\left\{G_n, \left\|E_1^{md}(t_n)\right\| \geq \delta\right\} \leq 2d^2 \exp\left(-\frac{\delta^2}{d^3 (L_1^{md})^2 a_n}\right) \tag{82}$$

and

$$\Pr\left\{G_n, \left\|E_2^{md}(s_n)\right\| \geq \delta\right\} \leq 2d^2 \exp\left(-\frac{\delta^2}{d^3 (L_2^{md})^2 b_n}\right). \tag{83}$$

Proof We only prove (82); (83) follows similarly.

Let $A_{k,n}$ be the matrix $\int_{t_k}^{t_{k+1}} e^{-X_1(t_n-\tau)} \mathrm{d}\tau$ with $A_{k,n}^{ij}$ denoting its $i,j-$th entry. Let $M_{k+1}^{(1)}(j)$ denote the $j-$th entry of $M_{k+1}^{(1)}$. On G_n, $1_{G_k} = 1$ for all $n_0 \leq k \leq n$. So

$$\begin{aligned}
\Pr\left\{G_n, \left\|E_1^{\mathrm{md}}(t_n)\right\| \geq \delta\right\} &= \Pr\left\{G_n, \left\|\sum_{k=n_0}^{n-1} A_{k,n} M_{k+1}^{(1)} 1_{G_k}\right\| \geq \delta\right\} \\
&\leq \Pr\left\{\left\|\sum_{k=n_0}^{n-1} A_{k,n} M_{k+1}^{(1)} 1_{G_k}\right\| \geq \delta\right\} \\
&\leq \sum_{i=1}^{d}\sum_{j=1}^{d} \Pr\left\{\left\|\sum_{k=n_0}^{n-1} A_{k,n}^{ij} M_{k+1}^{(1)}(j) 1_{G_k}\right\| \geq \frac{\delta}{d\sqrt{d}}\right\},
\end{aligned}$$

where the last relation is due to the union bound applied twice. On G_k, $K_1 \left\|M_{k+1}^{(1)}\right\| \leq L_1^{\mathrm{md}}$. Hence, on G_k, for any $i,j \in \{1,\ldots,d\}$, using (70),

$$|A_{k,n}^{ij}|\ |M_{k+1}^{(1)}(j)| \leq \|A_{k,n}\|\ \ \|M_{k+1}\| \leq K_1 L_1^{\mathrm{md}} \alpha_k e^{-q_1(t_n - t_{k+1})}.$$

Using $\sum_{k=n_0}^{n-1} \alpha_k^2 e^{-2q_1(t_n-t_{k+1})} \leq a_n$, the desired result now follows from the Azuma-Hoeffding inequality. ■

We finish with combining the above lemmas for proving our main technical result.

Proof of Theorem 4 Lemmas 14, 22, and 23 together show that, for any $n_0 \geq N_0(\epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$ and $n_1 \geq N_1(n_0, \epsilon_1, \epsilon_2, \{\alpha_k\}, \{\beta_k\})$,

$$\begin{aligned}
\mathcal{E}^c(n_0, T) \subseteq &\left[\bigcup_{n=n_0}^{\infty}\left[G_n \cap \left\{K_1 \left\|E_1^{\mathrm{md}}(t_n)\right\| \geq \frac{\epsilon_1}{4}\right\}\right]\right] \\
&\cup \left[\bigcup_{n=n_0}^{\infty}\left[G_n \cap \left\{K_2 \left\|E_2^{\mathrm{md}}(s_n)\right\| \geq \frac{\epsilon_2}{3}\right\}\right]\right] \cup \left[\bigcup_{n=n_0}^{\infty}\left[G_n \cap \left\{L_c^{\theta} K_2 \left\|E_2^{\mathrm{md}}(s_n)\right\| \geq \frac{\epsilon_1}{8}\right\}\right]\right].
\end{aligned}$$

The proof then follows from Lemma 24. ■

Appendix E. Proof of Theorem 6

Using Theorem 4, we are now ready to prove Theorem 6.

Proof of Theorem 6, Statement 1 First we claim that, under the choice of stepsize in the statement of the theorem, we have

$$n_0' \geq N_0(\epsilon, \epsilon, \{\alpha_k\}, \{\beta_k\}) \ , \tag{84}$$

where $N_0(\epsilon, \epsilon, \{\alpha_k\}, \{\beta_k\})$ is as in Lemma 22. The reason for this is that, due to our choice of n_0', (73) and (74) hold with

$$N_{0,a}(\epsilon, \epsilon, \{\alpha_k\}, \{\beta_k\}) = \left[8L^z \max\{L_c^\theta, 1\}/\epsilon\right]^{\frac{1}{\min\{\beta, \alpha-\beta\}}} ,$$
$$N_{0,b}(\epsilon, \epsilon, \{\beta_k\}) = \left[4L_b^\theta/\epsilon\right]^{1/\beta} .$$

Additionally, for any n_0, (78) holds with

$$N_{1,b}(n_0, \epsilon, \{\beta_k\}) = \left[(n_0+1)^{1-\beta} + \tfrac{1-\beta}{q_2} \ln\left[\tfrac{3K_2 R_2^{\text{in}}}{\epsilon}\right]\right]^{\frac{1}{1-\beta}} \quad .$$

This follows from the fact that

$$\begin{aligned} \sum_{k=n_0}^{N_{1,b}-1} (1+k)^{-\beta} \quad &\geq \quad \int_{n_0}^{N_{1,b}} (1+x)^{-\beta} dx \qquad (85) \\ &= \quad \frac{1}{1-\beta}\left[(N_{1,b}+1)^{(1-\beta)} - (N_{1,0}+1)^{(1-\beta)}\right] \quad . \qquad (86) \end{aligned}$$

Similarly, (77) holds with

$$N_{1,a}(n_0, \epsilon, \{\alpha_k\}) = \left[(n_0+1)^{1-\alpha} + \tfrac{1-\alpha}{q} \ln\left[\tfrac{4[K_1 R_1^{\text{in}} + L_a^\theta]}{\epsilon}\right]\right]^{\frac{1}{1-\alpha}} \quad .$$

For all $n_0 \geq 3$, we have $2n_0 \geq 1.5(n_0+1)$. Hence, if

$$n_0 \geq \max\left\{\left[\frac{1-\alpha}{((1.5)^{1-\alpha}-1)q} \ln \frac{4[K_1 R_1^{\text{in}} + L_a^\theta]}{\epsilon}\right]^{1/(1-\alpha)} , \ 3\right\},$$

then it is easy to see that $2n_0 \geq N_{1,a}(n_0, \epsilon, \{\beta_k\})$. Similarly, if

$$n_0 \geq \max\left\{\left[\frac{1-\beta}{((1.5)^{1-\beta}-1)q_2} \ln \frac{3K_2 R_2^{\text{in}}}{\epsilon}\right]^{1/(1-\beta)} , \ 3\right\},$$

then $2n_0 \geq N_{1,a}(n_0, \epsilon, \{\beta_k\})$. Thus, by our choice of n_0',

$$2n_0' \geq N_1(n_0', \epsilon, \epsilon, \{\alpha_k\}, \{\beta_k\}) \ , \tag{87}$$

where $N_1(n_0', \epsilon, \epsilon, \{\alpha_k\}, \{\beta_k\})$ is as in Lemma 23.

By (17), we have

$$\{w \in \mathbb{R}^d : \|w\| \leq R_2^{\text{in}}/2\} \subseteq \{w \in \mathbb{R}^d : \|w - \lambda(\theta)\| \leq R_2^{\text{in}} \ \ \forall \theta \text{ with } \|\theta\| \leq R_1^{\text{in}}/2\} \ .$$

Combining this with using (16), (7), and since n_0' is a power of 2, it follows from the definition of the projection operation that

$$\|\theta'_{n'_0} - \theta^*\| \leq R_1^{\text{in}}, \text{ and } \|z'_{n'_0}\| \leq R_2^{\text{in}} \ . \tag{88}$$

Let $(\theta_n, w_n)_{n \geq n'_0}$ be the iterates obtained by running the unprojected algorithm given in (1) and (2) with $\theta_{n'_0} = \theta'_{n'_0}$ and $w_{n'_0} = w'_{n'_0}$. Because of (88), it follows that (11) holds. Combining this with (84) and (87), it follows from Theorem 4 that

$$\begin{aligned}
&\Pr\{\|\theta_n - \theta^*\| \leq \epsilon, \|z_n\| \leq \epsilon, \forall\, n \geq 2n'_0\} \\
&\geq 1 - 2d^2 \sum_{n \geq n'_0} \left[\exp\left[\tfrac{-c_1\epsilon_1^2}{a_n}\right] + \exp\left[\tfrac{-c_2\epsilon_1^2}{b_n}\right] + \exp\left[\tfrac{-c_3\epsilon_2^2}{b_n}\right]\right] \\
&\geq 1 - 2d^2 \sum_{n \geq n'_0} \left[\exp\left[\tfrac{-c_1\epsilon_1^2}{a_n}\right] + 2\exp\left[\tfrac{-\min(c_2,c_3)\epsilon_1^2}{b_n}\right]\right].
\end{aligned} \tag{89}$$

As the next step, we claim that, for any n, the event

$$\{\|\theta_n - \theta^*\| \leq \epsilon, \|z_n\| \leq \epsilon\} \subseteq \{\theta_n = \Pi_{n, R_1^{\text{in}}/2}(\theta_n), w_n = \Pi_{n, R_2^{\text{in}}/2}(w_n)\} \ . \tag{90}$$

Indeed, due to (16) and the choice of ϵ, $\|\theta_n - \theta^*\| \leq \epsilon$ implies

$$\|\theta_n\| \leq \|\theta_n - \theta^*\| + \|\theta^*\| \leq \epsilon + R_1^{\text{in}}/4 \leq R_1^{\text{in}}/2 \tag{91}$$

and thus $\theta_n = \Pi_{n, R_1^{\text{in}}/2}(\theta_n)$. Separately, from the above relation and (17), we also have $\|\lambda(\theta_n)\| \leq R_2^{\text{in}}/4$. Because of this, (7), the fact that $\|z_n\| \leq \epsilon$, and the choice of ϵ, it then follows that $\|w_n\| \leq \|\lambda(\theta_n)\| + \|z_n\| \leq R_2^{\text{in}}/2$, and thus $w_n = \Pi_{n, R_2^{\text{in}}/2}(w_n)$.

An immediate consequence of (90) is that the event

$$\begin{aligned}
\mathcal{I} \ &:= \ \{\|\theta_j - \theta^*\| \leq \epsilon, \|z_j\| \leq \epsilon, \forall\, j \geq 2n'_0\} \\
&\subseteq \ \{\theta_j = \Pi_{j, R_1^{\text{in}}}(\theta_j), w_j = \Pi_{j, R_2^{\text{in}}}(w_j), \forall\, j \geq 2n'_0\} \ .
\end{aligned} \tag{92}$$

The statement of the theorem now follows by an easy coupling argument. For this, let

$$(\tilde{\theta}'_n, \tilde{w}'_n) := \begin{cases} (\theta'_n, w'_n), & \text{for } 0 \leq n < n'_0 \ , \\ (\theta_n, w_n), & \text{for } n \geq n'_0 \text{ on the event } \mathcal{I} \ , \\ (\theta'_n, w'_n), & \text{for } n \geq n'_0 \text{ on the complement of the event } \mathcal{I} \ . \end{cases} \tag{93}$$

Due to (92), $(\tilde{\theta}'_n, \tilde{w}'_n)_{n \geq 0}$ and $(\theta'_n, w'_n)_{n \geq 0}$ are distributed identically. This, together with (89) and Lemma 13, completes the proof of the claimed result. ■

Proof of Theorem 6, Statement 2 Let

$$N''_0(\epsilon, \delta, \alpha, \beta) = \max\left\{\left[\frac{1}{c_{6a}\epsilon^2} \log \frac{4d^2 c_{7a} e^{c_{5a}\epsilon^2}}{\epsilon^{2/\alpha}\delta}\right]^{1/\alpha}, \left[\frac{1}{c_{6b}\epsilon^2} \log \frac{8d^2 c_{7b} e^{c_{5b}\epsilon^2}}{\epsilon^{2/\beta}\delta}\right]^{1/\beta}\right\} \ .$$

Obviously, for any $n_0' \geq N_0''(\epsilon, \delta, \alpha, \beta)$,

$$2d^2 \frac{c_{7a}}{\epsilon^{2/\alpha}} \exp\left[c_{5a}\epsilon^2 - c_{6a}\epsilon^2 (n_0')^\alpha\right] + 4d^2 \frac{c_{7b}}{\epsilon^{2/\beta}} \exp\left[c_{5b}\epsilon^2 - c_{6b}\, \epsilon^2 (n_0')^\beta\right] \leq \delta \ .$$

Therefore, by Theorem 6, Statement 1,

$$\Pr\{\|\theta_n' - \theta^*\| \leq \epsilon, \|z_n'\| \leq \epsilon, \forall n \geq 2n_0'\} \geq 1 - \delta \tag{94}$$

for any $n_0' \geq \max\{N_0'(\epsilon, \alpha, \beta), N_0''(\epsilon, \delta, \alpha, \beta)\}$ such that n_0' is a power of 2. Thus,

$$\Pr\{\|\theta_n' - \theta^*\| \leq \epsilon, \|z_n'\| \leq \epsilon, \forall n \geq n_0'\} \geq 1 - \delta \tag{95}$$

for any $n_0' \geq 4 \max\{N_0'(\epsilon, \alpha, \beta), N_0''(\epsilon, \delta, \alpha, \beta)\}$. The factor 4 appears because the $2n_0'$ in (94) is replaced with n_0' in (95), and the fact that n_0' was earlier required to be a power of 2.

For any integer $n > 3$, we argue that there is some $\epsilon \equiv \epsilon(n)$ such that

$$n = 4 \max\{N_0'(\epsilon, \alpha, \beta), N_0''(\epsilon, \delta, \alpha, \beta)\} \ ;$$

indeed, as $N_0'(\epsilon, \alpha, \beta)$ and $N_0''(\epsilon, \delta, \alpha, \beta)$ are both defined to be the maximum of terms that strictly monotonically inrease as ϵ decreases—except for the constant 3 in (15)—such an $\epsilon(n)$ exists. Furthermore, it is also not difficult easy to see that

$$\epsilon(n) = O\left(\max\left\{n^{-\beta/2}\sqrt{\ln(n/\delta)},\ n^{\beta-\alpha)}\right\}\right) \ . \tag{96}$$

This, together with (95), implies

$$\Pr\{\|\theta_n' - \theta^*\| \leq \epsilon(n), \|z_n'\| \leq \epsilon(n)\} \geq 1 - \delta \tag{97}$$

for any $n > 3$, completing the proof. ■

Appendix F. Proofs from Section 5

Similarly to GTD(0) in Section 5, we now show how our assumptions hold, and with what constants, for GTD2 and TDC algorithms. Thus, in the same spirit as Corollary 12, similar results trivially follow for these algorithms as well.

F.1. GTD2

The GTD2 algorithm (Sutton et al., 2009b) minimizes the objective function

$$J^{\mathrm{MSPBE}}(\theta) = \tfrac{1}{2}(b - A\theta)^\top C^{-1}(b - A\theta). \tag{98}$$

The update rule of the algorithm takes the form of Equations (1) and (2) with

$$h_1(\theta, w) = A^\top w,$$
$$h_2(\theta, w) = b - A\theta - Cw,$$

and

$$M^{(1)}_{n+1} = \left(\phi_n - \gamma\phi'_n\right)\phi_n^\top w_n - A^\top w_n \ ,$$
$$M^{(2)}_{n+1} = r_n\phi_n + \phi_n[\gamma\phi'_n - \phi_n]^\top\theta_n - \phi_n\phi_n^\top w_n - [b - A\theta_n - Cw_n] \ .$$

That is, in case of GTD2 the relevant matrices in the update rules take the form $\Gamma_1 = 0$, $W_1 = -A^\top$, $v_1 = 0$, and $\Gamma_2 = A$, $W_2 = C$, $v_2 = b$. Additionally, $X_1 = \Gamma_1 - W_1W_2^{-1}\Gamma_2 = A^\top C^{-1}A$. By our assumptions, both W_2 and X_1 are symmetric positive definite matrices, and thus the real part of their eigenvalues are also positive. It is also clear that

$$\begin{aligned}
\|M^{(1)}_{n+1}\| &\leq (1+\gamma+\|A\|)\|w_n\|, \\
\|M^{(2)}_{n+1}\| &= \|r_n\phi_n - b + [A + \phi_n(\gamma\phi'_n - \phi_n)^\top]\theta_n - [\phi_n\phi_n^\top - C]w_n\| \\
&\leq 1 + \|b\| + (1+\gamma+\|A\|)\|\theta_n\| + (1+\|C\|)\|w_n\|.
\end{aligned}$$

Consequently, Assumption $\mathcal{A}_3$ is satisfied with constants $m_1 = (1+\gamma+\|A\|)$ and $m_2 = 1 + \max(\|b\|, \gamma+\|A\|, \|C\|)$.

F.2. TDC

The TDC algorithm is designed to minimize (98), just like GTD2.

The update rule of the algorithm takes the form of Equations (1) and (2) with

$$h_1\theta(\theta, w) = b - A\theta + [A^\top - C]w \ ,$$
$$h_2(\theta, w) = b - A\theta - Cw \ ,$$

and

$$M^{(1)}_{n+1} = r_n\phi_n + \phi_n[\gamma\phi'_n - \phi_n]^\top\theta_n - \gamma\phi'\phi^\top w_n - [b - A\theta_n + [A^\top - C]w_n] \ ,$$
$$M^{(2)}_{n+1} = r_n\phi_n + \phi_n[\gamma\phi'_n - \phi_n]^\top\theta_n - \phi_n\phi_n^\top w_n - [b - A\theta_n + Cw_n] \ .$$

That is, in case of TDC, the relevant matrices in the update rules take the form $\Gamma_1 = A$, $W_1 = [C - A^\top]$, $v_1 = b$, and $\Gamma_2 = A$, $W_2 = C$, $v_2 = b$. Additionally, $X_1 = \Gamma_1 - W_1W_2^{-1}\Gamma_2 = A - [C - A^\top]C^{-1}A = A^\top C^{-1}A$. By our assumptions, both W_2 and X_1 are symmetric positive definite matrices, and thus the real part of their eigenvalues are also positive. It is also clear that

$$\|M^{(1)}_{n+1}\| \leq 2 + (1+\gamma+\|A\|)\|\theta_n\| + (\gamma+\|A\|+\|C\|)\|w_n\|,$$
$$\|M^{(2)}_{n+1}\| = 2 + (1+\gamma+\|A\|)\|\theta_n\| + (1+\|C\|)\|w_n\| \ .$$

Consequently, Assumption $\mathcal{A}_3$ is satisfied with constants $m_1 = (2+\gamma+\|A\|+\|C\|)$ and $m_2 = (2+\gamma+\|A\|+\|C\|)$.