Fast and Robust Convergence Rate for TD(0) with Linear Function Approximation, Universal Learning Steps and I.I.D. Samples

Ziad Kobeissi
L2S, INRIA
Université Paris-Saclay, CentraleSupelec,
Gif-sur-Yvette, France

Éloïse Berthier
U2IS, ENSTA
Institut Polytechnique de Paris,
Palaiseau, France

Abstract

In this paper, we study the finite-time behavior of the TD(0) temporal-difference method with linear function approximation (LFA). We consider on-policy independent and identically distributed (i.i.d.) samples, a constant learning step, and the Polyak-Juditsky averaging method. We establish a new convergence rate, for the Mean-Square Error (MSE) on the approximated function, that is (i) *fast* in the sense that it admits an optimal dependency in the number of iterations k (i.e., of order $1/k$), (ii) is *robust* to ill-conditioning: it only depends on an initial error and model-independent constants and (iii) is *sharp* up to a multiplicative constant lower than 11. In particular, it does not depend on the smallest eigenvalue of the uncentered covariance matrix of the linear parametrization, unlike all pre-existing $O(1/k)$ rates in the TD(0) literature. We also introduce PCTD(0), a variant of TD(0), which benefits from better convergence properties under an additional assumption of strong mixing on the Markov Chain.

1 INTRODUCTION

Temporal Difference (TD) learning (Sutton, 1988; Sutton and Barto, 2018), is a fundamental algorithm in reinforcement learning (RL) for estimating the value function associated with a policy. Among the various TD variants, the TD(0) algorithm stands out for its simplicity and practical efficiency, particularly when combined with linear function approximation (LFA). Despite its widespread empirical success, obtaining sharp non-asymptotic convergence guarantees for TD(0) remains a significant theoretical challenge.

When used with LFA, TD(0) enters the more general framework of *linear stochastic approximation* (LSA) methods. Since the early work of Robbins and Monro (1951), it has been known that stochastic approximation (SA) schemes can achieve a $O(1/k)$ behavior under strong convexity assumptions, but these guarantees heavily depend on the constant of strong convexity. Later developments, such as the "robust SA" and mirror-descent approaches of Nemirovski et al. (2009) instead obtain non-asymptotic $O(1/\sqrt{k})$ bounds that hold for general convex problems and no longer depend on a strong convexity constant.

For practical SA algorithms, rates are in general either fast or robust, but rarely both. The most famous and impactful example of a fast and robust convergence rate applies to SGD in a particular setting. While many classical SGD analyses require strong convexity to get $O(1/k)$ rates, the breakthrough paper by Bach and Moulines (2013) showed that in least-squares regression, averaged SGD with a constant stepsize still attains a $O(1/k)$ convergence rate without any strong convexity assumption, thus providing *fast* yet *robust* guarantees in the non-strongly-convex setting.

In the whole field of RL, prior to our work, we are not aware of any rate on a non-tabular sample-based numerical method that is both fast and robust. For TD(0) with LFA, existing non-asymptotic convergent rates divide into two classes, aligned with the SA literature. The first one consists of fast $O(1/k)$ rates, Lakshminarayanan and Szepesvari (2018); Bhandari et al. (2018); Srikant and Ying (2019); Patil et al. (2023); Samsonov et al. (2024); Mitra (2024), where the constant in the $O(1/k)$ can be arbitrarily large in practice, especially for large dimensions where systems arising from modern ML are most often ill-conditioned. The dependence on the conditioning in the latter rates gen-

---

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

erally appears through the smallest eigenvalue of the uncentered covariance matrix of the linear parameterization, that we denote ω. We qualify this constant as being *model-dependent*, where the wording *model* refers, in the usual statistical sense, to the choice of the parameterization. It has to be noted that this ω admits the exact same definition as the strong convexity in Bach and Moulines (2013), making the comparison with this paper relevant.

The second class consists of robust rates, Bhandari et al. (2018); Liu and Olshevsky (2021); Lee and Orabona (2025). Even through they are of order $O(1/\sqrt{k})$, they are often thought to lead to tighter upper bounds in practice since they do not deteriorate when ω tends to zero. On the smallness of ω, Bach and Moulines (2013) write that *"typical ML problems are high dimensional and have correlated variables so that [ω] is zero or very close to zero, and in any case smaller than [$O(1/\sqrt{k})$]. This then makes the non-strongly convex methods better."*

In the relevant literature, the most popular conjecture on the open question of whether or not a fast and robust rate is attainable seems to be the negative answer. The second author in Bach and Moulines (2013), together with their coauthors in Samsonov et al. (2024), wrote: *"Note that the leading term of the bound in [our first main result for TD(0) (with i.i.d. samples)] includes factors of [$1/\omega$]. This dependence is generally unavoidable if one aims to obtain the MSE bound [on the approximated functions] that scales as* [$1/k$]*".* Furthermore, while the main purpose in Lakshminarayanan and Szepesvari (2018) (which appears in their title) is to obtain the sharpest rate on LSA and TD(0) using the same averaging method as in the present work, they did not attain a robust rate. They even provide a lower bound that scales as ω^{-2} on the MSE on the parameters (while we consider the MSE on the approximated functions, similarly to Samsonov et al. (2024), and others).

Contributions. Our main contributions are:

- Under standard assumptions, for learning rates lower than $\frac{2(1-\gamma)}{(1+\gamma)^2}$, we prove the first fast and robust convergence rate for TD(0), see Theorem 3.1. Proposition 3.2 shows that this rate is sharp up to a multiplicative constant lower than 11. This significantly improves on existing results as summarized in Table 1.
- In Theorem 3.3, we prove another convergence rate that benefits from the largest range of admissible learning rates in the literature up to our knowledge, namely $\alpha < \frac{2}{1+\gamma}$. Unlike the former rate, it is not robust but it remains faster than any existing rate prior to this work. We also prove that for $\alpha > \frac{2}{1+\gamma}$, convergence of TD(0) may fail.
- We propose two methods to reduce the dependence on $(1-\gamma)^{-1}$ of our robust rates when the transition operator of the Markov Chain admits a positive spectral gap $g > 0$. In the first method, we consider different learning rates for the constant part of the parametrization and the rest, see Theorem 3.5. For the second one, we introduce PCTD(0) which uses the differences of two independent copies of the Markov Chain to obtain centered feature maps. This allows one to: *(*i) consider γ larger than one; *(*ii) get a convergence rate independent of $(1-\gamma)^{-1}$ when $\gamma \leq 1$ (but dependent of g^{-1} that may be large in practice).
- For a more general class of LSA methods, which satisfy Assumptions **LS1-LS3**, we prove a convergence rate in $O(1/k)$ with a dependence onto the model only appearing in a faster converging $O(1/k^2)$ term, see Theorem 4.2.

Notations. The euclidean norm on $\mathbb{R}^d$ is denoted by $|\cdot|$, using the same symbol as the absolute value for real numbers and modulus for complex numbers, by a slight abuse of notations. The set of $d \times d$-sized square matrices is denoted by $\mathbb{R}^{d\times d}$. The operator norm on $\mathbb{R}^{d\times d}$ associated with the euclidean norm is denoted $\|\cdot\|_{\mathrm{op}}$, and the Frobenius norm is denoted $\|\cdot\|_F$. For S_1, S_2 two symmetric matrices, the notations $S_1 \leq S_2$ means that $S_2 - S_1$ is positive semi-definite (similarly we define $\geq$, $<$ and $>$ on the set of symmetric matrices). If S is a symmetric positive semi-definite matrix, its square root is denoted by $S^{\frac{1}{2}}$.

2 RELATED WORKS

2.1 TD(0) with LFA

The literature corpus of theoretical analyses of TD-learning is rich and has continuously developed over the past thirty years. The tabular version of the TD-learning algorithm was introduced by Sutton (1988), with first convergence results. Later, Tsitsiklis and Van Roy (1997) proved that TD(0) with LFA converges with probability one, in the case of linearly independent features, yet without an explicit convergence rate. Lakshminarayanan and Szepesvari (2018) proposed a non-asymptotic analysis in the i.i.d. sampling setting, while Bhandari et al. (2018) proposed a different analysis, applicable to the Markov sampling setting. Around the same period, Dalal et al. (2018) and Srikant and Ying (2019) derived alternative convergence rates, and Khamaru et al. (2021) introduced

Paper	Type of error	Step size	Averaging	Convergence rate
Bhandari et al. (2018)	$\mathbb{E}[\lvert v(X,\overline{\theta}_k) - v(X,\theta^*)\rvert^2]$	$O(K^{-\frac{1}{2}})$	yes	$O(K^{-\frac{1}{2}})$
	$\mathbb{E}\left[\lvert\theta_k - \theta^*\rvert^2\right]$	$O(k^{-1}\omega^{-1})$	no	$O(k^{-1}\omega^{-2})$
Srikant and Ying (2019)	$\mathbb{E}\left[\lvert\theta_k - \theta^*\rvert^2\right]$	$O(k^{-1}\omega^{-1})$	no	$O(k^{-1}\omega^{-1})$
Liu and Olshevsky (2021)	$\mathbb{E}[\lvert v(X,\overline{\theta}_K) - v(X,\theta^*)\rvert^2]$	$O(K^{-1/2})$	yes	$O(K^{-1/2})$
Lee and Orabona (2025)	$\mathbb{E}[\lvert v(X,\overline{\theta}_K) - v(X,\theta^*)\rvert^2]$	$O(K^{-1/2})$	yes	$O(K^{-1/2})$
Dalal et al. (2018)	$\mathbb{E}\left[\lvert\theta_k - \theta^*\rvert^2\right]$	$O(k^{-\beta})$	no	$O(k^{-\beta}\omega^{-1}e^{C(1-\beta)^{-1}\omega^{-\frac{1}{\beta}+1}})$
Lakshminarayanan and Szepesvari (2018)	$\mathbb{E}\left[\lvert\overline{\theta}_k - \theta^*\rvert^2\right]$	$O(1)$	yes	$O(k^{-1}\omega^{-2})$
Patil et al. (2023)	$\mathbb{E}\left[\lvert\overline{\theta}_k - \theta^*\rvert^2\right]$	$O(1)$	yes	$O(k^{-1}\omega^{-2})$
Mitra (2024)	$\mathbb{E}[\lvert v(X,\overline{\theta}_k) - v(X,\theta^*)\rvert^2]$	$O(1)$	yes	$O(k^{-1}\omega^{-2})$
Samsonov et al. (2024)	$\mathbb{E}[\lvert v(X,\overline{\theta}_k) - v(X,\theta^*)\rvert^2]$	$O(1)$	yes	$O(k^{-1}\omega^{-2})$
Our Theorem 3.1	$\mathbb{E}[\lvert v(X,\overline{\theta}_k) - v(X,\theta^*)\rvert^2]$	$O(1)$	yes	$O(k^{-1})$
Our Theorem 3.3	$\mathbb{E}[\lvert v(X,\overline{\theta}_k) - v(X,\theta^*)\rvert^2]$	$O(1)$	yes	$O(k^{-1}\omega^{-1})$

Table 1: **Different rates of convergence in expectation from the state of the art on TD(0) with i.i.d. samples.** On the first, fourth and fifth lines, K is the total number of iterations, that has to be known before starting the computations. For Dalal et al., $\beta \in (0,1)$ is a fixed parameter, observe that: letting β tend to one makes the term in the exponential blow up; taking β not near one implies that the dependency with respect to ω^{-1} is exponential. Some poly-logarithmic dependencies are omitted in the rates for simplicity. Note that some, but not all, results presented in the table also apply to the Markov sampling setting.

a variance-reduced variant of TD(0). Exploring different approximation schemes, Cai et al. (2019) proposed an analysis of a version of TD(0) where approximation is done using a finite-width one-hidden layer neural network, while Berthier et al. (2022) studied a non-parametric version of TD learning, using an infinite-dimensional linear approximation scheme. More recent analyses of TD(0) include those by Mitra (2024), Samsonov et al. (2024), Liu and Olshevsky (2021), Lee and Orabona (2025) and Patil et al. (2023).

These analyses essentially differ in a few crucial elements: the obtained convergence rate – fast or slow – often related to the choice of the learning rate, the dependence on unknown model-dependent quantities, the coverage or not of the Markov sampling scheme, and the potential need for projection. We detail below some of these elements.

2.2 Convergence rates

An important criterion to compare different analyses is whether or not, and how, the required learning rates and obtained convergence rates depend on quantities defined by the approximation model. While some dependence on the problem instance, like the initial error or the variance of the samples, is inevitable, a dependence on the model is usually unwanted. Indeed, such constants are typically unknown to the user, which might complicate the practical implementation of the method. Moreover, if such a dependence is present in the rate, it might become arbitrarily slow, typically as the dimension of the model grows. On the contrary, a rate is called robust if it does not depend on such model-dependent quantities. Such rates often extend to infinite-dimensional, universal approximation schemes, like in Berthier et al. (2022), which in turn remove approximation errors.

More specifically, in this paper, we study the mean squared error (MSE) in the setting of on-policy i.i.d. samples, which optimal convergence rates admit upper bounds proportional to $1/k$, where k is the number of iterations (Bhandari et al., 2018; Lakshminarayanan and Szepesvari, 2018; Patil et al., 2023; Samsonov et al., 2024). In Bhandari et al. (2018), the $1/k$-convergence rate is impractical because the learning rate depends on an intractable problem-specific constant, namely, ω that will be introduced later and was already discussed in Section 1. After that, it became an important issue for subsequent authors to use universal learning steps, that do not depend on ω or other intractable constants. Such a goal was achieved by Lakshminarayanan and Szepesvari (2018); Patil et al. (2023); Samsonov et al. (2024); Mitra (2024), whereas all their convergence rates kept on depending on ω.

Table 1 shows the explicit dependency on ω of some of the state-of-the-art convergence results. On the one hand, arbitrarily small ω may significantly weaken the rates from the above listed works. On the other hand, Bhandari et al. (2018) proved another convergence result with a rate independent of ω and any model-dependent constants, at the price of reducing the speed of convergence in k, to $O(1/\sqrt{k})$. Later works by Liu and Olshevsky (2021); Lee and Orabona (2025) obtained similar rates. Let us point out another common inconvenience of the latter three results: they require to know K the total number of iterations before starting the computations, in order to take a constant learning step of order $O(1/\sqrt{K})$. This implies that: (i) the rate is not guaranteed anymore if we increase the number of iterations afterwards; (ii) the learning can be slow at the beginning since the learning step is small for large K. As it will appear in our main result, Theorem 3.1, our rates do not suffer this issue.

2.3 Sampling schemes

In the literature, the convergence results for TD(0) with linear function approximation can be divided into two groups depending on the assumptions made on the data. In the first case, we assume that the samples used to compute the TD(0) iterates are i.i.d., directly sampled from the invariant distribution of the Markov chain. Whereas in the second situation, it is assumed that the samples are obtained from following the exponentially mixing Markov Chain associated with the dynamics. This case introduces a major difficulty: two consecutive samples are usually correlated, which introduces some complications in the analysis, handled by coupling arguments. Consequently, many analyses in the Markov sampling setting require to actively bound the iterates, which is usually ensured by adding a projection to the TD updates (Bhandari et al., 2018). Recent works suggest that this projection is not mandatory to obtain convergence (Samsonov et al., 2024; Mitra, 2024; Lee and Orabona, 2025).

In the present paper, we stick to the i.i.d. sampling setting. This choice is motivated by the simplicity of the analysis and the fact that it is often closer to practitioner's settings, where the samples can be obtained offline, or replayed using buffers, and do not necessarily strictly follow a trajectory. For instance, both AlphaGo Zero (Silver et al., 2017) and its open-source reimplementation (Tian et al., 2019) used a replay buffer with a size of 500,000 games (then, on average, a game contains more than 200 moves). In such a situation, two consecutively drawn samples have a probability of around 1/500,000 to be not independent, so the sampling is much closer to an i.i.d. model than to a single Markov trajectory. Finally, an important take-home message from the wide TD literature is that the convergence rates obtained in the Markov case do not significantly differ from the i.i.d. case, except for the introduction of a multiplicative factor due to mixing. For all these reasons, extending our present analysis to the Markov setting is left for future work.

3 CONTEXT AND MAIN RESULTS

3.1 Mathematical setting

For $(X_\ell, R_\ell)_{\ell\geq 0}$ a Markov Reward Process (MRP), we consider the problem of policy evaluation, consisting in approximating the value function V defined by

$$V_\gamma(x) = \mathbb{E}\left[\sum_{\ell=0}^{\infty} \gamma^\ell R_\ell \,\middle|\, X_0 = x\right], \tag{1}$$

for $\gamma \in [0,1)$, where R_ℓ is the reward at step k and the law of $(R_\ell, X_{\ell+1})$ is given by some probability kernel $((r,x') \mapsto P(x,r,x'))_{x\in\mathcal{X}} \in \mathcal{P}(\mathbb{R}\times\mathcal{X})^{\mathcal{X}}$. In particular, $(X_\ell)_\ell$ is a Markov chain of probability transition $(P_{x'}(x,\cdot))_{x\in\mathcal{X}} \in \mathcal{P}(\mathcal{X})^{\mathcal{X}}$, where $P_{x'}(x)$ is the second marginal of $P(x)$ and $P_r(x)$ its first one. The value function satisfies the Bellman equation

$$V_\gamma(x) = \mathbb{E}\left[R + \gamma V_\gamma(X') \,|\, (R, X') \sim P(x)\right]. \tag{2}$$

Our goal is to approximate V_γ using a parameterized function $v(\cdot,\theta)$ where $\theta\in\mathbb{R}^d$ is the parameter that we will learn. We make the following assumptions:

TD1 The Markov Chain induced by $P_{x'}$ admits an invariant probability measure m and we access i.i.d. samples (X_k, X'_k, R_k) with m being the law of X_k.

TD2 The second moment of $P_r(x)$ is uniformly bounded, i.e., $C_R := \sup_{x\in\mathcal{X}} \mathbb{E}_{R\sim P_r(x)}[R^2] < \infty$.

TD3 The parameterization is linear, i.e., $v(x,\theta) = \theta^\top \varphi(x)$ where $\varphi : \mathcal{X} \to \mathbb{R}^d$ are linearly independent on the support of m and satisfy $|\varphi| \leq 1$.

Let us define the uncentered covariance matrices:

$$\Sigma_0 = \mathbb{E}\left[\varphi(X)\varphi(X)^\top\right] \quad\text{and}\quad \Sigma_1 = \mathbb{E}\left[\varphi(X)(\varphi(X)')^\top\right].$$

The linear independence from Assumption **TD3** implies that Σ_0 is symmetric positive definite. Let $\omega > 0$ be its minimal eigenvalue, that was already discussed in Sections 1, 2. Consider $U = \Sigma_0^{-\frac{1}{2}}\Sigma_1\Sigma_0^{-\frac{1}{2}}$, we have

$$\Sigma_1 = \Sigma_0^{\frac{1}{2}} U \Sigma_0^{\frac{1}{2}} \quad\text{with}\quad \|U\|_{\text{op}} \leq 1,$$

as a consequence of the Cauchy-Schwarz inequality.

Under Assumptions **TD1-TD3**, the TD(0) algorithm consists in the following iterative method, from an initial parameter θ_0 and for $k\geq 1$:

$$\theta_k = \theta_{k-1} - \alpha\delta_k\varphi(X_k),$$

where $\delta_k = v(X_k,\theta_{k-1}) - \gamma v(X'_k,\theta_{k-1}) - R_k$ is named the temporal difference. As is usual in the literature, we do not consider the convergence of θ_k itself, but of its Polyak-Judistky averaging, denoted $\overline{\theta}_k$, defined by

$$\overline{\theta}_k = \frac{1}{k}\sum_{i=0}^{k-1}\theta_i.$$

Under convenient choice on the learning step α, it is known that $\overline{\theta}_k$ converge to θ^*, which satisfies

$$H_0\theta^* = b_0,$$

where $H_0 = (\Sigma_0 - \gamma\Sigma_1)$ and $b_0 = \mathbb{E}\left[R\,\varphi(X)\right]$. The value function associated with θ^* is the unique solution of the projected Bellman equation,

$$v(\cdot,\theta^*) = \Pi_{\{\theta^\top\varphi\}}\left(x \mapsto \mathbb{E}_{(X',R)\sim P(x)}\left[R + \gamma v(X',\theta^*)\right]\right),$$

which is similar to the Bellman equation (2) with an additional projection step onto the set of admissible functions. In particular, if the actual value function V belongs to the set of admissible functions, we get $V = v(\cdot,\theta^*)$ by uniqueness of the solution of the projected Bellman equation. Otherwise, the projection step generally introduces an approximation error.

3.2 Convergence rate of TD(0)

See Figure 1 for an overview of the results stated in this section. Our main result writes as follows.

Theorem 3.1. *Assume* **TD1**-**TD3**. *For* $\gamma \in [0,1)$, *define* $\alpha_0(\gamma) = \frac{2(1-\gamma)}{(1+\gamma)^2}$ *and* $\alpha_1(\gamma) = \frac{2}{1+\gamma} > \alpha_0(\gamma)$. *For* $0 < \alpha < \alpha_0(\gamma)$, *and* $k \geq 1$, *we have*

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \leq \frac{1}{(1-\gamma)^2 k}$$
$$\left(\left(\frac{|\theta_0-\theta^*|^2}{2\left(1-\frac{\alpha}{\alpha_1(\gamma)}\right)\frac{\alpha}{1-\gamma}}\right)^{\frac{1}{2}} + \left(\frac{de\sigma_0^2(1+\varepsilon_{k,\gamma})}{1-\frac{\alpha}{\alpha_0(\gamma)}}\right)^{\frac{1}{2}}\right)^2,$$

with $\varepsilon_{k,\gamma} = \frac{4}{k(1-\gamma)} + 3(1-\gamma) + \frac{10}{k} + \frac{2}{k(2-(1+\gamma)\alpha)}$, $\sigma_0^2 = \left\|x \mapsto \mathbb{E}[\delta(X,R,X',\theta^*)^2 \,|\, X = x]\right\|_\infty$ *and* $\delta(x,s,x',\theta^*) = v(x,\theta^*) - \gamma v(x',\theta^*) - s$.

As is customary in SA, the above rate is the sum of a bias term (which depends on $|\theta_0 - \theta^*|$) and a variance term (which depends on σ_0^2).

Given the above rate, the interesting regime is for $k \gg (1-\gamma)^{-2}$. In this case and for γ near one, $\varepsilon_{k,\gamma}$ is small.

As explained in the introduction, the above rate does not depend on ω the smallest eigenvalue of Σ_0, nor on any constant that might generally be qualified as model-dependent. The only two quantities in this rate that are unknown a priori come from the mathematical problem at hand. They are the initial error $|\theta_0 - \theta^*|^2$ and the quantity σ_0^2 that is often referred to as a variance (since $\mathbb{E}\left[\delta(X,R,X',\theta^*)\right] = (\Sigma_0 - \gamma\Sigma_1)\theta^* - b = 0$) or a residual of the Bellman equation, at the limit θ^*. We can easily bound σ_0^2 using $\sigma_0^2 \leq 4|\theta^*|^2 + 2C_R$, but we think that this inequality is in general very loose.

Up to our knowledge, this is the first convergence result for TD(0) that states a rate which is optimal in k and independent of ω at the same time. We refer to Table 1 for the dependency on ω of some state-of-the-art results in the setting of i.i.d. on-policy samples.

Observe that the upper bound in Theorem 3.1 gets unbounded for α near $\alpha_0(\gamma)$. In this case, it is standard in SA to restrain further the range of α by a factor $\frac{1}{2}$ and to consider only $\alpha \leq \frac{\alpha_0(\gamma)}{2}$. Under this standard regime, let us discuss the sharpness of Theorem 3.1.

Proposition 3.2. *Consider the constant Markov Chain on* $\mathcal{X} = \{1,\ldots,d\}$, *i.e.* $X' = X$ *almost surely. Take* $(X_k)_{k\geq 0}$ *i.i.d. with* $X_1 \sim m$, *where* $m \equiv \frac{1}{d}$ *is the uniform probability measure. Take* $(R_k)_{k\geq 1}$ *i.i.d. with* $R_1 \sim \mathcal{N}(0,\sigma_0^2)$ *for* $\sigma_0^2 > 0$, *and* $\varphi_i(x) = \sqrt{d\omega}\mathbb{1}_{\{i=x\}}$ *for* $\omega \in (0,\frac{1}{d}]$ *and* $1 \leq i,x \leq d$. *Assumptions* **TD1**-**TD3** *are satisfied and* ω *the smallest eigenvalue of* Σ_0. *Fix* $\lambda \in (0,\frac{1}{2}]$ *and* $\alpha = \lambda\alpha_0(\gamma)$, *in the regime* $k \to \infty$, $\gamma \to 1$ *and* $k(1-\gamma)\alpha\omega \to 5.5$, *we have*

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim \frac{1}{11(1-\gamma)^2 k}$$
$$\left(\left(\frac{|\theta_0-\theta^*|^2}{2\left(1-\frac{\alpha}{\alpha_1(\gamma)}\right)\frac{\alpha}{1-\gamma}}\right)^{\frac{1}{2}} + \left(\frac{de\sigma_0^2(1+\varepsilon_{k,\gamma})}{1-\frac{\alpha}{\alpha_0(\gamma)}}\right)^{\frac{1}{2}}\right)^2,$$

This implies that our upper bound in Theorem 3.1 is sharp up to at most equal to 11, for $0 < \alpha \leq \frac{\alpha_0(\gamma)}{2}$. We refer to Corollary C.3 where we prove that our upper bounds for the bias and variance terms are sharp up to constants 1.25 and $2e$ respectively.

Observe that this proposition holds for any $\omega \in (0,1)$ and does not rely on examples with specific or ill-conditioned structures; instead, our lower bound is obtained on one of the the simplest examples one can think of. Therefore, we believe that our upper bound in Theorem 3.1 captures the real convergence speed of practical computations of TD(0), on average and up to a multiplicative constant that should not be too big.

Now, let us describe what happens when $\alpha \geq \alpha_0(\gamma)$.

Theorem 3.3. *Assume* **TD1**-**TD3**. *For* $\gamma \in [0,1)$, $0 < \alpha < \alpha_1(\gamma) = \frac{2}{1+\gamma}$ *and* $k \geq 1$, *we have*

$$\mathbb{E}_m\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \leq \frac{|\theta_0-\theta^*|^2}{\left(1-\frac{\alpha}{\alpha_1(\gamma)}\right)\alpha(1-\gamma)k}$$
$$+ \frac{2d}{(1-\gamma)^2 k}\left(\frac{\alpha\sigma_1^2}{\left(1-\frac{\alpha}{\alpha_1(\gamma)}\right)\alpha_0(\gamma)\omega} + \sigma_0^2\right)$$
$$\times\left(3 + \frac{1}{2\left(1-\frac{\alpha}{\alpha_1(\gamma)}\right)\alpha(1-\gamma)\omega k}\right),$$

with $\sigma_1^2 = \mathrm{Var}(\delta(X,R,X',\theta^*)) \leq \sigma_0^2$. *Moreover, there*

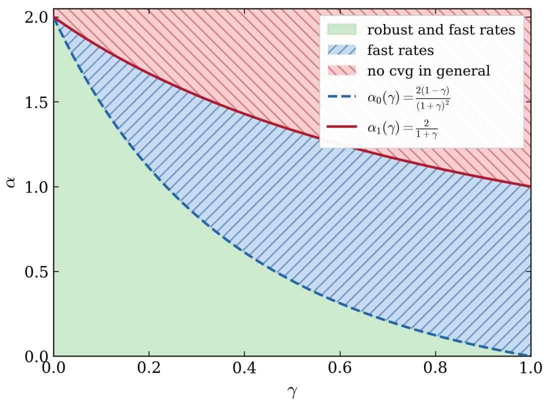


Figure 1: Proven convergence behaviors depending on the learning rate α. For $0 < \alpha < \alpha_0(\gamma)$, we proved a robust and fast convergence rate (Theorem 3.1) that is sharp up to a constant lower than 11 (Proposition 3.2). For $\alpha_0(\gamma) \leq \alpha < \alpha_1(\gamma)$, we proved faster convergence rate than any previous work (Theorem 3.3). For $\alpha > \alpha_1(\gamma)$, we proved that TD(0) may diverge (Theorem 3.3).

exist an MRP and feature functions such that Assumptions **TD1**-**TD3** *are satisfied and, for any* $\alpha > \alpha_1(\gamma)$,

$$\lim_{k\to\infty} \mathbb{E}_m\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] = \infty = \lim_{k\to\infty} \mathbb{E}\left[|\theta_k|^2\right].$$

Let us discuss the latter convergence result. In the regime $\alpha < \alpha_0(\gamma)$, the rate is slower than that of Theorem 3.1. However, it applies to a wider range of learning rates. Furthermore, it is both faster and valid for a broader range of α than any previously known rate. To our knowledge, Lakshminarayanan and Szepesvari (2018) allowed the widest range of learning rates, namely $\alpha < 1$, while Samsonov et al. (2024) achieved the fastest rate (for $\alpha \leq \frac{1-\gamma}{256}$), with a variance term of order $O(\frac{\alpha\sigma_0^2}{(1-\gamma)^3\omega^2 k})$. In comparison, we obtain $O(\frac{\alpha\sigma_1^2}{(1-\gamma)^3\omega k})$. Moreover, Theorem 3.3 shows that this range cannot be extended further in general.

3.3 Reducing the dependence on $1-\gamma$

In fact, $|\theta^*|$ generally grows as $O((1-\gamma)^{-1})$ when γ tends to 1, see Lemma C.5. Therefore, the bias term in Theorem 3.1 is of order $O(k^{-1}(1-\gamma)^{-4})$. Here, we propose a parameterization trick that allows us to keep θ^* bounded and get a $O(k^{-1}(1-\gamma)^{-2})$ rate.

td3 Let $\varphi : \mathcal{X} \to \mathbb{R}^d$ be linearly independent on the support of m, with $\varphi_1 \equiv c_\gamma$ and $|\varphi_{(-1)}| \leq 1$.

In fact, the latter assumption simply corresponds to having a different learning rate for the constant part of the parametrization and for the rest, as follows.

Lemma 3.4. *For* $C > 0$, *TD(0) computes the same value function in any of the following cases:*

(i) **td3** *with* $c_\gamma = C$, θ_0 *and* $\alpha > 0$.

(ii) **td3** *with* $c_\gamma = 1$, $\widetilde{\theta}_0 := (C\theta_{0,1}, \theta_{0,(-1)})$ *and* $\widetilde{\alpha} = \mathrm{diag}(C^2\alpha, \alpha, \ldots, \alpha) \in \mathbb{R}^{d\times d}$.

Let us introduce the following standard assumption.

TD4 The Markov Chain admits a positive spectral gap, i.e., there exists $g > 0$ such that, for any $f \in L^2(m)$ with $\mathbb{E}[f(X)] = 0$,

$$\mathbb{E}_m\left[\mathbb{E}[f(X') \mid X]^2\right] \leq (1-g)\mathbb{E}_m\left[f(X)^2\right]$$

Observe that g is a priori unknown and may be arbitrarily small in practice. However, it is model-independent: it does not depend on the choice of the feature maps φ but only on the Markov Chain itself.

Theorem 3.5. *Assume* **TD1**, **TD2**, **td3** *with* $c_\gamma = \frac{1}{1-\gamma}$ *and* **TD4**. *For* $\gamma \in [\frac{1}{2}, 1)$ *and* $0 < \alpha < \alpha_2(\gamma) := \frac{2(1-\gamma)}{1+(1+\gamma)^2}$, *the rate of Theorem 3.1 applies with* $\alpha_0(\gamma)$ *replaced by* $\alpha_2(\gamma)$ *and* θ^* *uniformly bounded in* γ.

3.4 Pairwise Centered TD(0) (PCTD(0))

Without loss of generality, the samples of Assumption **TD1** can be divided into two independent sequences (X_k, X'_k, R_k) and $(\widetilde{X}_k, \widetilde{X}'_k, \widetilde{R}_k)$. We then define the *pairwise centered temporal difference* (PCTD) as

$$\begin{aligned}\widetilde{\delta}_k =& \frac{1}{2}(v(X_k,\theta_{k-1}) - \gamma v(X'_k,\theta_{k-1}) - R_k) \\ &- (v(\widetilde{X}_k,\theta_{k-1}) - \gamma v(\widetilde{X}'_k,\theta_{k-1}) - \widetilde{R}_k).\end{aligned} \tag{3}$$

Then PCTD(0) is defined using the update rule

$$\theta_k = \theta_{k-1} - \alpha\widetilde{\delta}_k(\varphi(X_k) - \varphi(\widetilde{X}_k)). \tag{4}$$

Its eventual limit is $\widehat{\theta}^* = \widehat{H}^{-1}\mathbb{E}[R(\varphi(X) - \mu)]$ where $\widehat{H} := H - (1-\gamma)\mu\mu^\top$ is invertible under **TD4** if the set of parametrized function does not contain $\mathbb{1}$.

Proposition 3.6. *PCTD(0) is an instance of TD(0) on the MRP defined on* $\mathcal{X} \times \mathcal{X}$ *by* $((X_\ell, \widetilde{X}_\ell), R_\ell - \widetilde{R}_\ell)_{\ell\geq 0}$ *with feature maps* $\widetilde{\varphi}(x,\widetilde{x}) = \frac{1}{\sqrt{2}}(\varphi(x) - \varphi(\widetilde{x}))$, *where* $(X_\ell, R_\ell)_{\ell\geq 0}$ *and* $(\widetilde{X}_\ell, \widetilde{R}_\ell)_{\ell\geq 0}$ *are two independent copies of the MRP defined in Section 3.1. Therefore, all existing results for TD(0) apply to PCTD(0), for both i.i.d. and Markovian sampling schemes.*

However, PCTD(0) has an important advantage over TD(0): it is centered, i.e., $\mathbb{E}_{m\otimes m}[\widetilde{\varphi}(X, \widetilde{X})] = 0$.

Theorem 3.7. *Assume* **TD1**-**TD4**. *The rates from Theorems 3.1 and 3.3 hold with* γ *replaced by* $(1-g)\gamma$ *and the range of admissible* γ *becoming* $\gamma \in [0, \frac{1}{1-g})$.

Observe that γ can be larger than one in the latter theorem. Moreover, for $\gamma \leq 1$, it yields the first convergence rate independent of γ for a variant of TD(0) with general mixing Markov Chains up to our knowledge. Indeed, Liu and Olshevsky (2021) claim to obtain a slow rate of order $O(\frac{1}{\sqrt{k}})$ independent of γ, but it is in fact only of order $O((1-\gamma)^{-2}k^{-\frac{1}{2}})$ since it depends on $|\theta^*|^2$ which is of order $(1-\gamma)^{-2}$ in general, as proven in Lemma C.5.

Let us mention that most TD algorithms (such as $TD(\lambda)$) admit a PC variant with properties similar to Proposition 3.6 applying, allowing one to get convergence results without further analysis. We believe that in most cases, these variants benefit from better convergence behaviors than their original versions, as exemplified by PCTD(0) using Theorem 3.7.

4 RESULTS ON LSA

As is customary in the literature establishing convergence rates of TD(0) with LFA and i.i.d. samples, we will consider the larger class of *linear stochastic approximation* methods. To make a clear distinction with the framework of TD(0) introduced in the previous section, we are going to use different notations.

The goal here is to approximate $y^* \in \mathbb{R}^d$ satisfying

$$Hy^* = b,$$

for some invertible matrix $H \in \mathbb{R}^{d\times d}$ and vector $b \in \mathbb{R}^d$. To do so, we are going to use the iterative method described by, for some initial state y_0, for any $k \geq 1$,

$$y_k = y_{k-1} - \alpha(h_k y_{k-1} - b_k),$$

where $\alpha > 0$ is the learning step, (h_k, b_k) are couples of random variables in $\mathbb{R}^{d\times d} \times \mathbb{R}^d$. Let us assume

LS1 The symmetric part of H, denoted $S := \frac{H+H^\top}{2}$, satisfies $S \geq \mu I_d$ for some $\mu > 0$.

LS2 $(h_k, b_k)_{k\geq 1}$ is i.i.d. with $\mathbb{E}[h_k] = H$ and $\mathbb{E}[b_k] = b$.

LS3 There exist $\Sigma \in \mathbb{R}^{d\times d}$ symmetric definite positive and constants $c_\Sigma, \beta, \beta_1, \sigma^2$ such that $\Sigma \leq c_\Sigma S$ and

$$\mathbb{E}\left[h_k h_k^\top\right] \leq \beta\Sigma,$$
$$\mathbb{E}\left[(h_k y^* - b_k)(h_k y^* - b_k)^\top\right] \leq \sigma^2\Sigma,$$
$$\mathbb{E}\left[h_k^\top h_k\right] \leq \beta_1 S.$$

Observe that we can always take $\Sigma = S$ and $c_\Sigma = 1$ in the Assumption **LS3**; but, to obtain our sharper rates on TD(0), we will need to take Σ and S distinct.

We first state a proposition that will be useful to prove our main results on TD(0) as well as for our more general result on LSA, Theorem 4.2 below.

Proposition 4.1. *Under Assumptions* **LS1**-**LS3**, *for* $0 < \alpha < \min(2c_\Sigma^{-1}\beta^{-1}, 2\beta_1^{-1})$, *we have*

$$\mathbb{E}\left[(\overline{y}_k^\top - y^*)^\top \Sigma(\overline{y}_k - y^*)\right] \leq \left(\left(\frac{c_\Sigma|\theta_0 - \theta^*|^2}{\alpha(2-\alpha\beta_1)k}\right)^{\frac{1}{2}} + \left(\frac{2\sigma^2\mathcal{S}_k}{(2-\alpha\beta c_\Sigma)k}\right)^{\frac{1}{2}}\right)^2.$$

and $\mathcal{S}_k \leq c_\Sigma^2(3d + \widetilde{\mathcal{S}}_k)$, *where* $\mathcal{S}_k$ *and* $\widetilde{\mathcal{S}}_k$ *are defined by*

$$\mathcal{S}_k = \frac{1}{k}\sum_{j=0}^{k-1}\left\|\left(I_d - (I_d - \alpha\Sigma^{\frac{1}{2}}H\Sigma^{-\frac{1}{2}})^j\right)\Sigma^{\frac{1}{2}}H^{-1}\Sigma^{\frac{1}{2}}\right\|_F^2$$

$$\widetilde{\mathcal{S}}_k = \frac{1}{k}\sum_{j=0}^{k-1}\left\|\left(I_d - \alpha\Sigma^{\frac{1}{2}}H\Sigma^{-\frac{1}{2}}\right)^j\right\|_F^2.$$

Our main convergence result on LSA methods is:

Theorem 4.2. *Under Assumptions* **LS1**-**LS3**, *for* $0 < \alpha < \min(2c_\Sigma^{-1}\beta^{-1}, 2\beta_1^{-1})$, *we have*

$$\mathbb{E}\left[(\overline{y}_k^\top - y^*)^\top \Sigma(\overline{y}_k - y^*)\right] \leq \frac{2c_\Sigma|y_0 - y^*|^2}{(2-\alpha\beta_1)\alpha k} + \frac{2\sigma^2 c_\Sigma^2 d}{k(2-\alpha\beta c_\Sigma)}\left(3 + \frac{1}{\alpha(2-\alpha\beta_1)\mu k}\right).$$

Let us mention that this theorem is more general than Theorem 3.1 in the sense that it holds for a larger class of methods. However, its convergence rate is weaker: only its leading term with respect to k is model-independent, and there is an additional error term that is model-dependent but admits a faster decay in k^{-2}. In the end, the rate is of order $O(k^{-1} + \mu^{-1}k^{-2})$ where μ is typically the smallest eigenvalue of S. Up to our knowledge, the fastest convergence rates for LSA methods proved in the literature of TD(0) with similar assumptions are of order $O(k^{-1}\mu^{-2})$ (see Lakshminarayanan and Szepesvari (2018); Samsonov et al. (2024)), which makes our rate faster.

In the appendix, we present two extensions of this theorem. The first one, Theorem B.2, does not require the third inequality in Assumption **LS3**, applies to learning rates $\alpha < \frac{2}{c_\Sigma\beta}$ and admits a $O(k^{-1} + \mu^{-1}k^{-2})$ convergence rate, similar to that of Theorem 4.2. The second extension, Theorem B.3, allows to take $\alpha < \frac{2}{\beta_1}$ and admits a slower convergence rate of $O(k^{-1}\mu^{-1})$. Up to our knowledge, even this latter rate is faster than any existing one prior to this work.

5 THE SPECIFIC CASE OF TD(0)

5.1 A first Kreiss-like convergence rate

Given Proposition 4.1, the remaining of the proof of Theorem 3.1 relies on arguments from complex analysis and is highly technical. However, it is possible to get a robust $O(k^{-1})$ rate much more easily. This is what we will show in this subsection, in order to illustrate the type of arguments we used to prove Theorem 3.1. More precisely, this simpler result mainly relies on a classical result from the literature known as Kreiss' Theorem, that we state here in the same formulation as the one from Trefethen and Embree (2005).

Theorem 5.1 (Kreiss Theorem). *Let $Q \in \mathbb{R}^{d\times d}$ be a matrix such that $\sup_{k\geq 1} \left\|Q^k\right\|_{\mathrm{op}} < \infty$, then we have*

$$\mathcal{K}(Q) \leq \sup_{k\geq 1} \left\|Q^k\right\|_{\mathrm{op}} \leq d\, e\, \mathcal{K}(Q),$$

where $\mathcal{K}(Q)$ is the Kreiss constant defined by

$$\mathcal{K}(Q) = \sup_{|z|>1} (|z|-1) \left\|(Q-z)^{-1}\right\|_{\mathrm{op}}.$$

In general, the Kreiss constant may depend on the spectrum of Q, so the latter theorem cannot be used to directly obtain a robust rate on LSA methods. However, in the case of TD(0), we will be able to bound the Kreiss constant of $Q_0 := (I_d - \alpha \Sigma_0^{\frac{1}{2}} H_0 \Sigma_0^{-\frac{1}{2}})$ uniformly with respect to the linear model. To do so, we will use the specific structure of H_0, that is $H_0 = \Sigma_0^{\frac{1}{2}}(I_d - \gamma U)\Sigma_0^{\frac{1}{2}}$, with $\|U\|_{\mathrm{op}} \leq 1$. For $z \in \mathbb{C}$ with $|z| > 1$, we then get

$$\begin{aligned}(z - Q_0)^{-1} &= \left((z-1)I_d + \alpha\Sigma_0 - \gamma\alpha\Sigma_0 U\right)^{-1} \\ &= \left(I_d - \gamma\alpha((z-1)I_d + \alpha\Sigma_0)^{-1}\Sigma_0 U\right)^{-1} \\ &\quad \times \left((z-1)I_d + \alpha\Sigma_0\right)^{-1}.\end{aligned}$$

On the one hand, using that Σ_0 is symmetric positive definite and the triangular inequality on $\mathbb{C}$, we have

$$\begin{aligned}\left\|(z-1)I_d + \alpha\Sigma_0)^{-1}\right\|_{\mathrm{op}} &\leq \sup_{\lambda\in \mathrm{Sp}(\Sigma_0)} \frac{1}{|z-1+\alpha\lambda|} \\ &\leq \sup_{\lambda\in \mathrm{Sp}(\Sigma_0)} \frac{1}{|z| - (1-\alpha\lambda)} \\ &\leq \frac{1}{|z|-1}.\end{aligned}$$

On the second hand, observe that

$$\begin{aligned}&\left\|\alpha((z-1)I_d + \alpha\Sigma_0)^{-1}\Sigma_0 U\right\|_{\mathrm{op}} \\ &\qquad \leq \left\|\alpha((z-1)I_d + \alpha\Sigma_0)^{-1}\Sigma_0\right\|_{\mathrm{op}} \\ &\qquad = \sup_{\lambda\in \mathrm{Sp}(\Sigma_0)} \frac{\alpha\lambda}{|z-1+\alpha\lambda|} \\ &\qquad \leq \sup_{\lambda\in \mathrm{Sp}(\Sigma_0)} \frac{\alpha\lambda}{|z|-(1-\alpha\lambda)} \leq 1,\end{aligned}$$

which implies that

$$\left\|\left(I_d - \gamma\alpha((z-1)I_d + \alpha\Sigma_0)^{-1}\Sigma_0 U\right)^{-1}\right\|_{\mathrm{op}} \leq \frac{1}{1-\gamma}.$$

From the definition of the Kreiss constant and the above inequalities, we get

$$\mathcal{K}(Q_0) \leq \frac{1}{1-\gamma} \tag{5}$$

Assuming that Proposition 4.1 applies to TD(0) with $\Sigma = \Sigma_0$, $\beta = 4$, $c_\Sigma = \frac{1}{1-\gamma}$, $\beta_1 = 2$, and $\sigma^2 = \sigma_0^2$, we obtain $\widetilde{\mathcal{S}}_k \leq de\mathcal{K}(Q_0) \leq \frac{de}{1-\gamma}$ and the following robust $O(1/k)$-convergent rate:

Theorem 5.2. *Under Assumptions* **TD1-TD3***, for $0 < \alpha \leq \frac{1-\gamma}{8}$, for $k \geq 1$, we have*

$$\begin{aligned}&\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \\ &\qquad \leq \frac{2|\theta_0-\theta^*|^2}{(1-\gamma)\alpha k} + \frac{4\sigma_0^2 d}{(1-\gamma)^2 k}\left(1 + \frac{d^2e^2}{(1-\gamma)^2}\right).\end{aligned}$$

We do not consider this theorem as one of our main results, since its convergence rate is loose compared to the one of Theorem 3.1. We only presented it to show a simple proof of a first model-independent $O(k^{-1})$-rate.

5.2 The additional argument for Theorem 3.1

One may prove that our upper bound on the Kreiss constant (5) is sharp on the set of admissible matrices Q_0 obtained from instances of TD(0). Moreover, it is known, see Trefethen and Embree (2005), that the inequalities in Theorem 5.1 are sharp for general power-bounded matrices Q. So one may ask: how can the constants in the rate of Theorem 5.2 be improved to get the one in Theorem 3.1? The response lies in the fact that the right inequality from Theorem 5.1 is not sharp over the class of admissible Q_0. In the end, using the particular structure of H_0 once more, we obtain the subsequent proposition that is the last missing part to prove Theorem 3.1.

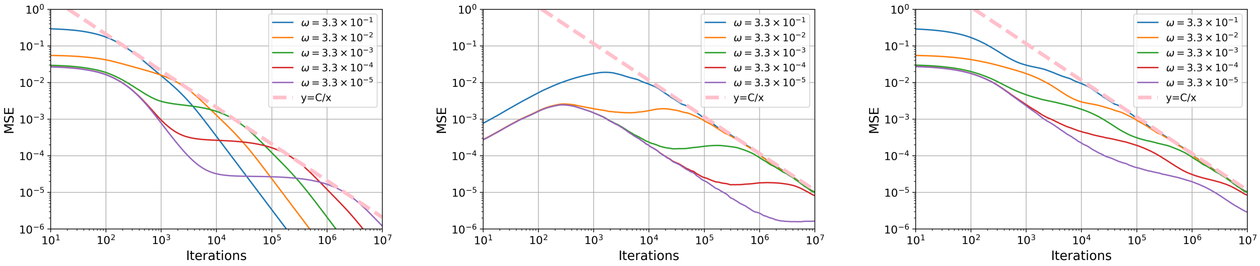
Proposition 5.3. *Assume* **TD1-TD3***. For $\varepsilon_{k,\gamma}$ and $\mathcal{S}_k$ defined as in Theorem 3.1 and Proposition 4.1, for $\alpha \in (0, \frac{1-\gamma}{1+\gamma})$, $k \geq 1$, we have $\mathcal{S}_k \leq \frac{de}{(1-\gamma)^2}\left(1 + \varepsilon_{k,\gamma}\right)$.*

Theorem 3.1 is then a direct consequence of the latter proposition and Proposition 4.1 with $\Sigma = \Sigma_0$, $c_\Sigma = (1-\gamma)^{-1}$, $\beta = (1+\gamma)^2$, $\beta_1 = 1+\gamma$ and $\sigma^2 = \sigma_0^2$.

6 NUMERICAL SIMULATIONS

Convergence independent of ω We consider $d = 3$ and draw a random MRP on $\mathcal{X} = \{1,2,3\}$ with $\omega = \frac{1}{3} \times 10^{-j}$ for $0 \leq j \leq 4$ as described in Appendix A.1. Figure 2 shows the convergence of the bias, the variance and the total MSE error on the values.

Focusing first on the bias, we observe that each curve exhibits a threshold, seemingly proportional to ω^{-1},



(a) Bias

(b) Variance

(c) Total error

Figure 2: We take $d = 3$ with $\lambda_1 = 1$ and $\lambda_2 = \lambda_3 = 10^{-j/2}$ for $0 \leq j \leq 4$, so that $\omega \in \{\frac{1}{3} \times 10^{-j},\ 0 \leq j \leq 4\}$. We use $\gamma = 0.9$ and $\alpha = \frac{1-\gamma}{4} = 0.025$. We draw the bias on the left (taking $\sigma_R^2 = 0$), the variance in the middle (taking $\theta_0 = \theta^* = 0$ and $\sigma_R^2 = 1$) and the total MSE error on the right. The curves are averaged over 10 simulations for 2a and 1000 simulations for 2b and 2c, always with the same transition matrix, initial parameter and matrices P, Q (that have been randomly drawn with $\alpha_w = 1$). In any case, the dashed pink line is the smallest curve of the form $y = C/x$ that upper bounds all curves from the same figure.

after which the bias decays quadratically in k. Before this threshold, the bias stays below a bound of the form $y = C/x$, with C apparently independent of ω.

For the variance term, it initially increases because too little noise has been observed, then it decreases. It also appears to switch between two envelopes of the form $y = C/x$ at an iteration count proportional to ω^{-1}.

For the total error (which is upper bounded by twice the sum of the bias and the variance terms), the bias dominates at the beginning; once it enters its quadratic decay regime, the variance eventually becomes dominant. The crossover again occurs after a number of iterations proportional to ω^{-1}. Finally, all curves remain bounded by a function C/x with C independent of ω, in agreement with Theorem 3.1.

7 CONCLUSIONS

In this paper, we derived a robust convergence rate of order $1/k$ for the TD(0) algorithm with i.i.d. samples, which uses a universal learning rate. The main argument behind these results lies in proving new Kreiss-like estimates adapted to the particular structure of TD(0). Such estimates describe the behavior of dynamical systems obtained by induction on a linear system with a non-symmetric associated matrix. We also derived a lower bound on the convergence of TD(0) that matches our convergence rate up to a multiplicative constant lower than 11. This lower bound is obtained on simple examples, so that we believe that the actual asymptotic behavior of TD(0) is accurately captured by our analysis. This belief is aligned with our numerical simulations.

As a consequence, we proved with theoretical arguments that the main limitation in the convergence of TD(0) is in fact the smallness of $(1-\gamma)$. Without further assumptions, under the regime of our main result, Theorem 3.1, the part of the error identified as the bias is of order $O((1-\gamma)^{-4}k^{-1})$, while the variance part is only of order $((1-\gamma)^{-2}k^{-1})$. When the Markov Chain is strongly mixing, we propose two methods to reduce the latter dependence. Among them, the second one is particularly promising. It consists of introducing a variant to TD(0) that we name PCTD(0) that can be computed for $\gamma = 1$ and which admits robust and fast convergence rate independent of $(1-\gamma)^{-1}$.

Along with this main result, we developed related bounds for general LSA schemes and obtained a convergence rate of order $1/k$ with a model-independent leading term, where the dependency on the model lies in a faster converging term in $1/k^2$.

While such results significantly improve our understanding of the behavior of TD(0), several open questions remain:

1. Defining $\alpha_{\text{rob}}(\gamma) > 0$ as the supremum of learning rates that allow robust and fast convergence in the present regime, we proved that $\alpha_1(\gamma) \leq \alpha_{\text{rob}}(\gamma) \leq \alpha_2(\gamma)$. Can we characterize α_{rob} more precisely?

2. Is it possible to adapt the main result to infinite-dimensional linear approximation, so that one could potentially combine optimal rates and universal approximation?

3. Can the proof be extended to the Markovian sampling setup, and is it possible to keep similar rates, only affected by mixing constants? Alternatively, is it possible to identify fundamental limitations in the Markovian setup hindering such results?

4. Do such proof techniques adapt to other, more complex algorithms in reinforcement learning beyond policy evaluation, such as SARSA, Q-learning, or actor-critic methods?

Bibliography

Francis Bach and Eric Moulines. Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$. *Advances in Neural Information Processing Systems*, 26:773–781, 2013.

Éloïse Berthier, Ziad Kobeissi, and Francis Bach. A non-asymptotic analysis of non-parametric temporal-difference learning. *Advances in Neural Information Processing Systems*, 2022.

Jalaj Bhandari, Daniel Russo, and Raghav Singal. A finite time analysis of temporal difference learning with linear function approximation. In *Conference on Learning Theory*, pages 1691–1692, 2018.

Qi Cai, Zhuoran Yang, Jason D. Lee, and Zhaoran Wang. Neural temporal-difference learning converges to global optima. *Advances in Neural Information Processing Systems*, 32, 2019.

Gal Dalal, Balázs Szörényi, Gugan Thoppe, and Shie Mannor. Finite sample analyses for TD(0) with function approximation. *AAAI'18/IAAI'18/EAAI'18*, 2018.

Koulik Khamaru, Ashwin Pananjady, Feng Ruan, Martin J. Wainwright, and Michael I. Jordan. Is temporal difference learning optimal? an instance-dependent analysis. *SIAM Journal on Mathematics of Data Science*, 3(4):1013–1040, 2021.

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.

Wei-Cheng Lee and Francesco Orabona. A finite-time analysis of TD learning with linear function approximation without projections nor strong convexity. *arXiv preprint arXiv:2506.01052*, 2025.

Rui Liu and Alex Olshevsky. Temporal difference learning as gradient splitting. In *International Conference on Machine Learning*, pages 6905–6913. PMLR, 2021.

Aritra Mitra. A simple finite-time analysis of TD learning with linear function approximation. *IEEE Transactions on Automatic Control*, 2024.

Arkadi Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

Gandharv Patil, LA Prashanth, Dheeraj Nagaraj, and Doina Precup. Finite time analysis of temporal difference learning with linear function approximation: Tail averaging and regularisation. In *International Conference on Artificial Intelligence and Statistics*, pages 5438–5448, 2023.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

Sergey Samsonov, Daniil Tiapkin, Alexey Naumov, and Eric Moulines. Improved high-probability bounds for the temporal difference learning algorithm via exponential stability. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4511–4547, 2024.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.

Rayadurgam Srikant and Lei Ying. Finite-time error bounds for linear stochastic approximation and TD learning. In *Conference on Learning Theory*, pages 2803–2830, 2019.

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

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: an Introduction*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, second edition, 2018.

Yuandong Tian, Jerry Ma, Qucheng Gong, Shubho Sengupta, Zhuoyuan Chen, James Pinkerton, and Larry Zitnick. Elf opengo: An analysis and open reimplementation of alphazero. In *International Conference on Machine Learning*, pages 6244–6253. PMLR, 2019.

Lloyd N. Trefethen and Mark Embree. Spectra and pseudospectra: The behavior of nonnormal matrices and operators, 2005.

John N. Tsitsiklis and Benjamin Van Roy. An analysis of temporal-difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42(5):674–690, 1997. doi: 10.1109/9.580874.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [**Yes**] Such a description is given in Section 3.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [**Yes**] Explicit convergence rates are obtained throughout the article.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [**Not Applicable**]
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [**Yes**] All assumptions are explicitly introduced before each result.
   (b) Complete proofs of all theoretical results. [**Yes**] All proofs are provided in the Appendix.
   (c) Clear explanations of any assumptions. [**Yes**] Specific assumptions are all discussed, in relation with the existing literature.
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [**Yes**]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [**Yes**]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [**Yes**]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [**Yes**]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [**Not Applicable**]
   (b) The license information of the assets, if applicable. [**Not Applicable**]
   (c) New assets either in the supplemental material or as a URL, if applicable. [**Not Applicable**]
   (d) Information about consent from data providers/curators. [**Not Applicable**]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [**Not Applicable**]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [**Not Applicable**]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [**Not Applicable**]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [**Not Applicable**]

A Additional information on the numerical simulations

All experiments were run on a standard personal laptop (Intel Core i7, 32 GB RAM) using only CPU. Total computation time amounted to a few hours.

A.1 Non-reversible MRP with fixed eigenvalues

In this section, we describe how we create MRP with the following properties:

1. we can fix the spectrum of Σ_0: it is particularly useful to illustrate the independence of the convergence with respect to ω the smallest eigenvalue of Σ_0.
2. It is not reversible. Indeed, we believe that reversibility is both unrealistic and overly simplistic. For reversible MRPs, the theoretical analysis becomes significantly simpler. In this case, most of the matrices involved are symmetric and commute with one another. As a result, one can obtain results similar to those proved in the main body of this article with considerably less effort. In particular, much of our theoretical analysis becomes unnecessary, especially the part relying on complex analysis.
3. It is random.

The Markov chain. We take $\{1, \ldots, n\}$ as the state space and $T \in \mathbb{R}^{n\times n}$ as the transition matrix. Recall that T is stochastic by definition, i.e., its coefficients are nonnegative and the sum of any row equals one. Here, we make the additional assumption that T is bistochastic, i.e., any column sums to one as well. This implies that the uniform distribution $m = \frac{1}{n}$ is invariant. For $n \leq 2$, any bistochastic matrix is in fact symmetric which implies that the Markov Chain is reversible. For this reason, we always assume $n \geq 3$ when working with bistochastic matrices.

It is well-known that the set of bistochastic matrices is exactly the convex hull of permutations matrices and therefore is of dimension $(n-1)^2$. Therefore, using Carathéodory's theorem, we can draw T of the form we can randomly draw a bistochastic T of the form $T = \sum_{i=1}^{(n-1)^2+1} w_i p_i$, where $(p_i)_i$ are permutation matrices and w follows a Dirichlet distribution $\mathrm{Dir}(\alpha_w, \ldots, \alpha_w)$ for some $\alpha_w > 0$. The support of the latter random variable is the whole set of bistochastic matrices.

The reward process. We simply take $(R_k)_{k\geq 0}$ i.i.d. with $R_1 \sim \mathcal{N}(0, \sigma_R^2)$. In particular, the noise does not depend on the state of the Markov Chain. This implies that $\theta^* = 0$.

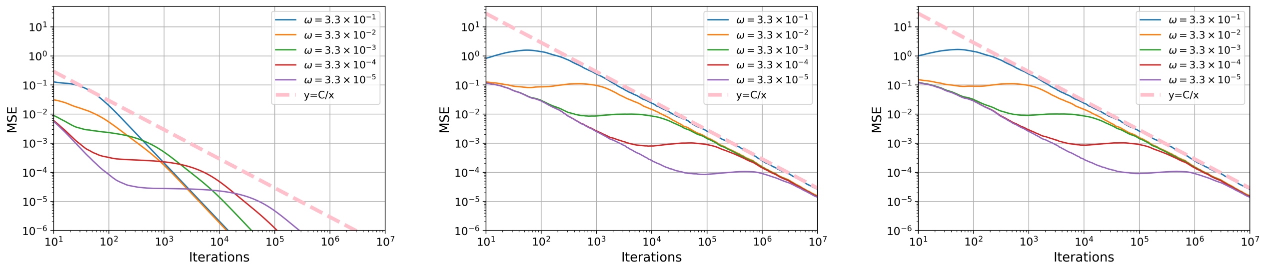
The linear parametrization. We have to take $d \leq n$ in order for the linear independence of the feature maps to be possible. Since the number of states is finite, the feature maps can be represented using the matrix $\Phi := (\varphi_i(j))_{1\leq i\leq d, 1\leq j\leq n}$. Consider its singular value decomposition $\Phi = P D_\varphi Q^\top$, where $P \in \mathbb{R}^{d\times d}, Q \in \mathbb{R}^{n\times n}$ are orthogonal and $D_\varphi = \mathrm{diag}(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^{d\times n}$ with $1 \geq \lambda_1 \geq \cdots \geq \lambda_d > 0$. Observe that, with the above assumption on the transition matrix, we have

$$\Sigma_0 = \Phi D_m \Phi^\top = \frac{1}{n} P D_\varphi D_\varphi^\top P^\top,$$

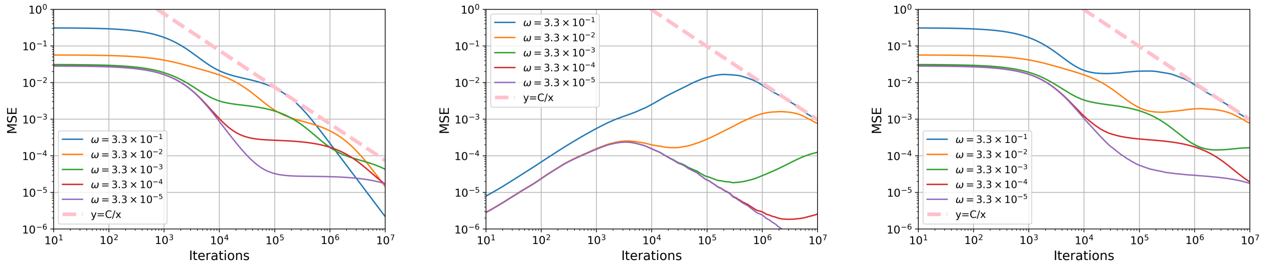
where $D_m = \frac{1}{n} I_n$ is the diagonal matrix with the invariant distribution as diagonal coefficients. Therefore, the eigenvalues of Σ_0 are given by λ_i^2/d for $1 \leq i \leq d$. Moreover, the boundedness condition of Assumption **TD3** holds since $|\varphi(j)|^2 = (Q D_\varphi^\top D_\varphi Q^\top)_{j,j} \leq \lambda_1^2 \leq 1$.

To randomly draw a matrix Σ_0 with fixed spectrum $(\frac{\lambda_i^2}{n})_{1\leq i\leq d}$, it is then sufficient to randomly draw P, Q. We do so by computing the singular value decomposition of a randomly drawn matrix in $\mathbb{R}^{d\times n}$ with i.i.d. coefficients distributed according to $\mathcal{U}(-1, 1)$.

The initial condition θ_0. When we are only interested in the variance part of the error, we take $\theta_0 = 0$. Otherwise, each coordinate of θ_0 is drawn uniformly over $[-1, 1]$.



(a) Bias (b) Variance (c) Total error



(d) Bias (e) Variance (f) Total error

Figure 3: We use similar parameters as in Figure 2 but with $\gamma = 0.9$ and $\alpha = 1$ for the first line and $\gamma = 0.99$ and $\alpha = \frac{1-\gamma}{4} = 2.5 \times 10^{-3}$ for the second line.

Figure 2 with $d = 3$. We take $d = 3$ because it is the smallest dimension for which there exists bistochastic matrices T that are not symmetric. We fix $\lambda_1 = 1$ and vary $\lambda_2 = \lambda_3$ so that $\omega = 5 \times 10^{-j}$ for $1 \leq j \leq 5$. Focusing first on the bias, we observe that each curve exhibits a threshold, seemingly proportional to ω^{-1}, after which the bias decays quadratically in k. Before this threshold, the bias stays below a bound of the form $y = C/x$, with C apparently independent of ω.

For the variance term, it initially increases because too little noise has been observed, then it decreases. It also appears to switch between two envelopes of the form $y = C/x$ at an iteration count proportional to ω^{-1}.

For the total error (which is upper bounded by twice the sum of the bias and the variance terms), the bias dominates at the beginning; once it enters its quadratic decay regime, the variance eventually becomes dominant. The crossover again occurs after a number of iterations proportional to ω^{-1}. Finally, all curves remain bounded by a function C/x with C independent of ω, in agreement with Theorem 3.1.

A.2 Additional numerical simulations

In the first line of Figure 3, we take $\alpha = 1$ so that Theorem 3.3 implies that the convergence is fast but does not ensure the robustness with respect to the smallness of ω. Yet, at least on our numerical simulations, we experience convergence speed that do not degrade when ω is small.

We observe that the bias term is much smaller than the one in Figure 2, but the variance and the total MSE are larger. Therefore, for bounded $|\theta_0 - \theta^*|$, the regime $\frac{1-\gamma}{4}$ may seem better at first sight. However, this statement has to be nuanced with the fact that θ^* can be of order $(1-\gamma)^{-1}$ in practice.

In the second line of Figure 3, we took $\gamma = 0.99$ and observe approximately the same behavior than in Figure 2 with a shift to the right. Observe that the coefficient of the dashed pink line is multiplied by a factor 100 in the three figures, which is aligned with the upper bound from Theorem 3.1.

B Proofs of the results on LSA

B.1 Proof of Proposition 4.1

Proof of Proposition 4.1. Define $z_k = y_k - y^*$, it satisfies

$$z_k = (I_d - \alpha h_k)z_{k-1} - \alpha(h_k y^* - b_k) \ \text{ with } \ z_0 = y_0 - y^*.$$

Observe that $-\alpha(h_k y^* - b_k)$ plays the role of a source term. Therefore, by linearity (or Duhamel's principle), we get that $z_k = w_k + \eta_k$, where w_k and η_k are defined by induction as

$$\begin{aligned} w_k &= (I_d - \alpha h_k)w_{k-1} && \text{with} && w_0 = y_0 - y^*, \\ \eta_k &= (I_d - \alpha h_k)\eta_{k-1} - \alpha(h_k y^* - b_k) && \text{with} && \eta_0 = 0. \end{aligned}$$

In the following, we will prove

$$\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k\right] \le \frac{c_\Sigma |w_0|^2}{\alpha(2-\alpha\beta_1)k} \ \text{ and } \ \mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}_k\right] \le \frac{2\sigma^2}{(2-\alpha\beta c_\Sigma)k}\mathcal{S}_k.$$

so that we will be able to conclude using the inequality $\mathbb{E}\left[\overline{z}_k^\top \Sigma \overline{z}_k\right] \le \left(\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k\right]^{\frac{1}{2}} + \mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}_k\right]^{\frac{1}{2}}\right)^2$.

First step: dealing with the initial condition (corresponding to w).

For $k \ge 1$, using $\mathbb{E}\left[h_k^\top h_k\right] \le \beta_1 S$, observe that

$$\begin{aligned} \mathbb{E}\left[|w_k|^2\right] &= \mathbb{E}\left[w_{k-1}^\top (I_d - \alpha h_k - \alpha h_k^\top + \alpha^2 h_k^\top h_k) w_{k-1}\right] \\ &\le \mathbb{E}\left[w_{k-1}^\top (I_d - 2\alpha S + \alpha^2\beta_1 S) w_{k-1}\right] \\ &\le \mathbb{E}\left[w_{k-1}^\top (I_d - \alpha(2-\alpha\beta_1)S) w_{k-1}\right] \\ &= \mathbb{E}\left[|w_{k-1}|^2\right] - \alpha(2-\alpha\beta_1)\mathbb{E}\left[w_{k-1}^\top S w_{k-1}\right]. \end{aligned}$$

Taking the sum from 1 to k in the latter inequality and using that $u \mapsto u^\top S u$ is convex, we obtain

$$\mathbb{E}\left[\overline{w}_k^\top S \overline{w}_k\right] \le \frac{1}{k}\sum_{i=0}^{k-1}\mathbb{E}\left[w_i^\top S w_i\right] \le \frac{1}{\alpha k}\sum_{i=1}^{k}\left(\mathbb{E}\left[|w_{k-1}|^2\right] - \mathbb{E}\left[|w_k|^2\right]\right) \le \frac{|w_0|^2}{\alpha(2-\alpha\beta_1)k}.$$

This and $\Sigma \le c_\Sigma S$ conclude the part on the bias.

Second step: getting bounds on the uncentered covariances of variance term (corresponding to η).

Using a straightforward induction, we have $\mathbb{E}[\eta_k] = 0$ for $k \ge 0$. We will now prove by induction that

$$\mathbb{E}\left[\eta_k \eta_k^\top\right] \le \frac{\sigma^2 \alpha c_\Sigma}{2-\alpha\beta c_\Sigma} I_d. \tag{6}$$

Recall that $\eta_0 = 0$ so the above inequality holds for $k = 0$. Then, for $k \ge 1$, we assume that it holds at index $k-1$. Since η_{k-1} and (h_k, b_k) are independent, we obtain

$$\mathbb{E}\left[(I_d - \alpha h_k)\eta_{k-1}(h_k y^* - b_k)^\top\right] = \mathbb{E}\left[(I_d - \alpha h_k)\mathbb{E}\left[\eta_{k-1}\right](h_k y^* - b_k)^\top\right] = 0.$$

Using the latter equality, Inequality (10), $\Sigma \le c_\Sigma S$, and the second inequality in Assumption **LS3**, we obtain

$$\begin{aligned} \mathbb{E}\left[\eta_k \eta_k^\top\right] &= \mathbb{E}\left[(I_d - \alpha h_k)\eta_{k-1}\eta_{k-1}^\top (I_d - \alpha h_k^\top)\right] + \alpha^2 \mathbb{E}\left[(h_k y^* - b_k)(h_k y^* - b_k)^\top\right] \\ &\le \frac{\sigma^2 \alpha c_\Sigma}{2-\alpha\beta c_\Sigma}\mathbb{E}\left[(I_d - \alpha h_k)(I_d - \alpha h_k^\top)\right] + \sigma^2\alpha^2\Sigma \\ &\le \frac{\sigma^2 \alpha c_\Sigma}{2-\alpha\beta c_\Sigma}(I_d - \alpha(2-\alpha\beta c_\Sigma)S) + \sigma^2\alpha^2 c_\Sigma S \\ &\le \frac{\sigma^2 \alpha c_\Sigma}{2-\alpha\beta c_\Sigma} I_d. \end{aligned}$$

This concludes the induction and the proof of Inequality (6).

Third step: introduce an appropriate decomposition for η.

Let us then define $(\eta_k^r)_{k\geq 0, 0\leq r\leq k+1}$ and $(\chi_k^r)_{1\leq r\leq k}$ by induction by, for $1 \leq r \leq k$,

$$\begin{aligned}\eta_k^r &= (I_d - \alpha H)\eta_{k-1}^r + \chi_k^r \quad \text{with} \quad \eta_{k-1}^0 = \eta_{k-1}^k = 0,\\ \chi_k^r &= -\alpha(h_k - H)\eta_{k-1}^{r-1} - \alpha(h_k y^* - b_k)\delta_{\{r=1\}}.\end{aligned}$$

Let us check by induction on $k \geq 1$ that $\eta_k = \sum_{r=1}^k \eta_k^r$. The case $k = 1$ is a simple consequence of the above initial condition. Now assume that it holds for $k-1$, for $k \geq 2$, and let us prove it for index k:

$$\begin{aligned}\eta_k &= (I_d - \alpha h_k)\sum_{r=1}^{k-1}\eta_{k-1}^r + \chi_k^1\\ &= (I_d - \alpha H)\sum_{r=1}^{k-1}\eta_{k-1}^r - \alpha(h_k - H)\sum_{r=2}^{k}\eta_{k-1}^{r-1} + \chi_k^1\\ &= \sum_{r=1}^{k}\left((I_d - \alpha H)\eta_{k-1}^r + \chi_k^r\right)\\ &= \sum_{r=1}^{k}\eta_k^r.\end{aligned}$$

In fact, η_k^r can be rewritten as a sum of terms, any of whom admits exactly r different multiplicative centered noises (of the form $h_j - H$ or $h_j y^* - b_j$ for some $1 \leq j \leq k$). Therefore, for $r' \neq r$, any term from the development of $\eta_k^{r'}(\eta_k^r)^\top$ contains at least one multiplicative centered noise that appears only once and is independent with the rest of this same term (i.e., is of the form $R_1(h_j - H)R_2$ or $R_1(h_j y^* - b_j)$ with neither R_1 or R_2 containing h_j or b_j). This implies that $\mathbb{E}\left[\eta_k^{r'}(\eta_k^r)\right] = 0$, so that we get

$$\mathbb{E}\left[\eta_k(\eta_k)^\top\right] = \sum_{r=1}^k\sum_{r'=1}^k \mathbb{E}\left[\eta_k^r(\eta_k^{r'})^\top\right] = \sum_{r=1}^k \mathbb{E}\left[\eta_k^r(\eta_k^r)^\top\right]. \tag{7}$$

Using similar arguments, we also have that, for $\overline{\eta}_k := \frac{1}{k}\sum_{j=0}^{k-1}\eta_k$ and $\overline{\eta}_k^r := \frac{1}{k}\sum_{j=0}^{k-1}\eta_k^r$,

$$\mathbb{E}\left[\overline{\eta}_k^\top\Sigma\overline{\eta}_k\right] = \sum_{r=1}^k \mathbb{E}\left[(\overline{\eta}_k^r)^\top\Sigma\overline{\eta}_k^r\right]. \tag{8}$$

Fourth step: obtaining the convergence rate.

For $1 \leq r \leq k$, the induction relation of η_k^r implies

$$\eta_k^r = \sum_{j=1}^k (I_d - \alpha H)^{k-j}\chi_j^r.$$

For $k \geq 2$, define $\overline{\eta}_k^r = \frac{1}{k}\sum_{i=1}^{k-1}\eta_k^r$, it satisfies

$$\overline{\eta}_k^r = \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{i}(I_d - \alpha H)^{i-j}\chi_j^r = \frac{1}{k}\sum_{j=1}^{k-1}\sum_{i=j}^{k-1}(I_d - \alpha H)^{i-j}\chi_j^r = \frac{1}{\alpha k}\sum_{j=1}^{k-1}M_{k-j}H^{-1}\chi_j^r = \frac{1}{\alpha k}\sum_{j=0}^{k-1}M_j H^{-1}\chi_{k-j}^r.$$

where $M_j = I_d - (I_d - \alpha H)^j$. Observe that M_j commutes with H and H^{-1} (but not with $H^\top$ and $H^{-\top}$). On the one hand, using Inequality (10), we get

$$\begin{aligned}
\mathbb{E}\left[(\overline{\eta}_k^1)^\top \Sigma \overline{\eta}_k^1\right] &= \frac{1}{\alpha^2 k^2} \sum_{j=0}^{k-1} \mathbb{E}\left[(\chi_{k-j}^1)^\top H^{-\top} M_j^\top \Sigma M_j H^{-1} \chi_{k-j}^1\right] \\
&= \frac{1}{\alpha^2 k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[\chi_{k-j}^1 (\chi_{k-j}^1)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&= \frac{1}{k^2} \sum_{j=1}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} y^* - b_{k-j})(h_{k-j} y^* - b_{k-j})^\top)\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&\leq \frac{\sigma^2}{k^2} \sum_{j=1}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \Sigma H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
\sum_{r=2}^{k} \mathbb{E}\left[(\overline{\eta}_k^r)^\top \Sigma \overline{\eta}_k^r\right] &= \frac{1}{\alpha^2 k^2} \sum_{r=2}^{k} \sum_{j=0}^{k-1} \mathbb{E}\left[(\chi_{k-j}^r)^\top H^{-\top} M_j^\top \Sigma M_j H^{-1} \chi_{k-j}^r\right] \\
&= \frac{1}{\alpha^2 k^2} \sum_{r=2}^{k} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[\chi_{k-j}^r \otimes \chi_{k-j}^r\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&= \frac{1}{k^2} \sum_{r=2}^{k} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H) \eta_{k-j-1}^{r-1} (\eta_{k-j-1}^{r-1})^\top (h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&= \frac{1}{k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H) \sum_{r=2}^{k} \mathbb{E}\left[\eta_{k-j-1}^{r-1} (\eta_{k-j-1}^{r-1})^\top\right] (h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&= \frac{1}{k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H) \mathbb{E}\left[\eta_{k-j-1} (\eta_{k-j-1})^\top\right] (h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&\leq \frac{\sigma^2 \alpha c_\Sigma}{(2 - \alpha\beta c_\Sigma) k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H)(h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&\leq \frac{\sigma^2 \alpha\beta c_\Sigma}{(2 - \alpha\beta c_\Sigma) k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \Sigma H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right),
\end{aligned}$$

where we used Equality (7) to get the fifth line, the second step of this proof to get the sixth line and the first inequality of Assumption **LS3** to obtain the last one. The latter two chains of inequalities and Equality (8) imply

$$\begin{aligned}
\mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}_k\right] &= \sum_{r=1}^{k} \mathbb{E}\left[(\overline{\eta}_k^r)^\top \Sigma \overline{\eta}_k^r\right] \\
&\leq \frac{\sigma^2}{k^2} \left(1 + \frac{\alpha\beta c_\Sigma}{2 - \alpha\beta c_\Sigma}\right) \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \Sigma H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&= \frac{2\sigma^2}{(2 - \alpha\beta c_\Sigma) k^2} \sum_{j=0}^{k-1} \left\| \Sigma^{\frac{1}{2}} \left(I_d - (I_d - \alpha H)^j\right) H^{-1} \Sigma^{\frac{1}{2}} \right\|_F^2 \\
&= \frac{2\sigma^2}{(2 - \alpha\beta c_\Sigma) k^2} \sum_{j=0}^{k-1} \left\| \left(I_d - (I_d - \alpha \Sigma^{\frac{1}{2}} H \Sigma^{-\frac{1}{2}})^j\right) \Sigma^{\frac{1}{2}} H^{-1} \Sigma^{\frac{1}{2}} \right\|_F^2 \\
&= \frac{2\sigma^2}{(2 - \alpha\beta c_\Sigma) k} \mathcal{S}_k.
\end{aligned}$$

To conclude, it only remains to prove that $\mathcal{S}_k \le c_\Sigma^2(3d+\widetilde{\mathcal{S}}_k)$, which is a consequence of the following computations,

$$\Sigma^{\frac{1}{2}}H^{-1}\Sigma H^{-\top}\Sigma^{\frac{1}{2}} \le c_\Sigma \Sigma^{\frac{1}{2}}H^{-1}SH^{-\top}\Sigma^{\frac{1}{2}} = c_\Sigma\Sigma^{\frac{1}{2}}\frac{H^{-1}+H^{-\top}}{2}\Sigma^{\frac{1}{2}} = c_\Sigma\Sigma^{\frac{1}{2}}S^{-1}\Sigma^{\frac{1}{2}} \le c_\Sigma^2 I_d,$$

where we used $\Sigma \le c_\Sigma S$, Heinze's inequality $H^{-1}+H^{-\top} \le 2S^{-1}$ and $S^{-1} \le c_\Sigma \Sigma^{-1}$; and

$$\left\| \left(I_d - Q^j\right)\Sigma^{\frac{1}{2}}H^{-1}\Sigma^{\frac{1}{2}} \right\|_F^2 \le c_\Sigma^2 \left\|I_d - Q^j\right\|_F^2 = c_\Sigma^2\left(\operatorname{tr}(I_d) - 2\operatorname{tr}(Q^j) + \left\|Q^j\right\|_F^2\right) \le c_\Sigma^2(3d + \left\|Q^j\right\|_F^2),$$

where we used $|\operatorname{tr}(Q^j)| = |\operatorname{tr}((I_d-\alpha H)^j)| \le d\|I_d - \alpha H\|_{\text{op}}^j \le d$. This concludes the proof. □

B.2 Proof of Theorem 4.2

Proof of Theorem 4.2. This is a consequence of Proposition 4.1 and Lemma B.1 below. □

Lemma B.1. *Under Assumptions* **LS1**-**LS3***, for* $0 < \alpha < \frac{2}{\beta_1}$*, we have*

$$\mathcal{S}_k \le c_\Sigma^2 d\left(3 + \frac{1}{\alpha(2-\alpha\beta_1)\mu k}\right).$$

Under Assumptions **LS1**,**LS2** *and* **LS3** *without the last inequality, for* $0 < \alpha < \frac{2}{\beta c_\Sigma}$*, we have*

$$\mathcal{S}_k \le c_\Sigma^2 d\left(3 + \frac{1}{\alpha(2-\alpha\beta c_\Sigma)\mu k}\right).$$

Proof. Let us prove the first inequality. Using $M_j = I_d - (I_d - \alpha H)^j$ and $N_j = (I_d - \alpha H)^j$ for $j \ge 1$, we have for $k \ge 1$:

$$\begin{aligned}
\mathcal{S}_k &= \frac{1}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(\Sigma^{\frac{1}{2}}M_jH^{-1}\Sigma H^{-\top}M_j^\top\Sigma^{\frac{1}{2}}\right)\\
&\le \frac{c_\Sigma}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(\Sigma^{\frac{1}{2}}M_jH^{-1}SH^{-\top}M_j^\top\Sigma^{\frac{1}{2}}\right)\\
&\le \frac{c_\Sigma}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(\Sigma^{\frac{1}{2}}M_j\frac{H^{-1}+H^{-\top}}{2}M_j^\top\Sigma^{\frac{1}{2}}\right)\\
&\le \frac{c_\Sigma}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(\Sigma^{\frac{1}{2}}M_jS^{-1}M_j^\top\Sigma^{\frac{1}{2}}\right)\\
&= \frac{c_\Sigma}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(S^{-\frac{1}{2}}M_j^\top\Sigma M_jS^{-\frac{1}{2}}\right)\\
&\le \frac{c_\Sigma^2}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(S^{-\frac{1}{2}}M_j^\top SM_jS^{-\frac{1}{2}}\right)\\
&= \frac{c_\Sigma^2}{k}\sum_{j=1}^{k-1}\operatorname{tr}\left(I_d + 2\operatorname{tr}(N_j) + S^{-\frac{1}{2}}N_j^\top SN_jS^{-\frac{1}{2}}\right)\\
&\le 3c_\Sigma^2 d + \frac{c_\Sigma^2}{k}\operatorname{tr}\left(S^{-\frac{1}{2}}\left(\sum_{j=1}^{k-1}N_j^\top SN_j\right)S^{-\frac{1}{2}}\right)\\
&\le c_\Sigma^2\left(3d + \frac{\operatorname{tr}(S^{-1})}{\alpha(2-\alpha\beta_1)k}\right)\\
&\le c_\Sigma^2 d\left(3 + \frac{1}{\alpha(2-\alpha\beta_1)\mu k}\right),
\end{aligned}$$

where we used $\Sigma \leq c_\Sigma S$ to get the second and sixth lines, Heinze's inequality (Lemma B.6) to get the fourth one, $\operatorname{tr}(N_j) \leq \operatorname{tr}(I_d)^{\frac{1}{2}} \operatorname{tr}(N_j^\top N_j)^{\frac{1}{2}} \leq \operatorname{tr}(I_d) = d$ (since $(I_d - \alpha H)^\top (I_d - \alpha H) \leq I_d$) and Inequality (12) to get the eighth one, and $S^{-1} \leq \mu^{-1} I_d$ to get the last one.

To get the second inequality, we write

$$\mathcal{S}_k = \frac{1}{k} \sum_{j=1}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j^\top H^{-\top} \Sigma H^{-1} M_j \Sigma^{\frac{1}{2}} \right),$$

using that H and M_j commutes. Then, we repeat similar arguments as the ones for the first inequality. This concludes the proof. □

B.3 Extensions of Theorem 4.2

Here, we present two extensions of Theorem 4.2. The first one, Theorem B.2, allows for learning steps up to $\alpha < \frac{2}{c_\Sigma \beta}$ and keep a $O\left(\frac{1}{k} + \frac{1}{k^2 \mu}\right)$ like Theorem 4.2.

The second extension is Theorem B.3 below that hold for $\alpha < \frac{2}{\beta_1}$, but admits a slower convergence rate of $O\left(\frac{1}{k\mu}\right)$. Still this rates remains better than any existing rate on LSA with similar assumptions prior to this work.

Theorem B.2. *Assume* **LS1**, **LS2** *and* **LS3** *without the third inequality. For* $0 < \alpha < \frac{2}{\beta c_\Sigma}$*, we have*

$$\mathbb{E}\left[(\overline{y}_k^\top - y^*)^\top \Sigma (\overline{y}_k - y^*) \right] \leq \frac{c_\Sigma |y_0 - y^*|^2}{(2 - \alpha\beta c_k)\alpha k} + \left(\left(\frac{\beta |y_0 - y^*|^2}{k} \right)^{\frac{1}{2}} + \left(\frac{2\sigma^2}{(2 - \alpha\beta c_\Sigma) k} \right)^{\frac{1}{2}} \right)^2 \left(3 + \frac{1}{\alpha(2 - \alpha\beta c_\Sigma)\mu k} \right).$$

Proof of Theorem B.2. It is a direct consequence of Proposition B.4 and Lemma B.1. □

Theorem B.3. *Assume* **LS1**-**LS3**. *Define* $\widetilde{\sigma}^2 = \mathbb{E}[|h_k y^* - b_k|^2]$. *For* $0 < \alpha < \frac{2}{\beta_1}$*, we have*

$$\mathbb{E}\left[(\overline{y}_k^\top - y^*)^\top \Sigma (\overline{y}_k - y^*) \right] \leq \frac{2 c_\Sigma |y_0 - y^*|^2}{\alpha(2 - \alpha\beta_1) k} + \frac{2 c_\Sigma^2 d}{k} \left(\sigma^2 + \frac{\alpha\beta\widetilde{\sigma}^2}{(2 - \alpha\beta_1)\mu} \right) \left(3 + \frac{1}{\alpha(2 - \alpha\beta_1)\mu k} \right).$$

Proof of Theorem B.3. Let us use $z_k = y_k - y^*$ and the decomposition $z_k = w_k + \eta_k$, similarly as defined in the proof of Proposition 4.1.

Repeating the first step of the proof of Proposition 4.1, we obtain

$$\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k \right] \leq \frac{c_\Sigma |w_0|^2}{\alpha(2 - \alpha\beta_1) k}.$$

Therefore, using inequality $\mathbb{E}\left[\overline{z}_k^\top \Sigma \overline{z}_k \right] \leq 2\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k \right] + 2\mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}_k \right]$, to conclude it only remains to prove that

$$\mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}_k \right] \leq \frac{2\sigma^2}{(2 - \alpha\beta c_\Sigma) k} \mathcal{S}_k.$$

First step: getting an upper bound on $\mathbb{E}[|\eta_k|^2]$

Recall that, using a straightforward induction, we have $\mathbb{E}[\eta_k] = 0$ for $k \geq 0$. We will now prove that

$$\mathbb{E}\left[|\eta_k|^2 \right] \leq \frac{\alpha \widetilde{\sigma}^2}{(2 - \alpha\beta_1)\mu}. \tag{9}$$

For $k = 0$, the result holds since $\eta_0 = 0$. Then, for $k \geq 1$, we assume that it holds for index $k - 1$. Since η_{k-1} and (h_k, b_k) are independent, we obtain

$$\mathbb{E}\left[\eta_{k-1}^\top (I_d - \alpha h_k)^\top (h_k y^* - b_k) \right] = \mathbb{E}\left[\eta_{k-1}^\top \right] \mathbb{E}\left[(I_d - \alpha h_k)^\top (h_k y^* - b_k) \right] = 0.$$

This and inequality (11) imply that

$$\begin{aligned}
\mathbb{E}\left[|\eta_k|^2\right] &= \mathbb{E}\left[|(I_d - \alpha h_k)\eta_{k-1}|^2\right] + \alpha^2 \mathbb{E}\left[|h_k y^* - b_k|^2\right] \\
&\leq \mathbb{E}\left[\eta_{k-1}^\top (I_d - \alpha(2-\alpha\beta_1)S)\eta_{k-1}\right] + \alpha^2 \widetilde{\sigma}^2 \\
&\leq (1 - \alpha(2-\alpha\beta_1)\mu)\mathbb{E}\left[|\eta_{k-1}|^2\right] + \alpha^2\widetilde{\sigma}^2 \\
&\leq \frac{\alpha\widetilde{\sigma}^2}{(2-\alpha\beta_1)\mu},
\end{aligned}$$

which concludes the induction.

Second step: obtaining the convergence rate.

Let us use the same decomposition $\eta_k = \sum_{r=1}^k \eta_k^r$ as introduced in the proof of Proposition 4.1. Similarly as in the third step of the latter proof, we get

$$\mathbb{E}\left[(\overline{\eta}_k^1)^\top \Sigma \overline{\eta}_k^1\right] \leq \frac{\sigma^2}{k^2} \sum_{j=1}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \Sigma H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) = \frac{\sigma^2}{k} \mathcal{S}_k,$$

and

$$\begin{aligned}
\sum_{r=2}^{k} \mathbb{E}\left[(\overline{\eta}_k^r)^\top \Sigma \overline{\eta}_k^r\right] &= \frac{1}{k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H)\mathbb{E}\left[\eta_{k-j-1}(\eta_{k-j-1})^\top\right](h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&\leq \frac{\alpha\widetilde{\sigma}^2}{(2-\alpha\beta_1)\mu k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H)(h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&\leq \frac{\alpha\beta\widetilde{\sigma}^2}{(2-\alpha\beta_1)\mu k^2} \sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \Sigma H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\
&= \frac{\alpha\beta\widetilde{\sigma}^2}{(2-\alpha\beta_1)\mu k} \mathcal{S}_k,
\end{aligned}$$

where we used $\mathbb{E}\left[\eta_j(\eta_j)^\top\right] \leq \mathbb{E}\left[|\eta_j|^2\right] I_d \leq \frac{\alpha\widetilde{\sigma}^2}{(2-\alpha\beta_1)\mu k^2} I_d$ to get the second line, and the first inequality of Assumption **LS3** for the third line.

Using the above inequalities and Lemma B.1, we obtain

$$\mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}\right] \leq \left(\frac{\sigma^2}{k} + \frac{\alpha\beta\widetilde{\sigma}^2}{(2-\alpha\beta_1)\mu k}\right) c_\Sigma^2 d \left(3 + \frac{1}{\alpha(2-\alpha\beta_1)\mu k}\right),$$

which concludes the proof.

□

The following propositions is an extension of Proposition 4.1 to learning rates up to $\alpha < \frac{2}{\beta c_\Sigma}$.

Proposition B.4. *Under Assumptions* **LS1**-**LS3** *without the third inequality, for* $0 < \alpha < \frac{2}{c_\Sigma \beta}$*, we have*

$$\mathbb{E}\left[(\overline{y}_k^\top - y^*)^\top \Sigma (\overline{y}_k - y^*)\right] \leq \frac{2|y_0 - y^*|^2 c_\Sigma d}{\alpha k} + \left(\left(\frac{\beta|y_0 - y^*|^2}{k}\right)^{\frac{1}{2}} + \left(\frac{2\sigma^2}{(2-\alpha\beta c_\Sigma)k}\right)^{\frac{1}{2}}\right)^2 \mathcal{S}_k.$$

where $\mathcal{S}_k$ *is defined as in Proposition 4.1.*

Proof of Proposition B.4. The three last point of the proof of Proposition 4.1 holds when removing the third inequality in **LS3**. Therefore, to prove Proposition B.4, it is sufficient to prove that

$$\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k\right] \leq \frac{c_\Sigma |y_0 - y^*|^2}{(2-\alpha\beta c_\Sigma)\alpha k} + \frac{\beta|y_0 - y^*|^2}{k}\mathcal{S}_k.$$

Similarly as in the third step of the proof of Proposition 4.1, we define $(w_k^r)_{k,r\geq 0}$ such that $w_k = \sum_{r=0}^k w_k^r$ and, for $r \geq 0$ and $k \geq 1$,

$$\begin{aligned} w_k^r &= (I_d - \alpha H) w_{k-1}^r + \chi_k^r \qquad \text{with} \qquad w_0^r = w_0 \mathbb{1}_{\{r=0\}}, \\ \chi_k^r &= -\alpha(h_k - H)\eta_{k-1}^{r-1}\mathbb{1}_{\{r\neq 0\}}. \end{aligned}$$

First step: getting a bound on $(\overline{w}_k^0)^\top S \overline{w}_k^0$.

Take $\widetilde{\alpha} = \frac{2}{\beta c_\Sigma}$, (10) implies that $(I_d - \widetilde{\alpha} H)(I_d - \widetilde{\alpha} H)^\top \leq I_d$. This implies that the operator norm of $(I_d - \widetilde{\alpha} H)$ is upper bounded by one and that

$$(I_d - \widetilde{\alpha} H)^\top (I_d - \widetilde{\alpha} H) \leq I_d.$$

This then implies

$$\begin{aligned} (I_d - \alpha H)^\top (I_d - \alpha H) &= \frac{\alpha^2}{\widetilde{\alpha}^2}(I_d - \widetilde{\alpha} H)^\top (I_d - \widetilde{\alpha} H) + \left(1 - \frac{\alpha^2}{\widetilde{\alpha}^2}\right) I_d - 2\left(\alpha - \frac{\alpha^2}{\widetilde{\alpha}}\right) S \\ &\leq I_d - \alpha(2 - \alpha\beta c_\Sigma) S. \end{aligned}$$

As a consequence, for $k \geq 1$, we obtain

$$\begin{aligned} |w_k^0|^2 &= (w_{k-1}^0)^\top (I_d - \alpha H)^\top (I_d - \alpha H) w_{k-1}^0 \\ &\leq |w_{k-1}^0|^2 - \alpha(2 - \alpha\beta c_\Sigma)(w_{k-1}^0)^\top S w_{k-1}^0 \end{aligned}$$

Taking the sum from 1 to k in the latter inequality and using that $u \mapsto u^\top S u$ is convex, we obtain

$$(\overline{w}_k^0)^\top S \overline{w}_k^0 \leq \frac{1}{k}\sum_{i=0}^{k-1} (w_i^0)^\top S w_i^0 \leq \frac{1}{\alpha k}\sum_{i=1}^{k}\left(|w_{k-1}^0|^2 - |w_k^0|^2\right) \leq \frac{|w_0|^2}{\alpha(2 - \alpha\beta c_\Sigma)k}.$$

This and $\Sigma \leq c_\Sigma S$ yield

$$(\overline{w}_k^0)^\top \Sigma \overline{w}_k^0 \leq \frac{c_\Sigma |w_0|^2}{\alpha(2 - \alpha\beta c_\Sigma)k}.$$

Second step: getting the bound on the bias term.

Then, using similar arguments as in the third step of the proof of Proposition 4.1, we have

$$\begin{aligned} \sum_{r=1}^k \mathbb{E}\left[(\overline{w}_k^r)^\top \Sigma \overline{w}_k^r\right] &= \frac{1}{k^2}\sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H)\mathbb{E}\left[w_{k-j-1}(w_{k-j-1})^\top\right](h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\ &\leq \frac{|y_0 - y^*|^2}{k^2}\sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \mathbb{E}\left[(h_{k-j} - H)(h_{k-j} - H)^\top\right] H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\ &\leq \frac{\beta |y_0 - y^*|^2}{k^2}\sum_{j=0}^{k-1} \operatorname{tr}\left(\Sigma^{\frac{1}{2}} M_j H^{-1} \Sigma H^{-\top} M_j^\top \Sigma^{\frac{1}{2}}\right) \\ &= \frac{\beta |y_0 - y^*|^2}{k}\mathcal{S}_k. \end{aligned}$$

In the end, we obtain

$$\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k\right] = \sum_{r=0}^k \mathbb{E}\left[(\overline{w}_k^r)^\top \Sigma \overline{w}_k^r\right] \leq \frac{c_\Sigma |w_0|^2}{\alpha(2 - \alpha\beta c_\Sigma)k} + \frac{\beta |y_0 - y^*|^2}{k}\mathcal{S}_k.$$

Third step: deal with the cross term and conclude.

We have

$$\mathbb{E}\left[\overline{z}_k^\top \Sigma \overline{z}_k\right] = \mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{w}_k\right] + \mathbb{E}\left[\overline{\eta}_k^\top \Sigma \overline{\eta}_k\right] + 2\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{\eta}_k\right],$$

where the last term is the cross term and satisfies

$$\begin{aligned}
\mathbb{E}\left[\overline{w}_k^\top \Sigma \overline{\eta}_k\right] &= \mathbb{E}\left[\left(\sum_{r=0}^{k} \overline{w}_k^r\right)^\top \Sigma \left(\sum_{r=1}^{k} \overline{\eta}_k^r\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{r=1}^{k} \overline{w}_k^r\right)^\top \Sigma \left(\sum_{r=1}^{k} \overline{\eta}_k^r\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{r=1}^{k} \overline{w}_k^r\right)^\top \Sigma \left(\sum_{r=1}^{k} \overline{w}_k^r\right)\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\sum_{r=1}^{k} \overline{\eta}_k^r\right)^\top \Sigma \left(\sum_{r=1}^{k} \overline{\eta}_k^r\right)\right]^{\frac{1}{2}} \\
&\le \left(\frac{\beta |y_0 - y^*|^2}{k} \mathcal{S}_k\right)^{\frac{1}{2}} \left(\frac{2\sigma^2 \mathcal{S}_k}{(2-\alpha\beta c_\Sigma)k}\right)^{\frac{1}{2}}.
\end{aligned}$$

The above inequalities imply the upper bound in Proposition B.4. □

B.4 Some results from linear algebra

Lemma B.5. *Under assumptions* **LS1-LS3***, for* $\alpha \le \frac{2}{\beta c_\Sigma}$*, we have*

$$(I_d - \alpha H)(I_d - \alpha H)^\top \le \mathbb{E}\left[(I_d - \alpha h)(I_d - \alpha h)^\top\right] \le I_d - \alpha(2 - \alpha\beta c_\Sigma)S, \tag{10}$$

Moreover, for $\alpha < \frac{2}{\beta_1}$*, we get*

$$(I_d - \alpha H)^\top (I_d - \alpha H) \le \mathbb{E}\left[(I_d - \alpha h)^\top (I_d - \alpha h)\right] \le I_d - \alpha(2 - \alpha\beta_1)S, \tag{11}$$

$$\sum_{i=0}^{k-1} (I_d - \alpha H^\top)^i S (I_d - \alpha H)^i \le \frac{1}{\alpha(2-\alpha\beta_1)} I_d. \tag{12}$$

Proof. The first inequality in (10) is straightforward. Then, we have,

$$\mathbb{E}\left[(I_d - \alpha h)(I_d - \alpha h)^\top\right] = I_d - 2\alpha S + \alpha^2 \mathbb{E}[hh^\top] \le I_d - \alpha(2 - \alpha\beta c_\Sigma)S,$$

using $\mathbb{E}[hh^\top] \le \beta\Sigma \le \beta c_\Sigma S$. Inequality (11) can be proved with a similar computation.

Then, using the notation $N_i = (I_d - \alpha H)^i$, we have

$$\begin{aligned}
I_d \ge I_d - N_k^\top N_k &= \sum_{i=0}^{k-1} N_i^\top N_i - N_{i+1}^\top N_{i+1} \\
&= \sum_{i=0}^{k-1} N_i^\top N_i - N_i^\top (I_d - \alpha H^\top)(I_d - \alpha H) N_i \\
&= \sum_{i=0}^{k-1} N_i^\top (I_d - (I_d - \alpha H^\top)(I_d - \alpha H)) N_i \\
&= \sum_{i=0}^{k-1} N_i^\top (2\alpha S - \alpha^2 H^\top H) N_i \\
&\ge \alpha(2 - \alpha\beta_1) \sum_{i=0}^{k-1} N_i^\top S N_i,
\end{aligned}$$

where we used $\alpha H^\top H \le (2 - \alpha\beta_1)S$ to get the last line. □

Lemma B.6 (Heinze's inequality)**.** *Let* $H \in \mathbb{R}^{d\times d}$ *be such that* $S := \frac{H+H^\top}{2} > 0$*, then* $H^{-1} + H^{-\top}$ *is positive definite and* $H^{-1} + H^{-\top} \le 2S^{-1}$*. Moreover, we have* $\operatorname{tr}(SH^{-\top}) \le d$*.*

Proof. First, let us prove that $H^{-1}+H^{-\top}$ is positive definite. Take $x \in \mathbb{R}^d \backslash \{0\}$, define $y = H^{-1}x$ we have

$$x^\top (H^{-1}+H^{-\top})x = 2x^\top H^{-\top} x = 2y^\top H y = y^\top (H + H^\top) y > 0.$$

Then, let us prove that $\mathrm{tr}(SH^{-\top}) \leq d$ is a consequence of $H^{-1}+H^{-\top} \leq 2S^{-1}$:

$$\mathrm{tr}(SH^{-\top}) = \frac{1}{2}\mathrm{tr}(S^{\frac{1}{2}}(H^{-1}+H^{-\top})S^{\frac{1}{2}}) \leq \mathrm{tr}(S^{\frac{1}{2}}S^{-1}S^{\frac{1}{2}}) = d.$$

Now, observe that

$$\begin{aligned}\left(S^{\frac{1}{2}}(H^{-1}+H^{-\top})S^{\frac{1}{2}}\right)^2 &= 2S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}} + 2S^{\frac{1}{2}}H^{-\top}SH^{-1}S^{\frac{1}{2}} \\ &\qquad - \left(S^{\frac{1}{2}}(H^{-1}-H^{-\top})S^{\frac{1}{2}}\right)\left(S^{\frac{1}{2}}(H^{-1}-H^{-\top})S^{\frac{1}{2}}\right)^\top \\ &\leq 2S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}} + 2S^{\frac{1}{2}}H^{-\top}SH^{-1}S^{\frac{1}{2}}.\end{aligned}$$

Therefore, to conclude, it is sufficient to prove $S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}} \leq I_d$ and $S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}} \leq I_d$. For the former inequality, invert the matrix and use the notation $A = \frac{H-H^\top}{2}$, we get

$$\begin{aligned}(S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}})^{-1} &= S^{-\frac{1}{2}}H^\top S^{-1}HS^{-\frac{1}{2}} \\ &= S^{-\frac{1}{2}}(S-A)S^{-1}(S+A)S^{-\frac{1}{2}} \\ &= I_d - S^{-\frac{1}{2}}AS^{-1}AS^{-\frac{1}{2}} \\ &= I_d + \left(S^{-\frac{1}{2}}AS^{-\frac{1}{2}}\right)\left(S^{-\frac{1}{2}}AS^{-\frac{1}{2}}\right)^\top \\ &\geq I_d,\end{aligned}$$

which, indeed, implies $S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}} \leq I_d$. Then, applying the later inequality to $\widetilde{H} = H^\top$, we obtain $S^{\frac{1}{2}}H^{-1}SH^{-\top}S^{\frac{1}{2}} \leq I_d$. This concludes the proof. $\square$

C Proofs specific to TD(0)

C.1 Proof of Proposition 5.3

Proof. For $k \geq 2$ fixed and take $c_k \in (0,1)$ such that $c_k^2 = 1 - \frac{1}{k}$. In particular, observe that

$$c_k^{-2(k-1)} = \left(\frac{k}{k-1}\right)^{k-1} = \left(1 + \frac{1}{k-1}\right)^{k-1} \leq e.$$

Define $\widetilde{Q} = c_k Q$, we have

$$c_k^{2(k-1)} \sum_{\ell=0}^{k-1} (I_d - Q^{\ell\top})(I_d - Q^\ell) \leq \sum_{\ell=0}^{k-1} c_k^{2\ell}(I_d - Q^{\ell\top})(I_d - Q^\ell)$$

Then, Lemma C.2 yields

$$\begin{aligned}&\sum_{\ell=0}^{k-1}(I_d - Q^{\ell\top})(I_d - Q^\ell) \\ &\qquad \leq \frac{1}{2i\pi c_k^{2(k-1)}} \int_{|z|=1} \overline{z}\left((\overline{z}-c_k)^{-1}I_d - (\overline{z}I_d - \widetilde{Q})^{-\top}\right)\left((z-c_k)^{-1}I_d - (zI_d - \widetilde{Q})^{-1}\right) dz \\ &\qquad = \frac{1}{2\pi c_k^{2(k-1)}} \int_{-\pi}^{\pi} \left((e^{-i\psi}-c_k)^{-1}I_d - (e^{-i\psi}I_d - \widetilde{Q})^{-\top}\right)\left((e^{i\psi}-c_k)^{-1}I_d - (e^{i\psi}I_d - \widetilde{Q})^{-1}\right) d\psi \\ &\qquad \leq \frac{e}{2\pi} \int_{-\pi}^{\pi} \left((e^{-i\psi}-c_k)^{-1}I_d - (e^{-i\psi}I_d - \widetilde{Q})^{-\top}\right)\left((e^{i\psi}-c_k)^{-1}I_d - (e^{i\psi}I_d - \widetilde{Q})^{-1}\right) d\psi\end{aligned}$$

In the following, we will use both the notation z or $e^{i\psi}$, depending on the situation. From the definition of $\mathcal{S}_k$ and the fact that $\Sigma_0^{\frac{1}{2}} H^{-1} \Sigma_0^{\frac{1}{2}} = (I_d - \gamma U)^{-1}$ This yields the following upper bound on $\mathcal{S}_k$

$$\mathcal{S}_k \leq \frac{e}{2\pi k} \int_{-\pi}^{\pi} \left\| \left((e^{i\psi} - c_k)^{-1} I_d - (e^{i\psi} I_d - \widetilde{Q})^{-1} \right) (I_d - \gamma U)^{-1} \right\|_F^2 d\psi. \tag{13}$$

We are going to separate the latter integral into three parts corresponding to:

(i) $\mathcal{R}(z) \geq 1 - (1-\gamma)^2$, i.e., $|\psi| \leq \psi_\gamma$

(ii) $|\mathcal{R}(z)| < 1 - (1-\gamma)^2$, i.e., $\psi_\gamma < |\psi| < \pi - \psi_\gamma$

(iii) $\mathcal{R}(z) \leq -1 + (1-\gamma)^2$, i.e., $|\psi| \geq \pi - \psi_\gamma$

where $\mathcal{R}(z)$ is the real part of z and ψ_γ is defined by $\psi_\gamma = \arccos(1 - (1-\gamma)^2)$.

First step: dealing with case (i).

Observe that

$$\begin{aligned} z - \widetilde{Q} &= (z - c_k) I_d + c_k \alpha \Sigma_0 (I_d - \gamma U) \\ &= (z - c_k) \left(I_d + \frac{c_k \alpha}{|z - c_k|} \frac{\overline{z} - c_k}{|z - c_k|} \Sigma_0 (I_d - \gamma U) \right) \\ &= (z - c_k) \left(I_d + \widetilde{z} \Sigma (I_d - \gamma U) \right), \end{aligned}$$

with $\widetilde{z} = \frac{\overline{z} - c_k}{|z - c_k|}$ and $\Sigma = \frac{c_k \alpha}{|z - c_k|} \Sigma_0$. Using Lemma C.1, for any $x \in \mathbb{C}^d$, we have

$$\begin{aligned} \left| \left((z - c_k)^{-1} I_d - (z - \widetilde{Q})^{-1} \right) (I_d - \gamma U)^{-1} x \right| &= |z - c_k|^{-1} \left| \left(I_d - (I_d + \widetilde{z} \Sigma (I_d - \gamma U))^{-1} \right) (I_d - \gamma U)^{-1} x \right| \\ &\leq \frac{2}{1 - \gamma^2} |z - c_k|^{-1}, \end{aligned}$$

which holds only if $-\mathcal{R}(\widetilde{z}) = -\mathcal{R}(\frac{\overline{z} - c_k}{|z - c_k|}) \leq \frac{1-\gamma}{\sqrt{2}}$, that we prove as follows,

$$\begin{aligned} -\mathcal{R}\left(\frac{\overline{z} - c_k}{|z - c_k|} \right) = \frac{c_k - \mathcal{R}(z)}{\sqrt{(c_k - \mathcal{R}(z))^2 + \mathcal{I}(z)^2}} &\leq \frac{1 - \mathcal{R}(z)}{\sqrt{(1 - \mathcal{R}(z))^2 + \mathcal{I}(z)^2}} \\ &= \frac{1 - \mathcal{R}(z)}{\sqrt{(1 - \mathcal{R}(z))^2 + 1 - \mathcal{R}(z)^2}} \\ &= \frac{1 - \mathcal{R}(z)}{\sqrt{2 - 2\mathcal{R}(z)}} \\ &= \sqrt{\frac{1 - \mathcal{R}(z)}{2}} \\ &\leq \sqrt{\frac{1 - (1 - (1-\gamma)^2)}{2}} = \frac{1-\gamma}{\sqrt{2}}, \end{aligned}$$

where the first line is due to $u \mapsto \frac{u}{u^2 + v^2}$ is non-decreasing for any $v \in \mathbb{R}$, and the last one comes from $\mathcal{R}(z) \geq 1 - (1-\gamma)^2$. Let us integrate for $\psi \in [-\psi_\gamma, \psi_\gamma]$ and use $\| \cdot \|_F^2 \leq d \| \cdot \|_{\text{op}}^2$,

$$\begin{aligned} \frac{e}{2\pi} \int_{-\pi}^{\pi} \left\| \left((e^{i\psi} - c_k)^{-1} I_d - (e^{i\psi} I_d - \widetilde{Q})^{-1} \right) (I_d - \gamma U)^{-1} \right\|_F^2 d\psi &\leq \frac{e}{2\pi} \int_{-\psi_k}^{\psi_k} \left\| \frac{(2-\gamma)^2 I_d}{(1-\gamma)^2 |e^{i\psi} - c_k|^2} \right\|_F^2 d\psi \\ &\leq \frac{de(2-\gamma)^2}{2\pi(1-\gamma)^2} \int_{-\pi}^{\pi} \frac{1}{|e^{i\psi} - c_k|^2} d\psi \\ &\leq \frac{4de}{(1-\gamma^2)^2 (1 - c_k^2)} \\ &= \frac{kde}{(1-\gamma)^2} \frac{4}{(1+\gamma)^2} \\ &\leq \frac{kde}{(1-\gamma)^2} (4 - 3\gamma), \end{aligned}$$

where the third line is obtained using the residue Theorem, and the last line uses $\frac{4}{(1+\gamma)^2} \leq 4-3\gamma$. This concludes the case (i).

Second step: dealing with case (ii)*.*

For cases (ii) and (iii), we will use the following inequalies

$$\begin{aligned}\left\|\left((z-c_k)^{-1}I_d-(z-\widetilde{Q})^{-1}\right)(I_d-\gamma U)^{-1}\right\|_F &\leq \left\|(z-c_k)^{-1}I_d-(z-\widetilde{Q})^{-1}\right\|_F\left\|(I_d-\gamma U)^{-1}\right\|_{\mathrm{op}}\\ &\leq \frac{\sqrt{d}}{1-\gamma}\left\|(z-c_k)^{-1}I_d-(z-\widetilde{Q})^{-1}\right\|_{\mathrm{op}}\\ &\leq \frac{\sqrt{d}}{1-\gamma}\left(|z-c_k|^{-1}+\left\|(z-\widetilde{Q})^{-1}\right\|_{\mathrm{op}}\right),\end{aligned} \tag{14}$$

where $\|\cdot\|_{\mathrm{op},\mathbb{R}^d}$ is the operator norm restricted to vectors on $\mathbb{R}^d$. Define $Q_R := \mathcal{R}(z)-\widetilde{Q} = \mathcal{R}(z)-c_k+\alpha c_k\Sigma_0(I_d-\gamma U)$, so that $z-\widetilde{Q} = Q_R + i\mathcal{I}(z)$. Take $x \in \mathbb{R}^d$ and $y_1+iy_2 = y = (z-\widetilde{Q})^{-1}x$, we have

$$\mathbb{R}^d \ni x = (z-\widetilde{Q})y = (Q_R+i\mathcal{I}(z))(y_1+iy_2) = Q_Ry_1-\mathcal{I}(z)y_2+i(Q_Ry_2+\mathcal{I}(z)y_1),$$

which implies $Q_Ry_2 = -\mathcal{I}(z)y_1$, so that we obtain

$$\begin{aligned}|(z-\widetilde{Q})y|^2 &= (y_1-iy_2)^\top(Q_R^\top-i\mathcal{I}(z))(Q_R+i\mathcal{I}(z))(y_1+iy_2)\\ &= |Q_Ry_1|^2+|Q_Ry_2|^2+\mathcal{I}(z)^2|y|^2-2\mathcal{I}(z)y_1^\top(Q_R^\top-Q_R)y_2\\ &= |Q_Ry_1|^2+\mathcal{I}(z)^2(2|y_1|^2+|y_2|^2)-2\alpha c_k\gamma\mathcal{I}(z)y_1^\top(\Sigma_0U-U^\top\Sigma_0)y_2\\ &\geq \mathcal{I}(z)^2(2|y_1|^2+|y_2|^2)-\frac{8\gamma\mathcal{I}(z)^2}{(1+\gamma)^2}|y_1||y_2|\\ &\geq \mathcal{I}(z)^2(2|y_1|^2+|y_2|^2)-2\mathcal{I}(z)^2|y_1||y_2|\\ &\geq \mathcal{I}(z)^2(2|y_1|^2+|y_2|^2)-\mathcal{I}(z)^2\left(\frac{3}{2}|y_1|^2+\frac{2}{3}|y_2|^2\right)\\ &= \mathcal{I}(z)^2\left(\frac{1}{2}|y_1|^2+\frac{1}{3}|y_2|^2\right) \geq \frac{\mathcal{I}(z)^2}{3}|y|^2\end{aligned} \tag{15}$$

where we used $|Q_Ry_2|^2 = \mathcal{I}(z)^2|y_1|^2$ and $Q_R^\top - Q_R = \alpha c_k\gamma(\Sigma_0U-U^\top\Sigma_0)$ to get the third line; $|\mathcal{I}(z)| \geq \sqrt{1-(1-(1-\gamma)^2)^2} \geq \sqrt{1-(1-(1-\gamma)^2)} = 1-\gamma$, $\alpha < \frac{2(1-\gamma)}{(1+\gamma)^2} \leq \frac{2|I(z)|}{(1+\gamma)^2}$, $c_k \leq 1$, $\|\Sigma_0U\|_{\mathrm{op}} \leq \|\Sigma_0\|_{\mathrm{op}}\|U\|_{\mathrm{op}} \leq 1$ to obtain the fourth line; $4\gamma \leq (1+\gamma)^2$ to get the fifth line; and the Young's inequality $2|y_1||y_2| \leq \frac{3}{2}|y_1|^2+\frac{2}{3}|y_2|^2$ to obtain the last one. This implies

$$\left\|(z-\widetilde{Q})^{-1}\right\|_{\mathrm{op}} \leq \sqrt{\frac{3}{\mathcal{I}(z)^2}} = \frac{\sqrt{3}}{|\sin(\psi)|}$$

for $z = e^{i\psi}$ such that $|\mathcal{R}(z)| \leq \gamma(2-\gamma)$. Similarly, we have $|z-c_k|^{-1} \leq \frac{1}{|\sin(\psi)|}$. Then using the above computations and $(\sqrt{3}+1)^2 \leq 8$ we obtain

$$\begin{aligned}
\frac{e}{2\pi}\int_{|\psi|\in(\psi_\gamma,\pi-\psi_\gamma)} &\left\|\left((e^{i\psi}-c_k)^{-1}I_d-(e^{i\psi}I_d-\widetilde{Q})^{-1}\right)(I_d-\gamma U)^{-1}\right\|_F^2 d\psi\\
&\leq \frac{8de}{\pi(1-\gamma)^2}\int_{\psi_\gamma}^{\pi-\psi_\gamma}\frac{1}{\sin(\psi)^2}d\psi\\
&\leq \frac{8de}{\pi(1-\gamma)^2}\left[-\mathrm{cotan}(\psi)\right]_{\psi_\gamma}^{\pi-\psi_\gamma}\\
&= \frac{16de}{\pi(1-\gamma)^2}\mathrm{cotan}(\psi_\gamma)\\
&= \frac{16de}{\pi(1-\gamma)^2}\frac{\gamma(2-\gamma)}{\sqrt{1-\gamma^2(2-\gamma)^2}}\\
&\leq \frac{16de}{\pi(1-\gamma)^2}\frac{1+(1-\gamma)}{\sqrt{(1-\gamma)^2(1+2\gamma-\gamma^2)}}\gamma\\
&\leq \frac{16de}{\sqrt{2}\pi(1-\gamma)^3}(1+(1-\gamma))\\
&\leq \frac{4de}{(1-\gamma)^3}(1+(1-\gamma)),
\end{aligned}$$

where we used that the primitive of $\frac{-1}{\sin^2}$ is cotan, $\mathrm{cotan}(\arccos(a)) = \frac{a}{\sqrt{1-a^2}}$ for $a\in(0,1)$, $\frac{\gamma}{\sqrt{1+2\gamma-\gamma^2}} \leq \frac{\gamma}{\sqrt{1+\gamma}} \leq \frac{1}{\sqrt{2}}$ and $\frac{16}{\sqrt{2}\pi} \leq 4$.

Third step: dealing with case (iii).

Once again take $x\in\mathbb{C}^d$ and $y=(z-\widetilde{Q})^{-1}x\in\mathbb{C}^d$, we have

$$\begin{aligned}
|x| = |(z-\widetilde{Q})y| &\geq |(z-c_k+c_k\alpha\Sigma_0)y| - \alpha\gamma|\Sigma_0 U y|\\
&\geq \left(\min_{\lambda\in\mathrm{Sp}(\Sigma_0)}|z-c_k+c_k\alpha\lambda|-\alpha\gamma\right)|y|\\
&\geq \left(\min_{u\in[0,1]}|z-c_k+c_k\alpha u|-\alpha\gamma\right)|y|
\end{aligned}$$

Then, let us focus on the term $|z-c_k+c_k\alpha u|$. Define $s = 1-c_k^{-1}\mathcal{R}(z)\in[1,1+c_k^{-1}]$, we get

$$\begin{aligned}
|z-c_k+c_k\alpha u|^2 &= (\mathcal{R}(z)-c_k+c_k\alpha u)^2+\mathcal{I}(z)^2\\
&= c_k^2(\alpha u-s)^2+1-c_k^2(s-1)^2\\
&= c_k^2(\alpha^2u^2-2s\alpha u+2s-1)+1 = f(u,s),
\end{aligned}$$

where $f(u,s) := c_k^2(\alpha^2u^2+2s(1-\alpha u)-1)+1$ for $u\in[0,1]$ and $s\in[1,1+c_k^{-1}]$. We make up two cases. First, if $\alpha u\leq 1$, we have

$$\min_s f(u,s) = f(u,1) = c_k^2(\alpha^2u^2+2(1-\alpha u)-1)+1 = c_k^2(\alpha u-1)^2+1 \geq 1,$$

which implies

$$|x| \geq (1-\alpha\gamma)|y| \geq \left(1-\frac{2\gamma(1-\gamma)}{(1+\gamma)^2}\right)|y| \geq \left(1-\frac{1-\gamma}{2}\right)|y| \geq \frac{|y|}{2},$$

where we used the above computations, $\alpha < \frac{2(1-\gamma)}{(1+\gamma)^2}$ and $4\gamma \leq (1+\gamma)^2$. Second, assume $\alpha u > 1$, we have

$$\begin{aligned}\min_s f(u,s) = f(u, 1+c_k^{-1}) &= c_k^2(\alpha^2u^2 + 2(1+c_k^{-1})(1-\alpha u) - 1) + 1\\ &= c_k^2(\alpha^2u^2 - 2(1+c_k^{-1})\alpha u + 2 + 2c_k^{-1} - 1 + c_k^{-2})\\ &= c_k^2(\alpha^2u^2 - 2(1+c_k^{-1})\alpha u + (1+c_k^{-1})^2)\\ &= c_k^2(\alpha u - (1+c_k^{-1}))^2\\ &= (1 - c_k(\alpha u - 1))^2\\ &\geq (1-(\alpha u-1))^2\\ &= (2-\alpha u)^2 \geq (2-\alpha)^2,\end{aligned}$$

where we used $1 - c_k(\alpha u - 1) \geq 0$, $\alpha u - 1 > 0$ and $c_k \leq 1$ to get the sixth line; $\alpha < 2$ and $u \leq 1$ to get the last one. This implies

$$|x| \geq (2 - \alpha - \gamma\alpha)|y| = (2 - (1+\gamma)\alpha)|y|,$$

where we have $(1+\gamma)\alpha < \frac{2(1-\gamma)}{1+\gamma} \leq 2$. In all cases, we obtain

$$|(z-\widetilde{Q})^{-1}x| \leq \min\left(\frac{1}{2}, 2-(1+\gamma)\alpha\right)^{-1}|x| = \max\left(2, \frac{1}{2-(1+\gamma)\alpha}\right)|x| \leq \left(2 + \frac{1}{2-(1+\gamma)\alpha}\right)|x|$$

This, (14) and $|z - c_k| \geq \sqrt{1+c_k^2} \geq 1$, since $\mathcal{R}(z) \leq 0$, imply

$$\begin{aligned}\left\|\left((z-c_k)^{-1}I_d - (z-\widetilde{Q})^{-1}\right)(I_d-\gamma U)^{-1}\right\|_F^2 &\leq \frac{2d}{(1-\gamma)^2}\left(|z-c_k|^{-2} + \left\|(z-\widetilde{Q})^{-1}\right\|_{\text{op}}^2\right)\\ &\leq \frac{2d}{(1-\gamma)^2}\left(3 + \frac{1}{2-(1+\gamma)\alpha}\right),\end{aligned}$$

and then

$$\frac{e}{2\pi}\int_{|\psi|\in[\pi-\psi_\gamma,\pi]}\left\|\left((e^{i\psi}-c_k)^{-1}I_d - (e^{i\psi}I_d-\widetilde{Q})^{-1}\right)(I_d-\gamma U)^{-1}\right\|_F^2 d\psi \leq \frac{2de}{(1-\gamma)^2}\left(3+\frac{1}{2-(1+\gamma)\alpha}\right).$$

Final step: concluding. From (13) and the three previous steps of the proof, we obtain

$$\begin{aligned}\mathcal{S}_k &\leq \frac{de}{(1-\gamma)^2}(1+3(1-\gamma)) + \frac{4de}{(1-\gamma)^3k}(1+(1-\gamma)) + \frac{2de}{(1-\gamma)^2k}\left(3+\frac{1}{2-(1+\gamma)\alpha}\right)\\ &\leq \frac{de}{(1-\gamma)^2}\left(1 + \frac{4}{k(1-\gamma)} + 3(1-\gamma) + \frac{10}{k} + \frac{1}{k(1-\frac{(1+\gamma)\alpha}{2})}\right).\end{aligned}$$

□

Lemma C.1. *Let $\Sigma \in \mathbb{R}^{d\times d}$ be symmetric positive semi-definite, let $U \in \mathbb{R}^{d\times d}$ be such that $\|U\|_{\text{op}} \leq 1$, $\gamma \in [0,1)$ and $z \in \mathbb{C}$ with $|z| = 1$ and $\mathcal{R}(z) \geq -\frac{1-\gamma}{\sqrt{2}}$. We have, for $x \in \mathbb{C}^d$,*

$$\left|\left(I_d - (I_d + z\Sigma(I_d-\gamma U))^{-1}\right)(I_d-\gamma U)^{-1}x\right| \leq \frac{2}{1-\gamma^2}|x|.$$

Proof. Define $A = I_d - \gamma U$ and $M = I_d + z\Sigma A$, we have

$$\begin{aligned}\left(I_d - (I_d + z\Sigma(I_d-\gamma U))^{-1}\right)(I_d-\gamma U)^{-1} &= (I_d - M^{-1})A^{-1}\\ &= M^{-1}(M-I_d)A^{-1}\\ &= z(I_d+z\Sigma A)^{-1}\Sigma\\ &= z\Sigma(I_d+zA\Sigma)^{-1}.\end{aligned}$$

Therefore, to conclude, it is sufficient to prove that, for an arbitrary $x \in \mathbb{C}^d$,

$$|\Sigma(I_d + zA\Sigma)^{-1}x| = |z\Sigma(I_d + zA\Sigma)^{-1}x| \leq \frac{2}{1-\gamma^2}|x|, \tag{16}$$

The remaining of the proof is dedicated to prove the latter inequality. Take $u = (I_d + zA\Sigma)^{-1}x$ and $r = \Sigma u$. Let us start with the subsequent observation,

$$|\overline{r}^\top A r - |r|^2| = \gamma |\overline{r}^\top U r| \leq \gamma |r|^2,$$

so that $\frac{\overline{r}^\top Ar}{|r|^2}$ belongs to $\overline{B}_{\mathbb{C}}(1,\gamma)$ the closed ball of center 1 and radius γ. Then, using $x = \Sigma^{-1}r + zAr$, we have

$$\frac{|\overline{r}^\top x|}{|r|^2} = \left|\frac{\overline{r}^\top \Sigma^{-1} r}{|r|^2} + z\frac{\overline{r}^\top A r}{|r|^2}\right| = \left|\frac{\overline{r}^\top A r}{|r|^2} + \overline{z}\frac{\overline{r}^\top \Sigma^{-1} r}{|r|^2}\right| \geq d_{|\cdot|}(\overline{B}_{\mathbb{C}}(1,\gamma), -\overline{z}\mathbb{R}_+),$$

where we used $\overline{r}^\top \Sigma^{-1} r \geq 0$, and $d_{|\cdot|}\left(\overline{B}_{\mathbb{C}}(1,\gamma), -\overline{z}\mathbb{R}_+\right)$ is the distance between the sets $\overline{B}_{\mathbb{C}}(1,\gamma)$ and $-\overline{z}\mathbb{R}_+$. Then, using Cauchy-Schwarz' inequality $|\overline{r}^\top x| \leq |r||x|$, we obtain

$$|x| \geq \frac{|\overline{r}^\top x|}{|r|} \geq d_{|\cdot|}(\overline{B}_{\mathbb{C}}(1,\gamma), -\overline{z}\mathbb{R}_+)|r|.$$

It only remains to derive an appropriate lower bound for the latter distance. Let us consider two cases. First, if $\mathcal{R}(z) \geq 0$, it is easy to check that this distance is equal to $1-\gamma$ that is reached for $1-\gamma \in \overline{B}_{\mathbb{C}}(1,\gamma)$ and $0 \in \overline{z}\mathbb{R}_-$. Second, if $0 \geq \mathcal{R}(z) \geq -\frac{1-\gamma}{\sqrt{2}}$. we have

$$\begin{aligned}
d_{|\cdot|}\left(\overline{B}_{\mathbb{C}}(1,\gamma), -\overline{z}\mathbb{R}_+\right) &= d_{|\cdot|}\left(1, -\overline{z}\mathbb{R}_+\right) - \gamma \\
&= \inf_{t\in\mathbb{R}_+} |1 + t\overline{z}| - \gamma \\
&= \inf_{t\in\mathbb{R}_+} \sqrt{|1 + t\mathcal{R}(z)|^2 + t^2(1 - |\mathcal{R}(z)|^2)} - \gamma \\
&= \inf_{t\in\mathbb{R}_+} \sqrt{1 + 2t\mathcal{R}(z) + t^2} - \gamma \\
&\geq \inf_{t\in\mathbb{R}_+} \sqrt{1 - \sqrt{2}(1-\gamma)t + t^2} - \gamma \\
&= \sqrt{1 - \frac{(1-\gamma)^2}{2}} - \gamma \\
&\geq 1 - \frac{(1-\gamma)^2}{2} - \gamma \\
&= \frac{1-\gamma^2}{2}
\end{aligned}$$

where we used the fact that the minimum of $t \in \mathbb{R}_+ \mapsto 1 - 2at + t^2$ equals $1 - a^2$ for $a \geq 0$ to get the sixth line, and $1 - \frac{a}{2} \geq (1+a)^{-1}$ for $a \in [0,1]$ to get the last one. Using the above calculations, we obtain

$$|\Sigma(I_d + zA\Sigma)^{-1}x| = |r| \leq d_{|\cdot|}\left(\overline{B}_{\mathbb{C}}(1,\gamma), -\overline{z}\mathbb{R}_+\right)^{-1}|x| \leq \frac{2}{1-\gamma^2}|x|,$$

which is exactly Inequality (16). This concludes the proof. □

Lemma C.2. *Let $M \in \mathbb{R}^{d\times d}$ be a matrix with a spectrum included in the open unit disk, and $c \in (0,1)$. We have*

$$\sum_{\ell=0}^{\infty} c^{2\ell}(I_d - M^{\ell\top})(I_d - M^\ell) = \frac{1}{2i\pi}\int_{|z|=1} z^{-1}\left((z^{-1}-c)^{-1} - (z^{-1}I_d - cM)^{-\top}\right)\left((z-c)^{-1} - (zI_d - cM)^{-1}\right)dz.$$

Proof. Since the spectral radius of M is lower than 1, it is power-bounded, i.e., $\sup_{k\geq 0}\|M^k\|_{\text{op}} < \infty$. Therefore, the series

$$F_0(z) = \sum_{\ell=0}^{\infty} c^\ell z^{-\ell-1} M^\ell,$$

is uniformly convergent on $\Gamma := \{z \in \mathbb{C}, |z| = 1\}$. This implies that $(I_d - cz^{-1}M)F_0(z) = z^{-1}I_d$ and then

$$F_0(z) = z^{-1}(I_d - cz^{-1}M)^{-1} = (zI_d - cM)^{-1}.$$

Similarly, we obtain

$$F(z) := (z-c)^{-1} - (zI_d - cM)^{-1} = \sum_{\ell=0}^{\infty} c^{\ell} z^{-\ell-1}(I_d - M^{\ell}),$$

so that, we get

$$\begin{aligned} z^{-1}F(z^{-1})^{\top}F(z) &= z^{-1}\left(\sum_{\ell=0}^{\infty} c^{\ell} z^{\ell+1}(I_d - M^{\ell})\right)^{\top}\left(\sum_{\ell=0}^{\infty} c^{\ell} z^{-\ell-1}(I_d - M^{\ell})\right) \\ &= \sum_{k,\ell\geq 0} c^{k+\ell} z^{k-\ell-1}(I_d - M^{k})^{\top}(I_d - M^{\ell}). \end{aligned}$$

For similar arguments as above, the latter double series is uniformly convergent on Γ which allows us to permute sums and integral as follows,

$$\begin{aligned} \int_{|z|=1} z^{-1}F(z^{-1})^{\top}F(z)dz &= \sum_{k,\ell\geq 0} c^{k+\ell}(I_d - M^{k})^{\top}(I_d - M^{\ell})\int_{|z|=1} z^{k-\ell-1}dz \\ &= 2i\pi \sum_{k,\ell\geq 0} c^{k+\ell}(I_d - M^{k})^{\top}(I_d - M^{\ell}), \end{aligned}$$

where we used that $\int_{|z|=1} z^{k-\ell-1} = 2i\pi$ by Residue theorem. This concludes the proof. ☐

C.2 Proof of Theorem 3.1

Proof of Theorem 3.1. This is a straightforward consequence of Propositions 4.1 and 5.3, with $\beta = (1+\gamma)^2$, $\beta_1 = 1+\gamma$, $c_{\Sigma} = (1-\gamma)^{-1}$ and $\sigma^2 = \sigma_0^2$. ☐

C.3 Proof of Proposition 3.2

Proof of Proposition 3.2. In this case, the matrices Σ_0, Σ_1, H, S are straightfoward to compute. In particular, we have

$$\Sigma_0 = \Sigma_1 = \omega I_d \qquad \text{and} \qquad H = S = (1-\gamma)\omega I_d.$$

Moreover, we have $b = \mathbb{E}[R_k \varphi(X_k)] = 0$ so that $\theta^* = 0$. Therefore, we have

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] = \omega \mathbb{E}\left[|\overline{\theta}_k|^2\right].$$

For $1 \leq i \leq d$, we have $\theta_{k,i}$ satisfies

$$\theta_{k,i} = (1 - d\alpha(1-\gamma)\omega B_{k,i})\theta_{k-1,i} + \alpha\sqrt{d\omega}B_{k,i}R_k,$$

where $(B_k)_{k\geq 1}$ are i.i.d. d-dimensional random variables, with $B_{1,i} \sim$Ber$(\frac{1}{d})$ and $\sum_{i=1}^{d} B_{1,i} = 1$ a.e..

On the one hand, from Lemma C.4, under the regime described in Proposition 3.2, for $C = 5.5$,

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim \frac{(1-e^{-C})^2|\theta_0|^2}{C\alpha(1-\gamma)k} + \frac{\sigma_0^2}{(1-\gamma)^2 k}\left(1 - \frac{3 - 4e^{-C} + e^{-2C}}{2C}\right). \tag{17}$$

On the other hand, under the same convergence regime, observe that the upper bound in Theorem 3.1 is equivalent to

$$\begin{aligned} E &= \left(\left(\frac{|\theta_0|^2}{2\alpha(1-\gamma)k}\right)^{\frac{1}{2}} + \left(\frac{e\sigma_0^2}{(1-\lambda)(1-\gamma)^2 k}\right)^{\frac{1}{2}}\right)^2 \\ &\leq \frac{(1+A)|\theta_0|^2}{2\alpha(1-\gamma)k} + \frac{2(1+A^{-1})e\sigma_0^2}{(1-\gamma)^2 k}, \end{aligned}$$

where we take $A = 2.7$. Then, we can easily check that $1.85 = \frac{1+A}{2} < 11\frac{(1-e^{-C})^2}{C} \approx 1.98$ and $7.45 \approx 2e(1+A^{-1}) < 11\left(1 - \frac{3-4e^{-C}+e^{-2C}}{2C}\right) \approx 8.02$. Therefore we obtain

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim 11E,$$

which concludes the proof. $\square$

Corollary C.3. *Consider the same MRP and feature functions as in Proposition 3.2. Take* $\alpha = \lambda\alpha_0(\gamma)$ *for* $\lambda \in (0,1)$. *For* $k \to \infty$, $\gamma \to 1$ *and* $k(1-\gamma) \to 1.25$, *we have*

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim \frac{1}{1.25}\frac{|\theta_0 - \theta^*|^2}{2\left(1 - \frac{\alpha}{\alpha_1(\gamma)}\right)\alpha(1-\gamma)k}.$$

Moreover, for $\gamma \to 1$ *and* $k(1-\gamma) \to \infty$, *we have*

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim \frac{1 - \frac{\alpha}{\alpha_0(\gamma)}}{e}\frac{de\sigma_0^2(1+\varepsilon_{k,\gamma})}{(1-\frac{\alpha}{\alpha_0(\gamma)})(1-\gamma)^2k}.$$

Proof. Similarly as in the proof of Proposition 3.2, for $k \to \infty$, $\gamma \to 1$ and $k(1-\gamma) \to 1.25$, the asymptotic inequality 17 holds for $C = 1.25$. This implies

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim \frac{(1-e^{-C})^2|\theta_0|^2}{C\alpha(1-\gamma)k}$$

which implies the first inequality in Corollary C.3, since $\frac{2(1-e^{-C})^2}{C} \geq \frac{1}{1.25}$.

Now consider for $\gamma \to 1$ and $k(1-\gamma) \to \infty$. Repeating the arguments in the proof of Lemma C.4, we obtain

$$\mathbb{E}_{X\sim m}\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] \gtrsim \frac{d\sigma_0^2}{(1-\gamma)^2k} \sim \frac{1 - \frac{\alpha}{\alpha_0(\gamma)}}{e}\frac{de\sigma_0^2(1+\varepsilon_{k,\gamma})}{(1-\frac{\alpha}{\alpha_0(\gamma)})(1-\gamma)^2k}.$$

This concludes the proof. $\square$

Lemma C.4. *Let* $(Y_k)_{k\geq 0}$ *be a stochastic process with value in* $\mathbb{R}$ *such that* $Y_0 = y_0 \in \mathbb{R}$ *is fixed and, for* $k \geq 1$

$$Y_k = (1-\lambda B_k)Y_{k-1} + B_kZ_k,$$

for $p \in [0,1]$ *and* $\lambda \in (0,2)$, *where* $(B_k, Z_k)_{k\geq 1}$ *are i.i.d. with* B_1 *independent from* Z_1, $B_1 \sim$*Ber*(p), $\mathbb{E}[Z_1] = 0$ *and* $\operatorname{Var}(Z_1) = \sigma^2 > 0$. *Define* $\overline{Y}_k = \frac{1}{k}\sum_{i=1}^{k-1}$. *In the regime* $k \to \infty$, $\lambda \to 0$ *and* $pk\lambda \to C$, *for some* $C > 0$, *we have*

$$\mathbb{E}\left[\overline{Y}_k^2\right] \gtrsim \frac{(1-e^{-C})^2y_0^2}{Cp\lambda k} + \frac{\sigma^2}{p\lambda^2k}\left(1 - \frac{3-4e^{-C}+e^{-2C}}{2C}\right).$$

Proof of Lemma C.4. Using superposition principle, we get $Y_k = Y_k^b + Y_k^v$, with

$$\begin{aligned} Y_k^b &= (1-\lambda B_k)Y_{k-1}^b, & Y_0^b &= y_0 \\ Y_k^v &= (1-\lambda B_k)Y_{k-1}^v + B_kZ_k, & Y_0^v &= 0. \end{aligned}$$

Moreover, using that Z_k is centered and independent with Y_k^b, we get $\mathbb{E}[Y_k^v] = 0$ and $\mathbb{E}\left[Y_k^bY_k^v\right] = 0$. Therefore, we obtain

$$\mathbb{E}\left[Y_k^2\right] = \mathbb{E}\left[(Y_k^b)^2\right] + \mathbb{E}\left[(Y_k^v)^2\right]$$

First step: the bias.

In this step, we only consider Y_k^b. We have

$$\mathbb{E}\left[Y_k^b\right] = p(1-\lambda)\mathbb{E}\left[Y_{k-1}^b\right] + (1-p)\mathbb{E}\left[Y_{k-1}^b\right] = (1-p\lambda)\mathbb{E}\left[Y_{k-1}^b\right] = (1-p\lambda)^ky_0,$$

so that we get

$$\mathbb{E}\left[(\overline{Y}^b_k)^2\right] \geq \mathbb{E}\left[\overline{Y}^b_k\right]^2 = \left(\frac{1}{k}\sum_{i=0}^{k-1}(1-p\lambda)^i y_0^2\right) = \frac{(1-(1-p\lambda)^k)^2 y_0^2}{\lambda^2 p^2 k^2}.$$

Therefore, for $k \to \infty$, $\lambda \to 0$ and $k\lambda p \to C > 0$, using $(1-p\lambda)^k \to e^{-C}$, we obtain

$$\mathbb{E}\left[(\overline{Y}^b_k)^2\right] \gtrsim \frac{(1-e^{-C})^2 y_0^2}{C\lambda p k}.$$

Second step: the variance term.

Recall that $\mathbb{E}[Y^v_k] = 0$. Then, for $\lambda' = 2\lambda - \lambda^2 \in (0,1)$, we have

$$\begin{aligned}
\mathbb{E}\left[(Y^v_k)^2\right] &= \left(p(1-\lambda)^2 + (1-p)\right)\mathbb{E}\left[(Y^v_{k-1})^2\right] + p\sigma^2 \\
&= \left(1 - p(2\lambda-\lambda^2)\right)\mathbb{E}\left[(Y^v_{k-1})^2\right] + p\sigma^2 \\
&= p\sigma^2 \sum_{i=0}^{k-1}(1-p\lambda')^i \\
&= \frac{(1-(1-p\lambda')^k)\sigma^2}{\lambda'}.
\end{aligned}$$

Let $(\mathcal{F}_k)_{k\geq 1}$ be the filtration associated to $(Z_k, B_k)_{k\geq 1}$. Using similar arguments as in the first step, for $1 \leq i \leq j$, we get

$$\mathbb{E}\left[Y^v_j \mid \mathcal{F}_i\right] = (1-p\lambda)^{j-i} Y^v_i.$$

This implies that

$$\mathbb{E}\left[Y^v_j Y^v_i\right] = (1-p\lambda)^{j-i}\mathbb{E}\left[(Y^v_i)^2\right] = (1-p\lambda)^{j-i}\frac{(1-(1-p\lambda')^i)\sigma^2}{\lambda'}.$$

Therefore, we obtain

$$\begin{aligned}
\mathbb{E}\left[(\overline{Y}^v_k)^2\right] &= \frac{1}{k^2}\left(2\sum_{0\leq i\leq j\leq k-1}\mathbb{E}\left[Y^v_j Y^v_i\right] - \sum_{i=0}^{k-1}\mathbb{E}\left[(Y^v_i)^2\right]\right) \\
&= \frac{\sigma^2}{\lambda' k^2}\sum_{i=0}^{k-1}\left((1-(1-p\lambda')^i)\left(2\sum_{j=i}^{k-1}(1-p\lambda)^{j-i} - 1\right)\right) \\
&= \frac{\sigma^2}{\lambda' k^2}\sum_{i=0}^{k-1}\left((1-(1-p\lambda')^i)\left(\frac{2(1-(1-p\lambda)^{k-i})}{p\lambda} - 1\right)\right) \\
&= \frac{2\sigma^2}{p\lambda\lambda' k^2}\sum_{i=0}^{k-1}\left(1-(1-p\lambda')^i - (1-p\lambda)^{k-i} + (1-p\lambda)^k\left(\frac{1-p\lambda'}{1-p\lambda}\right)^i\right) - \frac{\sigma^2}{\lambda' k^2}\sum_{i=0}^{k-1}(1-(1-p\lambda')^i) \\
&= \frac{2\sigma^2}{p\lambda\lambda' k^2}\left(k - \frac{1-(1-p\lambda')^k}{p\lambda'} - \frac{1-(1-p\lambda)^k}{p\lambda} + (1-p\lambda)\frac{(1-p\lambda)^k - (1-p\lambda')^k}{p(\lambda'-\lambda)}\right) \\
&\qquad - \frac{\sigma^2}{\lambda' k^2}\left(k - \frac{1-(1-p\lambda')^k}{p\lambda'}\right).
\end{aligned}$$

Therefore, for $k \to \infty$, $\lambda \to 0$ and $k\lambda p \to C > 0$, we obtain the following asymptotic equivalent

$$\begin{aligned}
\mathbb{E}\left[(\overline{Y}^v_k)^2\right] &\sim \frac{\sigma^2}{p\lambda^2 k}\left(1 - \frac{1-e^{-2C}}{2C} - \frac{1-e^{-C}}{C} + \frac{e^{-C}-e^{-2C}}{C}\right) - \frac{\sigma^2}{2\lambda k}\left(1 - \frac{1-e^{-2C}}{2C}\right) \\
&= \frac{\sigma^2}{p\lambda^2 k}\left(1 - \frac{3-4e^{-C}+e^{-2C}}{2C}\right) - \frac{\sigma^2}{2\lambda k}\left(1 - \frac{1-e^{-2C}}{2C}\right) \\
&\sim \frac{\sigma^2}{p\lambda^2 k}\left(1 - \frac{3-4e^{-C}+e^{-2C}}{2C}\right),
\end{aligned}$$

where we used $\lambda' \sim 2\lambda$, $(1-p\lambda)^k \sim e^{-C}$ and $(1-p\lambda')^k \sim e^{-2C}$ to get the first line, $\lambda \to 0$ and $1-\frac{3-4e^{-C}+e^{-2C}}{2C} > 0$ for $C > 0$ to get the last one.

This concludes the proof. $\square$

C.4 Proof of Theorem 3.3

Proof. The convergence rate is a consequence of Theorem B.3 with $\beta = (1+\gamma)^2$, $\beta_1 = 1+\gamma$, $c_\Sigma = (1-\gamma)^{-1}$ and $\sigma^2 = \sigma_0^2$.

Therefore, it only remains to prove the second part of the theorem. Take the Markov Process on $\mathcal{X} = \{1, -1\}$ such that $X' = -X$ a.e.. Take $m \equiv \frac{1}{2}$ the uniform probability distribution, $d = 1$ and $\phi(x) = x$. Consider $(R_k)_{k\geq 1}$ i.i.d. with $R_1 \sim \mathcal{N}(0, \sigma_0^2)$. For $\alpha \geq \alpha_1(\gamma) = \frac{2}{1+\gamma}$, $k \geq 1$, we have

$$\theta_k = (1-\alpha(1+\gamma))\theta_{k-1} + \alpha R_k = r^k\theta_0 + \alpha \sum_{i=0}^{k-1} r^i R_{k-i},$$

with $r = 1-\alpha(1+\gamma) < -1$. Therefore, we obtain

$$\mathbb{E}\left[|\theta_k|^2\right] = r^{2k}\theta_0^2 + \alpha^2\sigma_0^2 \sum_{i=0}^{k-1} r^{2i} = r^{2k}\theta_0^2 + \alpha^2\sigma_0^2 \frac{r^{2k}-1}{r^2-1} \xrightarrow[k\to\infty]{} +\infty.$$

Similarly, we have

$$\mathbb{E}_m\left[|v(X,\overline{\theta}_k) - v(X,\theta^*)|^2\right] = \mathbb{E}\left[|\overline{\theta}_k|^2\right] = \frac{(r^k-1)^2}{(1-r)^2k^2}|\theta_0|^2 + \frac{\alpha^2\sigma_0^2}{(1-r)^2k^2}|\theta_0|^2 \sum_{i=1}^{k-1} |r^i - 1|^2 \xrightarrow[k\to\infty]{} +\infty.$$

This concludes the proof. $\square$

C.5 Proof of Lemma 3.4

Proof of Lemma 3.4. We consider φ as in the first case in Lemma 3.4 and $(\theta_k)_{k\geq 0}$ the induced sequence obtained using TD(0). Similarly, we define $\widetilde{\varphi}$ and $(\widetilde{\theta}_k)_{k\geq 0}$ according to the second case. We are going to prove by induction on $k \geq 0$

$$\varphi(x)^\top\theta_k = \widetilde{\varphi}(x)^\top\widetilde{\theta}_k \quad \text{for any } x \in \mathcal{X}.$$

First for $k = 0$, we have

$$\widetilde{\varphi}(x)^\top\widetilde{\theta}_0 = \widetilde{\theta}_{0,1} + \widetilde{\varphi}_{(-1)}(x)^\top\widetilde{\theta}_{0,(-1)} = C\theta_{0,1} + \varphi_{(-1)}(x)^\top\theta_{0,(-1)} = \varphi(x)^\top\theta_0.$$

Then, for $k \geq 1$, assume that the equality holds at index $k-1$. Observe that $\widetilde{\varphi}(x)^\top\widetilde{\alpha}\widetilde{\varphi}(y) = \alpha\varphi(x)^\top\varphi(y)$, so that we obtain

$$\begin{aligned}
\widetilde{\varphi}(x)^\top\widetilde{\theta}_k &= \widetilde{\varphi}(x)^\top\left(\widetilde{\theta}_{k-1} - \widetilde{\alpha}\widetilde{\varphi}(X_k)\left((\widetilde{\varphi}(X_k) - \gamma\widetilde{\varphi}(X_k'))^\top\widetilde{\theta}_{k-1} - R_k\right)\right)\\
&= \varphi(x)^\top\theta_{k-1} - \alpha\varphi(x)^\top\varphi(X_k)\left((\varphi(X_k) - \gamma\varphi(X_k'))^\top\theta_{k-1} - R_k\right)\\
&= \varphi(x)^\top\left(\theta_{k-1} - \alpha\varphi(X_k)\left((\varphi(X_k) - \gamma\varphi(X_k'))^\top\theta_{k-1} - R_k\right)\right)\\
&= \varphi(x)^\top\theta_k.
\end{aligned}$$

$\square$

C.6 Asymptotic behavior of θ^* as γ tends to 1

Lemma C.5. *Assume* **TD1-TD3** *and that there exists $\theta_{\mathbb{1}} \in \mathbb{R}^d$ such that $v(\cdot, \theta_{\mathbb{1}}) = \mathbb{1}$, where $\mathbb{1}$ is the constant function equal to one. Then, we have*

$$\mu^\top\theta^* = \mathbb{E}\left[v(X,\theta^*)\right] = \frac{\mathbb{E}[R]}{1-\gamma}.$$

In particular, if $\mathbb{E}[R] \neq 0$*, we obtain*

$$|\theta^*| \geq \frac{|\mathbb{E}[R]|}{|\mu|(1-\gamma)} = O((1-\gamma)^{-1}).$$

Proof. Observe that $\Sigma_0\theta_1 = \mu = \Sigma_1^\top\theta_1$, which implies $\theta_1^\top(\Sigma_0 - \gamma\Sigma_1) = \theta_1^\top H = (1-\gamma)\mu^\top$, so that we get

$$\mathbb{E}\left[v(X,\theta^*)\right] = \mu^\top\theta^* = (1-\gamma)^{-1}\theta_1^\top H\theta^* = (1-\gamma)^{-1}\theta_1^\top b = (1-\gamma)^{-1}\mathbb{E}[R].$$

Recall that $\mu \neq 0$, otherwise $\theta_{\mathbb{1}}$ cannot exist. The second inequality in the lemma is straightforward. □

Lemma C.6. *Under the same assumptions as Theorem 3.5,* θ^* *is uniformly bounded with respect to* γ*.*

Proof. Observe that H and b write as

$$H = \begin{pmatrix} (1-\gamma)^{-1} & \mu_{(-1)}^\top \\ \mu_{(-1)} & \widetilde{H}, \end{pmatrix} \qquad\qquad b = \begin{pmatrix} (1-\gamma)^{-1}\mathbb{E}[R] \\ \widetilde{b} \end{pmatrix}$$

with $\widetilde{H} = \mathbb{E}\left[\varphi_{(-1)}(X)(\varphi_{(-1)}(X) - \gamma\varphi_{(-1)}(X'))^\top\right]$ and $\widetilde{b} = \mathbb{E}[R\varphi_{(-1)}(X)]$. From the first line in $H\theta^* = b$, we obtain

$$\theta_1 = \mathbb{E}[R] - (1-\gamma)\mu_{(-1)}^\top\theta_{(-1)}.$$

Using this and the other lines from $H\theta^* = b$, we get

$$\widetilde{H}\theta_{(-1)} = \widetilde{b} - \mathbb{E}[R]\mu_{(-1)},$$

where $\widetilde{H} = H - (1-\gamma)\mu_{(-1)}\mu_{(-1)}^\top$.

On the one hand, using similar arguments as in the proof of Theorem 3.7, we have

$$\theta_{(-1)}^\top\widetilde{H}\theta_{(-1)} \geq (1-\gamma(1-g))\theta_{(-1)}^\top\widetilde{\Sigma}_0\theta_{(-1)} \geq (1-\gamma(1-g))\widetilde{\omega}|\theta_{(-1)}|^2,$$

where $\widetilde{\Sigma}_0 = \mathbb{E}\left[(\varphi_{(-1)}(X) - \mu_{(-1)})(\varphi_{(-1)}(X) - \mu_{(-1)})^\top\right]$ and $\widetilde{\omega} > 0$ is the smallest eigenvalue of $\widetilde{\Sigma}_0$ which is positive using the linear independence of the features and the fact that φ_1 is constant.

On the other hand, we have

$$\theta_{(-1)}^\top\widetilde{H}\theta_{(-1)} = \theta_{(-1)}^\top(\widetilde{b} - \mathbb{E}[R]\mu_{(-1)}) \leq |\theta_{(-1)}|\mathbb{E}[R(\varphi_{(-1)} - \mu_{(-1)})]| \leq C_R^{\frac{1}{2}}|\theta_{(-1)}|,$$

where C_R is defined in Assumption **TD2**.

Consequently, we obtain

$$|\theta_{(-1)}| \leq \frac{C_R^{\frac{1}{2}}}{\widetilde{\omega}(1-\gamma(1-g))} \leq \frac{C_R^{\frac{1}{2}}}{\widetilde{\omega}g},$$

and then $|\theta_1| \leq C_R^{\frac{1}{2}} + \frac{C_R^{\frac{1}{2}}}{\widetilde{\omega}}$. This concludes the proof. □

C.7 Proof of Theorem 3.5

Proof of Theorem 3.5. The upper bound is a direct consequence of Proposition C.7 and Proposition B.4 with $\Sigma = \Sigma_0$, $\beta = 1 + (1+\gamma)^2$, $c_\Sigma = (1-\gamma)^{-1}$, and $\sigma^2 = \sigma_0^2$. The uniform boundedness of θ^* with respect to γ is given by Lemma C.6. □

Proposition C.7. *Under Assumptions* **TD1**,**TD2**, **td3***, the upper bound in Proposition 5.3 applies.*

Proof. We will repeat most of the arguments of the proof of Proposition 5.3.

First observe that $Q = \Sigma_0^{\frac{1}{2}}(I_d - \alpha H)\Sigma_0^{-\frac{1}{2}}$, thus it admits the same spectrum as $I_d - \alpha H$. Moreover, using (22), (23) we can take $c_\Sigma = (1-\gamma)^{-1}$ $\beta = 1 + (1+\gamma)^2$ so that (10) applies for $\alpha < \frac{2}{\beta c_\Sigma} = \frac{2(1-\gamma)}{1+(1+\gamma)^2} = \alpha_2(\gamma)$. Then (10) implies that the spectrum of $I_d - \alpha H$ is contained in the complex unit sphere. Therefore, Lemma C.2 applies to Q and Inequality (13) holds.

Then, the first step of the proof of Proposition 5.3 can be repeated without any change. Concerning the second step, the only change occurs to obtain the fourth line in the chain of inequalities (15): it requires that $\left\|\Sigma_0 U - U^\top \Sigma_0\right\|_{\mathrm{op}} \leq 2$ is indeed satisfied under the present assumption as stated in (27).

Therefore, there only remains the third step, i.e., for $\Re(z) \leq -1 + (1-\gamma)^2$, that we handle with a different approach using $\gamma \geq \frac{1}{2}$. First, we have

$$\alpha < \frac{2(1-\gamma)}{1+(1+\gamma)^2} \leq \frac{2(1-\gamma)}{1+\left(1+\frac{1}{2}\right)^2} = \frac{8}{13}(1-\gamma),$$

then, using Inequality (26), we get

$$\begin{aligned}
\alpha^2 \|\Sigma_0(I_d - \gamma U)\|_{\mathrm{op}}^2 &\leq \frac{(1-\gamma)^2}{4} \|\Sigma_0(I_d - \gamma U)\|_F^2 \\
&\leq \frac{8^2(1-\gamma)^2}{13^2}\left((1-\gamma)^2(((1-\gamma)^{-2}+1)^2+1) + 2(1+\gamma)^2\right) \\
&= \frac{8^2}{13^2}\left((1+(1-\gamma)^2)^2 + (1-\gamma)^4) + 2(1-\gamma^2)^2\right) \\
&\leq \frac{8^2}{13^2}\left(\frac{25}{16} + \frac{1}{16} + \frac{9}{8}\right) \leq 1.01^2.
\end{aligned}$$

Moreover, using $-\mathcal{R}(z) \geq 1 - (1-\gamma)^2 \geq \frac{3}{4}$, we obtain

$$|z - c_k|^2 = (c_k - \mathcal{R}(z))^2 + \mathcal{I}(z)^2 = c_k^2 - 2\mathcal{R}(z)c_k + 1 \geq c_k^2 + \frac{3}{2}c_k + 1.$$

Therefore, the above inequalities imply that, for $y \in \mathbb{C}^d$ with $|y| = 1$, we have

$$\begin{aligned}
|(z - \widetilde{Q})y| &= |\left((c_k - z)I_d - \alpha c_k \Sigma_0(I_d - \gamma U)\right) y| \\
&\geq |c_k - z||y| - \alpha c_k \|\Sigma_0(I_d - \gamma U)\|_{\mathrm{op}}|y| \\
&\geq \sqrt{c_k^2 + \frac{3}{2}c_k + 1} - 1.01 c_k = f(c_k),
\end{aligned}$$

where $f : x \in [0,1] \to \sqrt{x^2 + \frac{3}{2}x + 1} - 1.01x$. We have

$$f'(x) = \frac{x + \frac{3}{4}x}{\sqrt{1 + \frac{3}{2}x + x^2}} - 1.01 = \sqrt{\frac{x^2 + \frac{3}{2}x + \frac{9}{16}}{1 + \frac{3}{2}x + x^2}} - 1.01 \leq 0.$$

Consequently, we obtain $\left\|(z - \widetilde{Q})^{-1}\right\|_{\mathrm{op}} \leq f(1)^{-1} = \left(\sqrt{\frac{7}{2}} - 1.01\right)^{-1} \leq 2.$

The remaining of the proof is similar to that of Proposition 5.3. □

C.8 Proof of Theorem 3.7

Proof of Theorem 3.7. Define $\widehat{\varphi} = \varphi - \mu$, $\widehat{\Sigma}_0 = \mathbb{E}[\widehat{\varphi}(X)\widehat{\varphi}(X)^\top] = \Sigma_0 - \mu\mu^\top$, $\widehat{\Sigma}_1 = \mathbb{E}[\widehat{\varphi}(X)\widehat{\varphi}(X')^\top] = \Sigma_1 - \mu\mu^\top$, $\widehat{H} = \widehat{\Sigma}_0 - \gamma\widehat{\Sigma}_1$, $\widehat{S} = \frac{\widehat{H}+\widehat{H}^\top}{2}$ and $\widehat{b} = \mathbb{E}[R\widehat{\varphi}(X)] = b - \mathbb{E}[R]\mu$. Observe that, using Assumption **TD4**, we have, for any $\theta, \theta' \in \mathbb{R}^d$,

$$\theta^\top \widehat{\Sigma}_1 \theta' \leq (1-g)|\widehat{\Sigma}_0^{\frac{1}{2}}\theta||\widehat{\Sigma}_0^{\frac{1}{2}}\theta'|.$$

Therefore, $\widetilde{U}$ defined by $\widetilde{U} = (1-g)^{-1}\widehat{\Sigma}_0^{-\frac{1}{2}}\widehat{\Sigma}_1\widehat{\Sigma}_0^{-\frac{1}{2}}$ satisfies $\left\|\widetilde{U}\right\|_{\mathrm{op}} \leq 1$. Then we obtain

$$\widehat{H} = \widehat{\Sigma}_0^{\frac{1}{2}}(I_d - (1-g)\gamma\widetilde{U})\widehat{\Sigma}_0^{\frac{1}{2}},$$

which is the same structure as for TD(0) but with γ replaced with $(1-g)\gamma$.

To prove Theorem 3.1, the only property that we used for U was that it operator norm was bounded by one. Therefore, we can repeat all the proof in the case of $\widehat{H}$ with γ replaced by $(1-g)\gamma$. This proves Theorem 3.7. ☐

C.9 Some results on complex integration

Lemma C.8. *For $u \in [0,1)$, we have*

$$\int_0^{2\pi} \frac{d\theta}{|e^{-i\theta}-u|^2} = \frac{2\pi}{1-u^2}$$

Proof. Let $G = \{z \in \mathbb{C}, |z| = 1\}$, we have

$$\begin{aligned}\int_0^{2\pi} \frac{d\theta}{|e^{-i\theta}-u|^2} &= \int_G \frac{1}{|z-u|^2}\frac{dz}{iz} \\ &= \int_G \frac{dz}{(z-u)(z^{-1}-u)iz} \\ &= \int_G \frac{dz}{i(z-u)(1-zu)}.\end{aligned}$$

Recall that the function $\frac{1}{i(z-u)(1-zu)}$ is holomorph on $D(0,1)\backslash\{u\}$ with a pole in u. Therefore, we can use the Residue Theorem to get

$$\int_0^{2\pi} \frac{d\theta}{|e^{-i\theta}-u|^2} = 2\pi i \operatorname{Res}_u\left(z \mapsto \frac{1}{i(z-u)(1-zu)}\right) = 2\pi i \frac{1}{i(1-u^2)} = \frac{2\pi}{1-u^2}.$$

This conclude the proof. ☐

C.10 Few results from linear algebra

Lemma C.9. *Assume* **TD1-TD3**. *Using the notations of the algorithm TD(0) from Section 1, define $h^{\mathrm{TD}} = \varphi(X)(\varphi(X)-\gamma\varphi(X'))$ and $S^{\mathrm{TD}} = \Sigma_0 - \gamma\Sigma_1$, we have*

$$\begin{aligned}(1-\gamma)\Sigma_0 \le S^{\mathrm{TD}} &\le (1+\gamma)\Sigma_0, &(18)\\ \mathbb{E}\left[h^{\mathrm{TD}}(h^{\mathrm{TD}})^\top\right] &\le (1+\gamma)^2\Sigma_0 &(19)\\ \mathbb{E}\left[(h^{\mathrm{TD}})^\top h^{\mathrm{TD}}\right] &\le (1+\gamma)S^{\mathrm{TD}}, &(20)\\ \mathbb{E}\left[(h^{\mathrm{TD}}\theta^* - b^{\mathrm{TD}})(h^{\mathrm{TD}}\theta^* - b^{\mathrm{TD}})^\top\right] &\le \sigma_0^2\Sigma_0. &(21)\end{aligned}$$

Proof. To get (18), it is sufficient to observe

$$S^{\mathrm{TD}} = \Sigma_0 - \frac{\gamma}{2}\mathbb{E}\left[\varphi(X)\varphi(X')^\top + \varphi(X')\varphi(X)^\top\right],$$

and use Cauchy-Schwarz inequality.

To get (19), we compute

$$\begin{aligned}\mathbb{E}\left[h^{\mathrm{TD}}(h^{\mathrm{TD}})^\top\right] &= \mathbb{E}\left[|\varphi(X)-\gamma\varphi(X')|^2\varphi(X)\varphi(X)^\top\right] \\ &\le (1+\gamma)^2\mathbb{E}\left[\varphi(X)\varphi(X)^\top\right] = (1+\gamma)^2\Sigma_0.\end{aligned}$$

Now let us prove (20)

$$\begin{aligned}
\mathbb{E}\left[(h^{\mathrm{TD}})^\top h^{\mathrm{TD}}\right] &= \mathbb{E}\left[|\varphi(X)|^2(\varphi(X)-\gamma\varphi(X'))(\varphi(X)-\gamma\varphi(X'))^\top\right] \\
&\le \mathbb{E}\left[(\varphi(X)-\gamma\varphi(X'))(\varphi(X)-\gamma\varphi(X'))^\top\right] \\
&= \mathbb{E}\left[(1+\gamma^2)\varphi(X)\varphi(X)^\top - \gamma(\varphi(X)\varphi(X')^\top + \varphi(X')\varphi(X))\right] \\
&= (1+\gamma^2)\Sigma_0 - \gamma(\Sigma_1+\Sigma_1^\top) \\
&= (1+\gamma)S^{\mathrm{TD}} - \gamma(1-\gamma)\left(\Sigma_0 - \frac{\Sigma_1+\Sigma_1^\top}{2}\right) \\
&\le (1+\gamma)S^{\mathrm{TD}}.
\end{aligned}$$

We conclude the proof with (21):

$$\mathbb{E}\left[(h^{\mathrm{TD}}\theta^* - b^{\mathrm{TD}})(h^{\mathrm{TD}}\theta^* - b^{\mathrm{TD}})^\top\right] = \mathbb{E}\left[|\delta(X,X',\theta^*)|^2\varphi(X)\varphi(X)^\top\right] \le \sigma_0^2\Sigma_0.$$

□

Lemma C.10. *Assume* **TD1**, **TD2** *and* **td3**. *We have*

$$(1-\gamma)\Sigma_0 \le S^{\mathrm{TD}} \le (1+\gamma)\Sigma_0, \tag{22}$$

$$\mathbb{E}\left[h^{\mathrm{TD}}(h^{\mathrm{TD}})^\top\right] \le ((1-\gamma)^2c_\gamma^2 + (1+\gamma)^2)\Sigma_0, \tag{23}$$

$$\mathbb{E}\left[(h^{\mathrm{TD}}\theta^* - b^{\mathrm{TD}})(h^{\mathrm{TD}}\theta^* - b^{\mathrm{TD}})^\top\right] \le \sigma_0^2\Sigma_0, \tag{24}$$

$$\left\|\Sigma_0 U - U^\top\Sigma_0\right\|_{\mathrm{op}} \le 2, \tag{25}$$

$$\|\Sigma_0(I_d-\gamma U)\|_F^2 \le (1-\gamma)^2((c_\gamma^2+1)^2+1) + 2(1+\gamma)^2, \tag{26}$$

Proof. The proofs of (22) and (24) are similar as the ones of (18) and (21). To get (23), we compute

$$\begin{aligned}
\mathbb{E}\left[h^{\mathrm{TD}}(h^{\mathrm{TD}})^\top\right] &= \mathbb{E}\left[|\varphi(X)-\gamma\varphi(X')|^2\varphi(X)\varphi(X)^\top\right] \\
&\le ((1-\gamma)^2c_\gamma^2 + (1+\gamma)^2)\Sigma_0,
\end{aligned}$$

where we used that $|\varphi(X)-\gamma\varphi(X')|^2 = (1-\gamma)^2c_\gamma^2 + |\varphi_{(-1)}(X)-\gamma\varphi_{(-1)}(X')|^2 \le (1-\gamma)^2c_\gamma^2 + (1+\gamma)^2$.

It only remains to prove (25). Let us define $\widetilde{\Sigma}_0, \widetilde{\Sigma}_1 \in \mathbb{R}^{(d-1)\times(d-1)}$ by

$$\widetilde{\Sigma}_0 = \mathbb{E}\left[(\varphi_{(-1)}(X)-\mu_{(-1)})(\varphi_{(-1)}(X)-\mu_{(-1)})^\top\right] \qquad \text{and} \qquad \widetilde{\Sigma}_1 = \mathbb{E}\left[(\varphi_{(-1)}(X)-\mu_{(-1)})(\varphi_{(-1)}(X')-\mu_{(-1)})^\top\right].$$

Define $\widetilde{U} = \widetilde{\Sigma}_0^{-\frac{1}{2}}\widetilde{\Sigma}_1\widetilde{\Sigma}_0^{-\frac{1}{2}}$. Consider

$$L = \begin{pmatrix} c_\gamma & 0 \\ \mu_{(-1)} & \widetilde{\Sigma}_0^{\frac{1}{2}} \end{pmatrix} \text{ so that } LL^\top = \widetilde{\Sigma}_0, \quad L\begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix}L^\top = \widetilde{\Sigma}_1, \quad L^\top L = Q^\top\widetilde{\Sigma}_0 Q \text{ and } U = Q\begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix}Q^\top,$$

where $Q := \widetilde{\Sigma}_0^{-\frac{1}{2}}L \in \mathcal{O}(\mathbb{R}^{d-1})$. This implies that

$$\begin{aligned}
Q^\top(\Sigma_0U - U^\top\Sigma_0)Q &= L^\top L\begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix} - \begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}^\top\end{pmatrix}L^\top L \\
&= \begin{pmatrix} c_\gamma^2 + |\mu_{(-1)}|^2 & \mu_{(-1)}^\top\widetilde{\Sigma}_0^{\frac{1}{2}} \\ \widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{\Sigma}_0 \end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix} - \begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix}\begin{pmatrix} c_\gamma^2 + |\mu_{(-1)}|^2 & \mu_{(-1)}^\top\widetilde{\Sigma}_0^{\frac{1}{2}} \\ \widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{\Sigma}_0 \end{pmatrix}\begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix} \\
&= \begin{pmatrix} c_\gamma^2 + |\mu_{(-1)}|^2 & \mu_{(-1)}^\top\widetilde{\Sigma}_0^{\frac{1}{2}}\widetilde{U} \\ \widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{\Sigma}_0\widetilde{U} \end{pmatrix} - \begin{pmatrix} c_\gamma^2 + |\mu_{(-1)}|^2 & \mu_{(-1)}^\top\widetilde{\Sigma}_0^{\frac{1}{2}} \\ \widetilde{U}^\top\widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{U}^\top\widetilde{\Sigma}_0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & \mu_{(-1)}^\top\widetilde{\Sigma}_0^{\frac{1}{2}}(\widetilde{U}-I_d) \\ (I_d-\widetilde{U}^\top)\widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{\Sigma}_0\widetilde{U} - \widetilde{U}^\top\widetilde{\Sigma}_0 \end{pmatrix}. \\
&= \begin{pmatrix} 0 & -y^\top \\ y & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \widetilde{\Sigma}_0\widetilde{U} - \widetilde{U}^\top\widetilde{\Sigma}_0 \end{pmatrix},
\end{aligned}$$

where $y := (I_d - \widetilde{U}^\top)\widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)}$. Therefore, we obtain

$$\left\|\Sigma_0 U - U^\top \Sigma_0\right\|_{\mathrm{op}} \le |y|^2 + \left\|\widetilde{\Sigma}_0 \widetilde{U} - \widetilde{U}^\top \widetilde{\Sigma}_0\right\|_{\mathrm{op}}. \tag{27}$$

On the one hand, we have

$$\begin{aligned}|y|^2 \le 2\left\|\widetilde{\Sigma}_0^{\frac{1}{2}}\right\|_{\mathrm{op}}|\mu_{(-1)}| \le \left\|\widetilde{\Sigma}_0\right\|_{\mathrm{op}} + |\mu_{(-1)}|^2 \le \mathrm{tr}(\widetilde{\Sigma}_0) + |\mu_{(-1)}|^2 \\ = \mathbb{E}\left[|\varphi_{(-1)}(X) - \mu_{(-1)}|^2\right] + \mu_{(-1)}^2 = \mathbb{E}\left[|\varphi_{(-1)}(X)|^2\right] \le 1.\end{aligned}$$

On the other hand, take $(\lambda_i, v_i)_{1 \le i \le d-1}$ the couples of eigenvalues and eigenvectors of $\widetilde{\Sigma}_0$, we have

$$\begin{aligned}\left\|\widetilde{\Sigma}_0 \widetilde{U} - \widetilde{U}^\top \widetilde{\Sigma}_0\right\|_{\mathrm{op}} &= \sum_{i=1}^{d-1} \lambda_i (v_i (\widetilde{U}^\top v_i)^\top - (\widetilde{U}^\top v_i) v_i^\top) \\ &\le \sum_{i=1}^{d-1} \lambda_i \left\|v_i (\widetilde{U}^\top v_i)^\top - (\widetilde{U}^\top v_i) v_i^\top\right\|_{\mathrm{op}} \\ &\le \sum_{i=1}^{d-1} \lambda_i |v_i||\widetilde{U}^\top v_i| \\ &\le \sum_{i=1}^{d-1} \lambda_i = \mathrm{tr}(\widetilde{\Sigma}_0) \le 1.\end{aligned}$$

where we used Lemma C.11 to get the third line, $\|\widetilde{U}\|_{\mathrm{op}} \le 1$ to get the last line. We conclude using (27) and the above inequalities.

Let us now prove Inequality (26). Using similar computations as above, we have

$$\begin{aligned}Q\Sigma_0(I_d - \gamma U)Q^\top = L^\top L\left(I_d - \gamma\begin{pmatrix}1 & 0 \\ 0 & \widetilde{U}\end{pmatrix}\right) &= \begin{pmatrix} c_\gamma^2 + |\mu_{(-1)}|^2 & \mu_{(-1)}^\top \widetilde{\Sigma}_0^{\frac{1}{2}} \\ \widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{\Sigma}_0\end{pmatrix}\begin{pmatrix}1-\gamma & 0 \\ 0 & I_{d-1} - \gamma\widetilde{U}\end{pmatrix} \\ &= \begin{pmatrix}(1-\gamma)(c_\gamma^2 + |\mu_{(-1)}|^2) & \mu_{(-1)}^\top \widetilde{\Sigma}_0^{\frac{1}{2}}(I_{d-1} - \gamma\widetilde{U}) \\ (1-\gamma)\widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)} & \widetilde{\Sigma}_0(I_{d-1} - \gamma\widetilde{U})\end{pmatrix}.\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\left\|\Sigma_0(I_d - \gamma U)Q^\top\right\|_F^2 &= (1-\gamma)^2(c_\gamma^2 + |\mu_{(-1)}|^2)^2 + \left|(I_d - \gamma\widetilde{U}^\top)\widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)}\right|^2 \\ &\quad + \left|(1-\gamma)\widetilde{\Sigma}_0^{\frac{1}{2}}\mu_{(-1)}\right|^2 + \left\|\widetilde{\Sigma}_0(I_{d-1} - \gamma\widetilde{U})\right\|_F^2 \\ &\le (1-\gamma)^2(c_\gamma^2 + 1)^2 + (1+\gamma)^2 + (1-\gamma)^2 + (1+\gamma)^2 \\ &= (1-\gamma)^2((c_\gamma^2 + 1)^2 + 1) + 2(1+\gamma)^2,\end{aligned}$$

where we used $\|\widetilde{U}\|_{\mathrm{op}}, \|\widetilde{\Sigma}\|_{\mathrm{op}} \le 1$ and $\left\|\widetilde{\Sigma}_0(I_{d-1} - \gamma\widetilde{U})\right\|_F^2 \le (1+\gamma)^2\mathrm{tr}(\widetilde{\Sigma}_0^2) \le (1+\gamma)^2\mathrm{tr}(\widetilde{\Sigma}_0) \le (1+\gamma)^2$ since $\mathrm{tr}(\widetilde{\Sigma}_0) = \mathbb{E}[|\varphi_{(-1)}(X)|] \le 1$. □

Lemma C.11. *For $u, v \in \mathbb{R}^d$, we have $\left\|uv^\top - vu^\top\right\|_{\mathrm{op}} \le |v||u|$.*

Proof. It is sufficient to prove the result for $|u| = |v| = 1$. Take $v = \lambda_1 u + \lambda_2 w$, with $\lambda_1^2 + \lambda_2^2 = 1$, $w \in u^\perp$ and $|w| = 1$. We have

$$uv^\top - vu^\top = \lambda_2(uw^\top - wu^\top).$$

Moreover, since $u \perp w$, for any $x \in \mathbb{R}^d$, we have

$$|(uw^\top - wu^\top)x|^2 = |w^\top x|^2 + |u^\top x|^2 = \left|\Pi_{\mathrm{Span}(u,w)}x\right|^2 \le |x|^2,$$

where $\Pi_{\mathrm{Span}(u,w)}$ is the orthogonal projection on $\mathrm{Span}(u, w)$. Therefore, we obtain

$$\left\|uv^\top - vu^\top\right\|_{\mathrm{op}} \le \lambda_2 \left\|uw^\top - wu^\top\right\|_{\mathrm{op}} \le \lambda_2 \le 1.$$

This concludes the proof. □