Linear Stochastic Approximation: How Far Does Constant Step-Size and Iterate Averaging Go?

Chandrashekar Lakshminarayanan **Csaba Szepesvári**
DeepMind

Abstract

In this paper we study study constant step-size averaged linear stochastic approximation. With an eye towards linear value estimation in reinforcement learning, we ask whether for a given class of linear estimation problems *i)* a single *universal* constant step-size with *ii)* a C/t worst-case expected error with a class-dependent constant $C > 0$ can be guaranteed when the error is measured via an appropriate weighted squared norm. Such a result has recently been obtained in the context of linear least squares regression. We give examples that show that the answer to these questions in general is *no.* On the positive side, we also characterize the instance dependent behavior of the error of the said algorithms, identify some conditions under which the answer to the above questions can be changed to the positive, and in particular show instance-dependent error bounds of magnitude $O(1/t)$ for the constant step-size iterate averaged versions of TD(0) and a novel variant of GTD, where the stepsize is chosen independently of the value estimation instance. Computer simulations are used to illustrate and complement the theory.

1 Introduction

Various estimation problems in supervised, unsupervised, or reinforcement learning and beyond are formulated as the problem of finding the unique solution $\theta_* \in \mathbb{R}^d$ to the linear equation $\mathbf{E}[A_t]\,\theta = \mathbf{E}[b_t]$, where $\{(A_t, b_t)\}_{t\geq 1}$ is an $\mathbb{R}^{d\times d} \times \mathbb{R}^d$-valued random sequence with a common distribution P and the expectation $\mathbf{E}[A_t]$ of the matrix A_t is non-singular [e.g., 2, 13, 17, 6, 19, 24, 11, 21, 20, 12]. Oftentimes, the matrices A_t are rank-1, $\mathbf{E}[A_t]$ is Hurwitz (its eigenvalues have positive real parts) and the dimensionality d is large. Then, for any positive-valued, user-chosen stepsize sequence $\{\alpha_t\}_{t\geq 1}$, the updates

$$\theta_t = \theta_{t-1} + \alpha_t(b_t - A_t\theta_{t-1}) \qquad (1)$$

can be implemented in $O(d)$ time and space, making such *linear stochastic approximation (LSA) algorithms* an appealing alternative to directly computing the solution to $\bar{A}_t\theta = \bar{b}_t$, where $\bar{A}_t = \frac{1}{t}\sum_{s=1}^{t} A_s$, $\bar{b}_t = \frac{1}{t}\sum_{s=1}^{t} b_s$ by inverting $\bar{A}_t$ (in which case the computational and storage costs are $O(d^2)$ or more).

Assuming sufficient regularity of $\{(A_t, b_t)\}_{t\geq 1}$, e.g., independence, or mixing, in addition to bounded moments, if the stepsize sequence converges to zero at an appropriate rate, convergence of $\{\theta_t\}_{t\geq 0}$ to θ_* can be guaranteed in various senses [2, 3]. In applications, one often starts from some additional broad properties of the common distribution P underlying $\{(A_t, b_t)\}$, i.e., $P \in \mathcal{P}$ for a known family of instances $\mathcal{P}$. For example, in *linear regression under the squared loss criterion* (LS), $A = \mathbf{E}[A_t]$ (the expectation of A_t) is symmetric and positive definite and $\mathbf{E}\left[\|b_t\|^2\right]$, $\mathbf{E}\left[\|A_t\|^2\right] \leq B$ with B known. The goal then is not only to guarantee asymptotic convergence on a per-instance basis, but also to choose $\{\alpha_t\}_{t\geq 1}$ based on the knowledge of $\mathcal{P}$ only, so that the worst-case error is "small" over the whole class $\mathcal{P}$ and for any $t \geq 1$.

To overcome the difficulty of choosing such a "universally good" stepsize sequence, following the ideas of Ruppert [15], Polyak and Juditsky [14], in the context of linear regression with the squared loss (LS), Bach and Moulines suggested that (1) should be used with $\alpha_t = \alpha > 0$ $(t \geq 1)$ with some $\alpha > 0$ to be chosen based on $\mathcal{P}$, and use the average $\hat{\theta}_t \doteq \frac{1}{t+1}\sum_{s=0}^{t}\theta_s$ as the output [1]. Their main result is that for the LS problem, under the assumption that $\{(A_t, b_t)\}_{t\geq 1}$ is an independent sequence, the stepsize α can be chosen solely based on the above-mentioned upper bound B to guarantee that for some universal constant $C > 0$

Proceedings of the 21^{st} International Conference on Artificial Intelligence and Statistics (AISTATS) 2018, Lanzarote, Spain. PMLR: Volume 84.

the expected squared prediction error of using $\hat{\theta}_t$ is at most $C\,d\,B^2/t$ for any $t \geq 1$, which is information-theoretically near-optimal (e.g., [16]).

In this paper we ask to what extent the nice result of Bach and Moulines can be extended beyond LS; in other words, we are asking which aspects of the LS problem play a critical role. Our interest stems from the desire to reproduce this result for linear value-function estimation (LVE) in reinforcement learning (RL) where multiple members of the *temporal-difference (TD) family of algorithms* (cf. [19, 21, 20, 12] and Section 5) have been proposed as an analog of the "LMS algorithm" analyzed by Bach and Moulines [1]. The extension is not straightforward as there are a number of critical differences between the properties of the instances in LS and LVE. In particular, in LVE (*i*) $A = \mathbf{E}[A_t]$ is in general non-symmetric; (*ii*) the sequence $\{(A_t, b_t)\}_{t\geq 1}$ is "driven by Markov noise" [23]; (*iii*) and while the natural error metric in LS problems is $\|\hat{\theta}_t - \theta_*\|_A^2$ (here $\|x\|_Q^2 = x^\top Q x$ for Q symmetric, positive definite), this is not the case in TD. Of these differences we only consider (*i*) and (*iii*), assuming that $\{(A_t, b_t)\}_{t\geq 1}$ is an independent, identically distributed (i.i.d.) sequence. For our results that have a negative character there is no loss of generality in making this assumption. The question as to what extent our other results can be extended to the Markov noise case remains for future work.

Regarding the quest to extend the analysis of Bach and Moulines [1] to LVE we provide the following results:

1. **Finite-time Instance Dependent Bounds** (Section 3): When $\mathbf{E}[A_t]$ is Hurwitz, we show that under additional mild regularity assumptions there exists a constant $\alpha_P > 0$ such that for any $\alpha \in (0, \alpha_P)$, the mean-squared error (MSE), $\mathbf{E}\left[\|\hat{\theta}_t - \theta_*\|^2\right]$ is at most $\frac{C}{t} + \frac{C'}{t^2}$ with some positive constants C, C' that we explicitly compute from P and α (Theorem 1). Our result is an extension of the result by Polyak and Juditsky [14] who proved a similar result for the case when $A_t = A$. We also show that our upper bound is essentially tight up to a universal constant factor (Theorem 2), thus Theorem 1 captures the instance-dependent behavior of CALSA in a faithful manner.

2. **Problem Landscape** (Section 4): By means of a simple example we establish that Hurwitzness and uniform boundedness of $\{(A_t, b_t)\}_{t\geq 1}$ alone are insufficient for the existence of a single, *universal* stepsize (Proposition 3). Here, a universal stepsize is one that guarantees the convergence of the worst-case expected squared error over the class of problems to zero as $t \to \infty$. This result is complemented by Theorem 4, which distills the importance of various aspects of the LS problem, such as the positive definiteness of A, that the error is measured in norm $\|\cdot\|_A$, or the so-called "structured noise" property, in governing the worst-case error. The strength of our results is that they give the exact behavior of the worst-case error (i.e., matching lower and upper bounds).

3. **Reinforcement Learning** (Section 5): In the context of reinforcement learning we establish that the constant stepsize averaged TD(0) and a novel version of GTD assume universal stepsizes in a number of cases for various problem classes with bounded data (Theorem 5). In particular, this is first shown for averaged TD(0) for the "on-policy case", which we define using the so-called "second-order feature stationarity condition". This change is partially necessary because we consider the i.i.d. case. However, the new condition can also be viewed as the "true" condition to guarantee the stability of TD(0). Finally, we establish that this condition can be dropped for the novel version of GTD. The strength of these results is that a user who is concerned with achieving the $O(1/t)$ problem-dependent rate over broad classes of LVE problems is relieved from the burden of designing stepsize tuning methods. In Section 6 we illustrate these theoretical results by means of computer simulations.

After our results, we briefly discuss related work in Section 7. In connection to this, we also wish to mention that other computationally cheap methods such as those based on matrix sketching idea [26], could be a viable alternative to CALSA. However, regardless of this, we believe that understanding a simple method like CALSA remains an important and foundational challenge.

Notation: The set of reals is denoted by $\mathbb{R}$. The d-dimensional vector space over $\mathbb{R}$ is denoted by $\mathbb{R}^d$, while $\mathbb{R}^{p\times q}$ denotes the vector space of $p \times q$ matrices over the reals. Vectors are column vectors (i.e., $\mathbb{R}^d$ is identified with $\mathbb{R}^{d\times 1}$). The transpose of a matrix C is denoted by $C^\top$ (and of course the same notation applies to vectors, as well). We will use $\langle \cdot, \cdot \rangle$ to denote inner products: $\langle x, y \rangle = x^\top y$ and use $\|x\| = \langle x, x \rangle^{1/2}$ to denote the 2-norm. We call a matrix $A \in \mathbb{R}^{d\times d}$ *Hurwitz* (H) if all eigenvalues of A have strictly positive real parts. We call a matrix $A \in \mathbb{R}^{d\times d}$ *positive definite* (PD) if $\langle x, Ax \rangle > 0$ for all nonzero $x \in \mathbb{R}^d$. If $\inf_x \langle x, Ax \rangle \geq 0$ then A is *positive semi-definite* (PSD). We call a matrix $A \in \mathbb{R}^{d\times d}$ to be *symmetric positive definite* (SPD) if it is symmetric i.e., $A^\top = A$ and PD. For $C \in \mathbb{R}^{d\times d}$ SPD and $x \in \mathbb{R}^d$, we let $\|x\|_C^2 \doteq x^\top C\,x$. The spectral norm of the matrix A is given by $\|A\| \doteq \sup_{x\in\mathbb{R}^d:\|x\|=1} \|Ax\|$. The spectral radius of A is $\rho(A) \doteq \max\{|\lambda| : \lambda \in \Lambda(A)\}$ where $\Lambda(A)$ is the set of (complex) eigenvalues of A. For symmetric matrices $\rho(A) = \|A\|$. We use $\kappa(A) = \|A\|\|A^{-1}\|$ to denote the condition number of a non-singular matrix A. We use $\lambda_{\min}(A)$ to denote the minimum eigenvalue of a symmetric matrix A, while $\lambda_{\max}(A)$ denotes its maximum eigenvalue. We denote the identity matrix in $\mathbb{R}^{d\times d}$ by I. We use $Z\sim P$ to denote the fact that Z (which can be a number, or vector, or matrix) is distributed according to probability distribution P; $\mathbf{E}$ denotes mathematical expectation. For $t\geq 0$ let $a_t, b_t : X\to(0,\infty)$. We write $a_t \asymp b_t$ when there exists constants $0\leq c_1\leq c_2$ such that for any $x\in X$, $c_1 a_t(x)\leq b_t(x)\leq c_2 a_t(x)$.

2 Problem Setup and Examples

In this section we define what we mean by the *constant stepsize averaged linear stochastic approximation* (CALSA) algorithm, state and discuss the assumptions under which we study CALSA and then present two instance of learning problems where CALSA is applied. A CALSA algorithm sequentially processes the data $\{(A_t,b_t)\}_{t\geq 1}\subset\mathbb{R}^{d\times d}\times\mathbb{R}^d$ to produce in round t the parameter vectors $\theta_t,\hat{\theta}_t\in\mathbb{R}^d$ using

$$\text{LSA:} \qquad \theta_t=\theta_{t-1}+\alpha(b_t-A_t\theta_{t-1})\,, \tag{2a}$$

$$\text{Average:} \qquad \hat{\theta}_t=\frac{1}{t+1}\sum\nolimits_{s=0}^{t}\theta_s\,. \tag{2b}$$

The value of the initial parameter vector $\theta_0\in\mathbb{R}^d$ and the stepsize $\alpha>0$ are left to be chosen by the user. The iterate θ_t is treated as an internal state of the algorithm, while $\hat{\theta}_t$ is the output in round t. The update of θ_t alone is considered a form of constant stepsize LSA.

The data $\{(A_t,b_t)\}_{t\geq1}$ is assumed to be an i.i.d. sequence with common distribution P. Throughout the paper we will use $\mathcal{F}_t=\sigma(A_1,b_1,\dots,A_t,b_t)$ to denote the σ-algebra summarizing the history up to and including time step $t\geq 1$ and we let $\mathcal{F}_0$ denote the trivial σ-algebra. We will also assume that $A_P\doteq\mathbf{E}[A_t]$ is nonsingular and let $\theta_*=A_P^{-1}b_P$ where $b_P\doteq\mathbf{E}[b_t]$.

We are interested in the mean squared error (MSE) at time t given by $\mathbf{E}[\|\hat{\theta}_t-\theta_*\|_Q^2]$ for some SPD matrix Q. Our assumptions concerning $\{(A_t,b_t)\}_{t\geq1}$ are as follows:

Assumption 1.

1. *$\{(A_t,b_t)\}_{t\geq1}$ is an* i.i.d. *sequence with common distribution P. We further assume that $A_P=\mathbf{E}[A_t]$ is* Hurwitz.

2. *The "noise sequences" $\{M_t\}_{t\geq1}$, $\{N_t\}_{t\geq1}$, where $M_t=A_t-A_P$ and $N_t=b_t-b_P$, have uniformly bounded second conditional moments: For some $\sigma^2_{A_P}$ and $\sigma^2_{b_P}$ constants, $\mathbf{E}[\|M_t\|^2\mid\mathcal{F}_{t-1}]\leq\sigma^2_{A_P}$, $\mathbf{E}[\|N_t\|^2\mid\mathcal{F}_{t-1}]\leq\sigma^2_{b_P}$.*

Note that $\mathbf{E}[M_t|\mathcal{F}_{t-1}]=0$ and $\mathbf{E}[N_t|\mathcal{F}_{t-1}]=0$, i.e., $\{M_t\}_{t\geq1}$ and $\{N_t\}_{t\geq1}$ are $(\mathcal{F}_t)_{t\geq0}$-adapted martingale difference sequences. In fact, this property could replace the assumption that $\{(A_t,b_t)\}_{t\geq1}$ is an i.i.d. sequence without harming our results with the exception of the results on RL where some additional assumptions would also be necessary was the i.i.d. assumption removed. We stick to the i.i.d. assumption for the sake of simplicity.

Since a Hurwitz matrix is necessarily nonsingular, A_P is nonsingular as promised. Note that the assumption that A_P is Hurwitz is necessary for the boundedness of of the iterates $\{\theta_t\}_{t\geq1}$ in any reasonable sense (e.g., in the sense that $\mathbf{E}[\|\theta_t\|^2]$ is bounded). In general, Q is allowed to be dependent on the instance P. In particular, this is the case in linear regression, which we consider next.

Example 1 (Linear regression under squared loss and bounded data (LS))**.** Let $\{(x_t,y_t)\}_{t\geq1}$ be an $\mathbb{R}^d\times\mathbb{R}$-valued i.i.d. sequence so that $\|x_t\|,|y_t|\leq B$ with some $B>0$ that is given to the algorithm designer. In linear prediction under the squared loss criterion the problem is to find $\theta_*\in\mathbb{R}^d$ such that $\theta_*=\arg\min_{\theta\in\mathbb{R}^d}L(\theta)$ with $L(\theta)=\mathbf{E}\left[(\langle x_t,\theta\rangle-y_t)^2\right]=c+\|\theta-\theta_*\|_A^2$, where $A=\mathbf{E}\left[x_t x_t^\top\right]$, where c is a constant independent of θ (but c can depend on the joint distribution of (x_t,y_t)). The *constant stepsize averaged least-mean square* (CALMS) algorithm analyzed by Bach and Moulines [1] is given by $\theta_t=\theta_{t-1}+\alpha(x_t y_t-x_t x_t^\top\theta_{t-1})$, $\hat{\theta}_t=\frac{1}{t+1}\sum_{s=0}^{t}\theta_s$.

Example 2 (Linear value-function estimation (LVE))**.** The reader interested in the background of LVE can consult, e.g., [19, 22]. In i.i.d. discounted LVE the algorithm designer is given a so-called discount factor $\gamma\in(0,1)$, while the data is an i.i.d. sequence $\{(\phi_t,\phi'_t,r_t)\}_{t\geq1}\subset\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}$ and the goal is to find a solution $\theta_*\in\mathbb{R}^d$ to the equation $A\theta=b$ where $A=\mathbf{E}\left[\phi_t(\phi_t-\gamma\phi'_t)^\top\right]$ and $b=\mathbf{E}[\phi_t r_t]$. Note that when $\gamma=0$, the equation defining θ_* is the same as $\nabla L(\theta)=0$ in LS. Hence, in this sense LVE generalizes LS. Again, it is customary to assume that the data is bounded: $\|\phi_t\|,\|\phi'_t\|,|r_t|\leq B$ almost surely (a.s.) with some known constant $B>0$. Commonly, the loss of a parameter vector $\theta\in\mathbb{R}^d$ is either measured using $\|\theta-\theta_*\|^2_{\mathbf{E}[\phi_t\phi_t^\top]}$, which can be thought of as a generalization of $L(\theta)$, or just by the unweighted 2-norm, $\|\theta-\theta_*\|^2$. While it is not the purpose of this article to discuss these choices, we note in passing that these losses are nowhere near as natural as the squared loss in LS. In this paper we consider the constant stepsize version of Sutton's TD(0) [18], and a constant stepsize version of a novel variant of the so-called GTD algorithm [21, 20]. The novelty of our variant is that it updates the parameter vector θ_t using the updates auxiliary parameter y_t, rather than using y_{t-1} as in the original version. This small change will be instru-

CATD(0)	CAGTD
$\theta_t = \theta_{t-1} + \alpha(b_t - A_t\theta_{t-1}),$ $\hat\theta_t = \frac{1}{t+1}\sum_{s=0}^{t}\theta_s$	$y_t = y_{t-1} + \beta(b_t - A_t\theta_{t-1} - y_{t-1}),$ $\theta_t = \theta_{t-1} + \alpha A_t^\top y_t$ $\hat\theta_t = \frac{1}{t+1}\sum_{s=0}^{t}\theta_s, \hat y_t = \frac{1}{t+1}\sum_{s=0}^{t} y_s,$

Table 1: In the table $A_t = \phi_t(\phi_t - \gamma\phi_t')^\top$, and $b_t = \phi_t r_t$. The left column shows the updates of CATD(0), the constant stepsize averaged TD(0) algorithm, while the right column shows the updates of CAGTD, our variant of GTD that is combined with constant stepsizes and averaging. In this paper, we let $\beta = \alpha$ in CAGTD. It is straightforward to write the algorithms in the form given in (2). Both updates can be implemented in $O(d)$ time and space.

mental for our results in Section 5. The algorithms are summarized in Table 1.

3 Instance Dependent Bounds

Let $\hat e_t = \hat\theta_t - \theta_*$ be the error of estimating θ_*. To provide the foundation for the next sections, in this section we consider instance-dependent bounds on the expected squared parameter estimation error $\mathbf{E}[\|\hat e_t\|^2]$. The consideration of weighted norms is postponed until later. While the results presented here may not have appeared in the literature in exactly the form presented here and has some novelty in dealing with the general Hurwitz case, we borrow much from previous papers (e.g., [1, 14]).

For the next result fix an instance P so that $(A_t, b_t) \sim P$. To minimize clutter, we let $A = \mathbf{E}[A_t]$ and $b = \mathbf{E}[b_t]$, dropping the subindex P from A_P and b_P. Straightforward calculation gives that $\hat e_t = \frac{1}{t+1}\sum_{s=0}^{t} e_s$, where $e_s \doteq \theta_s - \theta_*$ is the error of the sth un-averaged iterate. Further algebra shows that

$$e_t = F_{t,1}e_0 + \alpha\sum_{i=1}^{t} F_{t,i+1}\zeta_i \tag{3}$$

where for $1 \le i \le t$, $F_{t,i} = (I - \alpha A_t)(I - \alpha A_{t-1})\dots(I - \alpha A_{i+1})(I - \alpha A_i)$ and $\zeta_i = b_i - b - (A_i - A)\theta_*$ is the "noise component" in (A_i, b_i).

Focusing on $F_{t,1}e_0$, using repeated conditioning one can see that $\phi \doteq \|\mathbf{E}\left[(I - \alpha A_1)^\top(I - \alpha A_1)\right]\| < 1$ is sufficient to guarantee that $\mathbf{E}\left[\|F_{t,1}e_0\|^2\right]$ vanishes as $t \to \infty$. Here, we can use A_1, because $(A_t)_t$ is an i.i.d. sequence. Let $R \doteq \mathbf{E}\left[(I - \alpha A_1)^\top(I - \alpha A_1)\right]$. From the definition of R we see that, on the one hand, R is positive definite, while, on the other hand, $R = I - \alpha\left\{(A + A^\top) - \alpha\mathbf{E}\left[A_1^\top A_1\right]\right\}$. Hence the eigenvalues of R are all real, nonnegative and are of the form $1 - \alpha\lambda$ for $\lambda \in \Lambda(S) \subset \mathbb{R}$, $S = (A + A^\top) - \alpha\mathbf{E}\left[A_1^\top A_1\right]$. Hence, if all eigenvalues of S are positive, we have $\phi < 1$. Let P_A denote the distribution of A_1 and note that $\lambda_{\min}(S) > 0$ is equivalent to

$$\rho_s(\alpha, P_A) \doteq \lambda_{\min}((A + A^\top) - \alpha\mathbf{E}\left[A_1^\top A_1\right]) > 0\,. \tag{4}$$

Now, $\rho_s(\alpha, P_A) \ge \lambda_{\min}(A + A^\top) - \alpha\lambda_{\max}(\mathbf{E}\left[A_1^\top A_1\right])$. Thus, if $\lambda_{\min}(A + A^\top) > 0$ then for $\alpha > 0$ small enough, $\rho_s(\alpha, P_A) > 0$ is guaranteed to hold. While A being Hurwitz is insufficient to guarantee $\lambda_{\min}(A + A^\top) > 0$, one can show that every Hurwitz matrix is similar to a real matrix B such that $B + B^\top$ is SPD (cf. Appendix A.1). Now, if U is the underlying similarity transformation, so that $B = U^{-1}AU$, one can check that $z_t = U^{-1}e_t$ satisfies (3) with A_s replaced by $B_s \doteq U^{-1}A_sU$ and ζ_i replaced by $U^{-1}\zeta_i$. Let P_U to denote the common distribution of $\{B_s\}_s$. Thanks to $\mathbf{E}[B_s] = B$, the expected squared norm of the first term in the analog of (3) can be shown to be bounded by $(1 - \alpha\rho_s(\alpha, P_U))^t\|e_0\|^2$, while the expected squared norm of the second term can be shown to be bounded by $c\alpha/\rho_s(\alpha, P_U)$ with some P-dependent constant $c > 0$. One can also show that $\mathbf{E}\left[\|\hat e_t\|^2\right] \le (1 + 4/\alpha\rho_d(\alpha, P_U))\sum_{i=0}^{t}\mathbf{E}\left[\|e_i\|^2\right]$, where $\rho_d(\alpha, P_U) \doteq \lambda_{\min}\left(B + B^\top - \alpha B^\top B\right) \ge \rho_s(\alpha, P_U)$. Putting things together, we get the following result:

Theorem 1. Let P be a distribution over $\mathbb{R}^{d\times d} \times \mathbb{R}^d$ satisfying Assumption 1. Then, for any $U \in \mathbb{R}^{d\times d}$ and P_U as in the previous paragraph there exists $\alpha_{P_U} > 0$ such that for all $\alpha \in (0, \alpha_{P_U})$ and for all $t \ge 0$,

$$\mathbf{E}\left[\|\hat\theta_t - \theta_*\|^2\right] \le \nu\left\{\frac{\|\theta_0 - \theta_*\|^2}{(t+1)^2} + \frac{v^2}{t+1}\right\},$$

where $\nu = \left(1 + \frac{4}{\alpha\rho_d(\alpha, P_U)}\right)\frac{2\kappa(U)^2}{\alpha\rho_s(\alpha, P_U)}$ and $v^2 = 2\alpha^2(\sigma^2_{A_P}\|\theta_*\|^2 + \sigma^2_{b_P})$.

Thus, the MSE in round t is bounded by a sum of two terms. The first, *bias* term, is given by $\nu\,\frac{\|\theta_0 - \theta_*\|^2}{(t+1)^2}$, bounding how fast the initial error $\|\theta_0 - \theta_*\|^2$ is forgotten. The second, *variance* term, $\nu\,\frac{v^2}{t+1}$ captures the rate at which noise is rejected. Note that ν depends on U, P_U and α.

As $\alpha \to 0$, the bias term blows up, due to the presence of α^{-1} there. This is unavoidable (see also Theorem 2 below) and is due to the slow forgetting of initial conditions for small α. Small step-sizes are however useful to suppress noise, as seen from that in our bound α^2 is seen to multiply the variances $\sigma^2_{A_P}$ and $\sigma^2_{b_P}$. In quantitative terms, we can see that the α^{-2} and α^2 terms are trading off the two types of errors. As α is increased to a critical value α_{P_U}, $\rho_s(\alpha, P_U) \to 0$ and the bounds blow up again. Indeed, too large stepsizes can lead to instability, though the upper bound of Theorem 1 is a bit loose in this respect. Finally, note that one can always take U in the result that leads to the smallest bound (including a U with complex entries, in which case, the analysis goes through with appropriate technical modifications). As promised, the next result shows that the bound of Theorem 1 is tight, at least for t large and α small:

Theorem 2 (Lower Bound)**.** There exists a distribution P over $\mathbb{R}^{d\times d} \times \mathbb{R}^d$ satisfying Assumption 1 and a constant $\alpha_P > 0$ so that $(\rho_d(\alpha, P) \geq)\ \rho_s(\alpha, P) > 0$ holds for all $0 < \alpha < \alpha_P$ and for any $t \geq 0$, $\mathbf{E}\left[\|\hat{\theta}_t - \theta_*\|^2\right] \geq \frac{1}{\alpha^2\,\rho_d(\alpha,P)\rho_s(\alpha,P)} \left\{ \frac{\beta_{t+1}^2\|\theta_0-\theta_*\|^2}{(t+1)^2} + \frac{\alpha^2\sigma_{b_P}^2\sum_{s=1}^t \beta_s^2}{(t+1)^2} \right\}$, where $\beta_t = 1 - (1-\frac{1}{2}\alpha\rho_s(\alpha,P))^t$.

Note that $\beta_t \geq 1/2$ when $t \geq 2\log(2)/(\alpha\rho_s(\alpha,P))$. Thus, for such t, Theorem 2 the coefficients of both the $1/t$ and $1/t^2$ terms inside $\{\cdot\}$ match the corresponding terms of Theorem 1 (here $U = I$). While $(\rho_d(\alpha,P)\rho_s(\alpha,P))^{-1}$ appears in the lower bound, careful inspection of the proof reveals that α_P is chosen in a conservative way and as a result this term fails to blow up as α approaches α_P from below.

4 Problem Landscape

While the previous section considered individual problem instances, in this section we start to consider classes $\mathcal{P}$ of problem instances. The first question that arises then is whether for a given class $\mathcal{P}$ it is possible to find a single *universal* stepsize that guarantees that the worst-case expected squared error, $\sup_{P\in\mathcal{P}} \mathbf{E}_P[\|\hat{e}_t\|^2]$, vanishes as $t \to \infty$. Here, $\mathbf{E}_P[\cdot]$ is used to signify that the randomness underlying $\mathbf{E}[\cdot]$ is governed by the instance P.

As can be seen from Theorem 2, a universal stepsize may fail to exist for multiple reasons: First, it will fail to exist if the noise variance is not uniformly bounded, e.g., when $\sup_{P\in\mathcal{P}} \sigma_{b_P}^2 = +\infty$ (while Theorem 2 does not show it, we believe that $\sup_P \sigma_{A_P}^2 = +\infty$ will also lead to the same conclusion). Hence, in what follows, we assume that the variance is uniformly bounded; in fact, we will often assume that the data $\{(A_t, b_t)\}_{t\geq 1}$ itself is uniformly bounded. We consider this as a mild assumption. The second mode of failure is more interesting: This happens because $\mathbf{E}\left[\|F_{t,1}e_0\|^2\right]$ is uncontrolled. In fact, when $A_t = A$, $F_{t,1} = (I-\alpha A)^t$ and so a necessary condition for controlling $\|F_{t,1}\|$ is that $\rho(I - \alpha A) < 1$. A simple example this cannot be satisfied uniformly over all instances regardless of the choice of α is the case of the $\mathrm{ROT}(2,B)$ class: which we define as the class when $d = 2$, $B > 0$ is a constant, and every instance P in $\mathrm{ROT}(2,B)$ is a Dirac distribution, putting a point mass on a pair (A, b), where $\|b\|^2 \leq B$ and A is a 2×2, scaled rotation matrix: $A = \begin{bmatrix} u & v \\ -v & u \end{bmatrix}$ such that $u^2 + v^2 \leq B$ and $u > 0$. Note that $u > 0$ guarantees that A is a Hurwitz matrix.

Proposition 3. *For any $\alpha > 0$, $B > 0$,* $\sup_{P\in\mathrm{ROT}(2,B)} \rho(I - \alpha A_P) = 1 + B > 1$.

Proof. Let A be the scaled rotation matrix given by u, v as in the description of $\mathrm{ROT}(2,B)$. Since $\rho(I - \alpha A) = (1-\alpha u)^2 + v^2$, we see that as $u \to 0+$, we can let $v^2 \to B$. Thus, $\sup_{P\in\mathrm{ROT}(2,B)} \rho(I - \alpha A_P) \geq 1 + B$. $\square$

Next, we consider the following classes:

$$\begin{aligned} &\mathrm{SPD} : P \text{ is such that } A_P \text{ is SPD }, \|A_P\| \leq 1, A_t = A_P, \\ &\qquad b_P = 0, \sigma_{b_P}^2 \leq \sigma_b^2\,; \\ &\mathrm{SPDSN} : P \text{ is in SPD and in addition } \mathbf{E}\left[b_t b_t^\top\right] \preceq A_P\,. \end{aligned}$$

Here, $A \preceq B$ is $B - A$ is PSD. The abbreviation SPD stands for symmetric positive definite (the property of A_P), while SPDSN stands for symmetric positive definite with *structured noise*.

For the next result define $\varepsilon_t(\mathcal{P}) = \sup_{P\in\mathcal{P}} \mathbf{E}_P[\|\hat{e}_t\|^2]$ and $\varepsilon_t'(\mathcal{P}) = \sup_{P\in\mathcal{P}} \mathbf{E}_P[\|\hat{e}_t\|_{A_P}^2]$.

Theorem 4. Any $\alpha \in (0,1)$ is a universal stepsize for both SPD and SPDSN. Furthermore, for any fixed $\alpha \in (0,1)$, $\theta_0 \in \mathbb{R}^d$,

$$\varepsilon_t(\mathrm{SPD}) \asymp \|e_0\|^2 + \alpha^2\sigma_b^2 t\,, \quad \varepsilon_t'(\mathrm{SPD}) \asymp \frac{\|e_0\|^2}{\alpha t} + \sigma_b^2\alpha\,,$$

$$\text{and} \qquad \varepsilon_t'(\mathrm{SPDSN}) \asymp \frac{\|e_0\|^2}{\alpha t} + \frac{d}{t}\,.$$

From the result stated for $\varepsilon_t(\mathrm{SPD})$, it follows that the SPD class is too broad in the sense that although any $\alpha \in (0,1)$ leads to an asymptotic $O(1/t)$ decrease of the error, for any choice of α, the class contains an instance which makes the error grow linearly with t. Intuitively, this happens because an adversary can choose A_P to be near zero, in which case CALSA accumulates the noise due to the randomness in $\{b_t\}$. When $\alpha = 0$, the linear term would vanish, but the initial error remains.

When the error is measured with respect to the SPD matrix A_P (as is the case in LS), the worst-case error, $\varepsilon_t'(\mathrm{SPD})$ is dramatically improved. This is because the adversarial choice of letting A_P approach zero also automatically reduces the error. Note that in this case for a *fixed* time step t, the best possible ($\|e_0\|$-independent) choice for α is $\alpha = 1/(\sigma_b\sqrt{t})$

and this choice gives the error $2\left\|e_0\right\|^2 \sigma_b/\sqrt{t}$, which decreases over time.

In the structured noise case and when the error is also scaled with A_P, the worst-case error improves to scale with $1/t$. This is because here the magnitude of the noise on a per-instance bases is also constrained by A_P. Thus, scaling down A_P will not hurt the CALSA algorithm anymore.

We note in passing that the results of Bach and Moulines [1] are very similar to this last result, in that, they use weighted-norm with respect to the A_P and the *structured noise* property. In fact, our intention was to capture the effect of various properties that are available in LS instances on the error of CALSA. Furthermore, it is clear that the special structures that helped us to achieve the $O(1/t)$ worst-case rate are not present in the case of TD algorithms for LVE problems.

5 Universal stepsizes in LVE

We now turn to the question of the existence of universal stepsizes for CATD(0) and CAGTD (cf. Table 1). In what follows, we define what we call *admissibility*, a sufficient condition for the existence of a universal stepsize.

Definition 1. *Call a problem class $\mathcal{P}$* admissible *if there exists a unique U and $\alpha_{\mathcal{P}_U} > 0$ such that $\rho_s(\alpha, P_U) > 0$ holds for all $P \in \mathcal{P}$ and $\alpha \in (0, \alpha_{\mathcal{P}_U})$.*

If $\mathcal{P}$ is admissible, it follows from Theorem 1 that an asymptotic "fast" rate of $O(\frac{1}{t})$ is achieved for any $P \in \mathcal{P}$. We now define three LVE problem classes. For the definitions introduce the entrywise max-norm for matrices: $\|A\|_{\max} = \max_{i,j} |A_{ij}|$. Recall that an LVE problem is given by the joint distribution of the i.i.d. sequence $\{(\phi_t, \phi'_t, r_t)\}_{t\geq 1} \subset \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$. The classes we define are as follows:

$$\text{SOFS}(B)\text{: } \mathbf{E}[\phi_t\phi_t^\top] = \mathbf{E}[\phi'_t(\phi'_t)^\top], \|\phi_t\| \leq B\,, t \geq 1\,.$$
$$\text{GTDOFF}(B)\text{: } \|A_t\|_{\max} \leq B, t \geq 1\,.$$

Here, SOFS stands for *second-order feature stationarity*, while GTDOFF stands for "off-policy", a specific nomenclature borrow from the RL literature. Note that the "second order feature stationarity" condition $\mathbf{E}[\phi_t\phi_t^\top] = \mathbf{E}[\phi'_t(\phi'_t)^\top]$ will hold when sampling (in the underlying Markov reward process) is started from the stationary distribution. Note that the constants B appearing in the two classes constraint the data in different ways.

Theorem 5. Let $B \geq 1$. The following hold: *i)* CATD(0) has a universal stepsize of $\alpha_{td} = \frac{1}{B^2}$ for the class SOFS(B). *ii)* CAGTD has a universal stepsize of $\alpha_{gtd} = \frac{1}{2B^2 d}$ for the class GTDOFF(B).

In the proof we show that the two classes are admissible for the respective algorithms.

From Table 1, the matrix $A_t = \phi_t(\phi_t - \gamma\phi'_t)^\top$ is key to both CATD(0) and CAGTD. In the case of CATD(0), the expression for $\rho_s(\alpha, P)$ involves $\alpha_{td}A_t^\top A_t$ and under the said assumptions here we have $\alpha_{td}A_t^\top A_t \preceq Q_t \doteq (\phi_t - \gamma\phi'_t)(\phi_t - \gamma\phi'_t)^\top$. Now, the proof for the CATD(0) case, follows by using the stationary property on top of simple algebra with matrices Q_t and A_t. In the case of CAGTD, owing to its composite structure with primal and dual variables, the expression for $\rho_s(\alpha, P)$ involves $A_t^\top A_t$, $(A_t^\top A_t)^2$ and $(A_t^\top A_t)^3$, and hence the proof uses a bound on $\|A_t\|_{\max}$. This small stepsize for CAGTD seems to be the price paid for *off-policy* stability. Note that the above result in particular implies that the respective algorithms with the proposed stepsizes achieve the instance-dependent errors $O(\frac{1}{t})$ on these three classes of LVE problems.

6 Numerical Experiments

Worst-case Error: The goal here was to illustrate Theorem 4, which proved results for the behavior of the worst-case errors $\varepsilon_t(\text{SPD})$, $\varepsilon'_t(\text{SPD})$, and $\varepsilon'_t(\text{SPDSN})$. To validate this result, we chose $d = 2$ and define the classes USN ("unscaled noise") and SN ("scaled noise") as subsets of SPD and SPDSN, respectively. To define these classes let $\{u_t\}_{t\geq 1} \subset \mathbb{R}^2$ be in i.i.d. sequence so that $u_{t,1}$ and $u_{t,2}$ are also independent and they are both uniformly distributed in $[-1, 1]$. Now, P is in USN when $A_P = \begin{bmatrix} 1 & 0 \\ 0 & a_P \end{bmatrix}$ for some $a_P \in (0, 1)$, $A_t = A_P$ for all $t \geq 1$ and $b_t = u_t$. Further, P is is in SN when A_P and $\{A_t\}_{t\geq 1}$ are as in USN and $b_t = A_P u_t$. The upper left subfigure in Figure 1 shows lower bounds on $\varepsilon_t(\text{USN})$, $\varepsilon'_t(\text{USN})$, and $\varepsilon'_t(\text{SN})$ as a function of the number of rounds, or iterations. The stepsize for producing $\varepsilon_t(\text{USN})$ is chosen to be $\alpha = 0.9$, while to obtain a lower bound on $\varepsilon_t(\text{USN})$ we let $a_P = 1/t$. We can observe that the lower bound increases linearly with t. For producing a lower bound on $\varepsilon'_t(\text{USN})$, we let $a_P = 1/\sqrt{t}$ and also $\alpha = 1/\sqrt{t}$. Observe that the lower bound decreases as $1/\sqrt{t}$, as expected. Finally, to produce a lower bound on $\varepsilon'_t(\text{SN})$, we chose $\alpha = 0.9$ and $a_P = \frac{1}{t}$. The lower bound decreases as $1/t$, as expected.

Mountain Car (setup): The mountain car is a widely used domain for illustrating control learning in RL. However, here, we use it for illustrating linear value estimation only. The domain consists of an underpowered car, that needs to swing from the bottom of a valley to the top by performing either one of the three possible actions: *forward, reverse, no* throttle. Since the car is underpowered, it cannot directly accelerate to the top from the bottom and needs to swing back and forth to reach the top. The state of the system is described by the position p and the velocity v of the car at a given time. For the purpose of *on-policy* evaluation, we sample from the policy π that accelerates in the direction of the velocity with probability $\frac{298}{300}$ and the other two actions with probability $\frac{1}{300}$ each. Since, we are also interested in the *off-policy* case, we sampled using a behavior policy π_b that accelerates in the direction of the velocity with probability $\frac{8}{10}$ and chooses the other two actions with probability $\frac{1}{10}$ each. We used *tile coding* and *Fourier* basis (un-normalized and normalized). We used 4 different tiling (4×4 and 7×7 grid for the two state-variables permuted with 5 and 10 tiles), and we also tried 4 different m^{th} Fourier basis function ($m = 3, 5, 7, 9$), with $d = (m+1)^2$. For a given state $s = (p, v)$,[1] the Fourier feature is given by $\phi(p, v) = \big(\cos(\pi[c_1 p + c_2 v])\big)_{c_1, c_2 = 0,1,\ldots,m} \in \mathbb{R}^{(m+1)^2}$. The normalized features were obtained by letting $\|\phi(s)\|_2 = 1$. We generated 100 trajectories for the *on/off*-policies, and the discount factor we used was $\gamma = 0.999$.

Before discussing the observations, we digress, to mention two important aspects related to LSA algorithms, which, while being out of the scope of this paper, nevertheless are important in practice.

Singularity: In Assumption 1 we assumed that the matrix A_P is Hurwitz and hence invertible. When the underlying matrix is singular, there could be two scenarios: either $A_P\theta = b_P$ has infinitely many solutions, or it has no solutions. In the former scenario, and under a further assumption that the null-space of A_P is diagonalizable (see [25]), the null space can be discarded after applying an appropriate linear transformation U (as in Theorem 1) to a obtain a reduced linear system $\tilde{A}_P\tilde{\theta} = \tilde{b}_P$. This reduced linear system has a unique solution so that Theorem 1 applies.

Design of Updates: Note that the CATD(0) and CAGTD have different underlying linear systems. This is evident by writing down (b_P, A_P) for TD(0) and GTD respectively. Let $A_t = \phi_t(\phi_t - \gamma\phi'_t)^\top$, $b_t = r_t\phi_t$. Then, for CATD(0), $A_{TD} = \mathbf{E}[A_t]$ and $b_{TD} = \mathbf{E}[b_t]$. For CAGTD we have $A_{GTD} = \begin{bmatrix} I & A_{TD} \\ -(1-\alpha)A_{TD}^\top & \alpha A_{TD}^\top A_{TD} \end{bmatrix}$ and $b_{GTD} = [b_{TD}^\top, \alpha b_{TD}^\top A_{TD}]^\top$. For CAGTD, the eigenvalues involve $A_{TD}^\top A_{TD}$, i.e., a small eigenvalue of A_{TD} gets squared. Consequently CAGTD can be poorly conditioned compared to CATD(0).

[1] We scale the states by subtracting the minimum value and dividing it by its range, so that $p, v \in (0,1)$ after scaling.

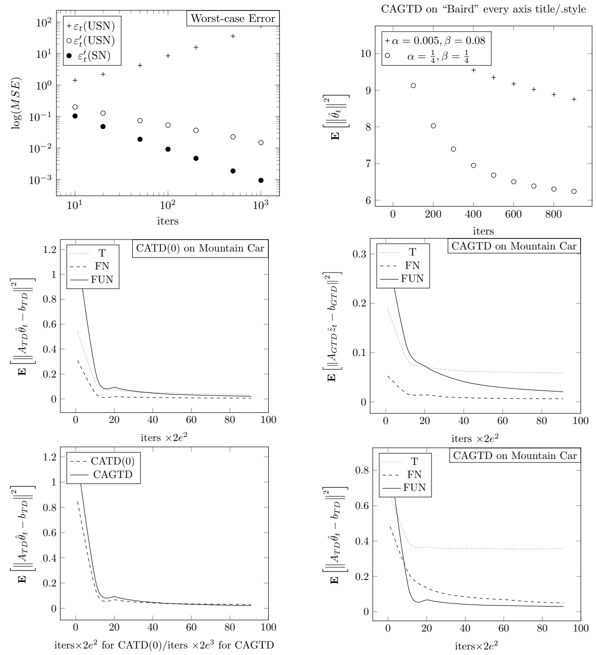

Figure 1: Experimental results. T, FUN, FN stand for tile coding, Fourier un-normalized and normalized, respectively. In the bottom left plot, the CAGTD curve is rescaled as shown on the label of the x axis.

- **Stability:** For CATD(0), we ran *on-policy* with all the three features and *off-policy* with normalized features. For CAGTD, we ran with all the features and both *on/off-policy*. In all the experiments, we chose the stepsize dictated by Theorem 5. All the experiments were stable (bottom two rows of Figure 1). The values are averaged over 10 runs and since the variance was observed to be small, to reduce clutter, error bars are not shown.

- **Near Singularity:** We observed in the case of tile coding and normalized Fourier basis functions that the underlying A_{TD} matrices were nearly singular, i.e., they had eigenvalues with positive real-parts close 0. However, we observed that the error (in the case of CATD(0)) $\mathbf{E}\left[\|A_{TD}\hat{\theta}_t - b_{TD}\|^2\right]$ converges to 0 (left plot in the second row of Figure 1). We also observed that, in the case of CAGTD, $\mathbf{E}\left[\|A_{GTD}\hat{z}_t - b_{GTD}\|^2\right]$ converges to 0 (right plot in the second row of Figure 1), where $\hat{z}_t = [\hat{y}_t^\top, \hat{\theta}_t^\top]^\top$. Here, $\hat{\theta}_t$ is the primal variable and $\hat{y}_t$ is the dual variable. However (in CAGTD), $\mathbf{E}\left[\|A_{TD}\hat{\theta}_t - b_{TD}\|^2\right]$ does not always converge to 0 (tile coding in right plot in the third row of Figure 1). This might be due to the fact that linear systems underlying CATD(0) and CAGTD are different.

- **Slowness of GTD:** Unnormalized Fourier basis were better conditioned in comparison to the other basis choices. In this case, for CATD(0) and CAGTD $\mathbf{E}\left[\|A_{TD}\hat{\theta}_t - b_{TD}\|^2\right]$ converges to 0. However, CAGTD is slower in comparison to CATD(0) (left plot in the third row of Figure 1).

BAIRD: In this domain there are $S = \{s_1, \ldots, s_7\}$ states and $A = \{a_1, a_2\}$ actions. Under, a_1 we have $p_{a_1}(s, s_1) = 1$ for all $s \in S$ and under a_2 we have $p_{a_2}(s, s') = \frac{1}{6}$ for all $s \in S, s' = 2, \ldots, 7$. The samples are collected using a behaviour policy π_b that performs action a_2 with probability $\frac{6}{7}$ and action a_1 with probability $\frac{1}{7}$, and the target policy that we are interested is π which performs action a_1 in all the states. The feature vector we chose was: $\phi(s_1) = [\frac{1}{2}\, 0\, 0\, 0\, 0\, 0\, 0\, 1]$, $\phi(s_i) = e_i + [0\, 0\, 0\, 0\, 0\, 0\, 0\, \frac{1}{2}], i = 2, \ldots, 7$, where e_i is the standard basis with i^{th} co-ordinate 1 and rest of the co-ordinates 0. Since ϕ'_t always corresponds to state 1 and is different from ϕ_t, in this example $\mathbf{E}[\phi_t \phi_t] \neq \mathbf{E}[\phi'_t {\phi'_t}^\top]$. We compared the performance of CAGTD with $\alpha = 0.005$ (and $\beta = 0.08$, see [12]) with the choice of $\alpha = \frac{1}{2\times 2}$ (2 is to normalize the features) and initial condition $\theta_0 = [1\,1\,1\,1\,1\,1\,1\,0\,1]$. The identical stepsize of $\frac{1}{4}$ performed better than choosing different stepsizes for the primal and dual variables. Please refer to the top right plot of Figure 1.

7 Related Work

Other stepsize methods: It is clear that for the LSA in (1) to be stable α_t should be non-increasing. In this paper, we showed the results for LSA with a constant stepsize and averaging of the iterates. This brings us to a brief discussion on the other two non-increasing choices for stepsize strategies namely the diminishing and adaptive strategies. An immediate choice could be $\alpha_t = \frac{1}{t}$. However, this is a poor choice because even for bounded 1-dimensional problems with no noise it leads to constant worst-case error for any finite time step t. Various heuristic methods exist for "adapting" the stepsize sequence; see, e.g., [4] and the references therein. The methods proposed do not have guarantees and given how difficult it is to experimentally validate a method whose main purpose is *robustness*, it is hard to assess how good these methods are (e.g., the method of Dabney and Barto [4] seemed to perform well in their experiments, but one can *show* that for finite state spaces it eventually settles on a constant step-size and hence it will fail to guarantee convergence).

Error bounds for GTD/TD The initial convergence analysis for GTD/GTD2/TDC was only asymptotic in nature [21, 20] with diminishing stepsizes. In the case of GTD/GTD2 diminishing stepsizes $\alpha_t = O(\frac{1}{\sqrt{t}})$, projection of iterates and PR-averaging leads to a rate of $O(\frac{1}{\sqrt{t}})$ for the prediction error $\|A_P\hat{\theta}_t - b_P\|^2$ with high probability [12]. Liu et al. [12] also suggest a new version of GTD based on stochastic mirror prox ideas, called the GTD-Mirror-Prox (GTDMP), with identical guarantees. Inspired by TD algorithms, Dalal et al. [5] provide a stochastic boundedness result, which does not even guarantee that the error vanishes as t increases.

CALSA analysis: Analysis of CALSA goes back to the work by Polyak and Juditsky [14], wherein they considered the additive noise setting i.e., $A_t = A$ for some deterministic Hurwitz matrix $A \in \mathbb{R}^{d\times d}$. A key improvement in our paper is that we consider the 'multiplicative' noise case, i.e., A_t is non-constant random matrix. To tackle the multiplicative noise we build on the newer analysis introduced by Dieuleveut et al. [7]. However, due to the generality of our setting (with Hurwitz assumption), diverging from the analysis of Dieuleveut et al., we make use of Lyapunov's equation and a similarity transformations in a critical way to prove our results.

Conclusion: Stepsize choice is critical in LSA algorithms, and especially in the case of TD algorithms. Stepsizes are often treated as hyper-parameters that need to be tuned in a problem instance specific manner. To avoid this tuning, it is desirable to choose a single *universal* stepsize rule that works for all the instances in a problem class. This paper investigated the promise of an approach called CALSA (constant stepsize averaged linear stochastic approximation), based on an idea that goes back to Ruppert [15] and Polyak and Juditsky [14]. For a given problem class, we asked *i*) whether a *universal* constant stepsize can be chosen and *ii*) whether a *uniform* rate of convergence for the MSE can be achieved, across the class. We showed that answers to these questions in general is *no*. However, we showed (under our assumptions) that any CALSA achieves an MSE of C_P/t, where the constant $C_P > 0$ is instance dependent. We then showed that TD algorithms with a problem independent universal constant stepsize and iterate averaging, achieve a problem-dependent error that decays as $O(\frac{1}{t})$ with the number of iterations t.

Acknowledgements

Part of the work was done at the University of Alberta. The authors greatly acknowledge the support of NSERC and the Alberta Innovates Technology Futures through the Alberta Machine Intelligence Institute (AMII).

References

[1] Bach, F. R. and Moulines, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$. In *NIPS*, pages 773–781.

[2] Benveniste, A., Métivier, M., and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer Verlag, New York.

[3] Borkar, V. S. (2008). *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press.

[4] Dabney, W. and Barto, A. G. (2012). Adaptive step-size for online temporal difference learning. In *AAAI*, pages 872–878.

[5] Dalal, G., Szörényi, B., Thoppe, G., and Mannor, S. (2017). Concentration bounds for two timescale stochastic approximation with applications to reinforcement learning. *arXiv preprint arXiv:1703.05376*.

[6] Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer-Verlag New York.

[7] Dieuleveut, A., Flammarion, N., and Bach, F. (2017). Harder, better, faster, stronger convergence rates for least-squares regression. *Journal of Machine Learning Research*, 18(1):3520–3570.

[8] Du, S. S., Chen, J., Li, L., Xiao, L., and Zhou, D. (2017). Stochastic variance reduction methods for policy evaluation. In Precup, D. and Teh, Y. W., editors, *ICML*, pages 1049–1058, International Convention Centre, Sydney, Australia.

[9] French, M., Szepesvári, C., and Rogers, E. (2003). *Performance of Nonlinear Approximate Adaptive Controllers*. Wiley.

[10] Gallier, J. (2010). The Schur complement and symmetric positive semidefinite (and definite) matrices. `www.cis.upenn.edu/~jean/schur-comp.pdf`.

[11] Konda, V. R. and Tsitsiklis, J. N. (2003). Linear stochastic approximation driven by slowly varying Markov chains. *Systems & Control Letters*, 50(2):95–102.

[12] Liu, B., Liu, J., Ghavamzadeh, M., Mahadevan, S., and Petrik, M. (2015). Finite-sample analysis of proximal gradient td algorithms. In *UAI*, pages 504–513.

[13] Ljung, L., Pflug, G., and Walk, H. (1992). *Stochastic approximation and optimization of random systems*. Birkhäuser.

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

[15] Ruppert, D. (1988). Efficient estimations from a slowly convergent Robbins-Monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

[16] Shamir, O. (2015). The sample complexity of learning linear predictors with the squared loss. *Journal of Machine Learning Research*, 16:3475–3486.

[17] Solo, V. and Kong, X. (1994). *Adaptive signal processing algorithms: stability and performance*. Prentice-Hall, Inc.

[18] Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Machine Learning*, 3(1):9–44.

[19] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT press.

[20] Sutton, R. S., Maei, H. R., Precup, D., Bhatnagar, S., Silver, D., Szepesvári, C., and Wiewiora, E. (2009a). Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *ICML*, pages 993–1000.

[21] Sutton, R. S., Maei, H. R., and Szepesvári, C. (2009b). A convergent $O(n)$ temporal-difference algorithm for off-policy learning with linear function approximation. In *NIPS*, pages 1609–1616.

[22] Szepesvári, Cs. (2010). *Algorithms for Reinforcement Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.

[23] Tsitsiklis, J. N. and Van Roy, B. (1997). An analysis of temporal difference learning with function approximation. *IEEE Transactions on Automatic Control*, 42:674–690.

[24] Tsitsiklis, J. N. and Van Roy, B. (2002). On average versus discounted reward temporal-difference learning. *Machine Learning*, 49(2):179–191.

[25] Wang, M. and Bertsekas, D. P. (2013). Stabilization of stochastic iterative methods for singular and nearly singular linear systems. *Mathematics of Operations Research*, 39(1):1–30.

[26] Woodruff, D. P. (2014). Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1–2):1–157.

A Proofs for Section 3

A.1 Hurwitz matrices and positive definiteness

The symmetric part of a real-valued matrix A is $(A+A^\top)/2$. The following lemma states that Hurwitz matrices are similar to some matrix whose symmetric part is positive definite. After the lemma we give an example that shows that the similarity cannot be chosen to be the identity matrix.

Lemma 6. Any Hurwitz matrix is similar to a real matrix whose symmetric part is positive definite. Furthermore, the similarity transformation can be chosen to be an SPD matrix.

Proof. Recall that if $A \in \mathbb{R}^{d\times d}$ is Hurwitz, then there exist a unique SPD matrix P such that the Lyapunov equation

$$A^\top P + PA = I$$

is satisfied (e.g., Lemma A.23 of French et al. [9]). Take $U = P^{1/2}$. Clearly, U is positive definite and symmetric. Let $B = U^{-1}AU$. Twice the symmetric part of B is

$$B + B^\top = P^{-1/2}AP^{1/2} + P^{1/2}A^\top P^{-1/2} = P^{-1/2}(AP + PA^\top)P^{-1/2} = P^{-1}\,,$$

which is positive definite, hence finishing the proof. $\square$

Now, consider the matrix

$$A = \begin{pmatrix} 1 & -10 \\ 0 & 1 \end{pmatrix} .$$

We claim that this is a Hurwitz matrix, but $A + A^\top$ is not positive definite. Indeed, A has a single eigenvalue of multiplicity two, which is equal to one. Thus, A is Hurwitz. On the other hand,

$$A + A^\top = \begin{pmatrix} 2 & -10 \\ -10 & 2 \end{pmatrix} .$$

The eigenvalues of this matrix solve $(2-\lambda)^2 - 100 = 0$, i.e., they are $\lambda_1 = 12$ and $\lambda_2 = -8$. Hence, $A + A^\top$ is indeed not SPD.

We note in passing that the change of coordinates argument that we follow here is inspired by the paper of Du et al. [8].

A.2 Error bound when $\rho_s(\alpha, P_A) \geq 0$

In this section we show a bound on the MSE, $\mathbf{E}[\|\hat{e}_t\|^2]$, for the case when $\rho_s(\alpha, P_A) \geq 0$. Following the suggestion in the main body of the paper, the general case will be reduced to this case by using Lemma 6. The relations presented below that connect expressions that involve conditional expectations hold almost surely (a.s.). However, to minimize clutter, the "a.s." qualifier will be omitted from their statement.

Let $b = b_P$. Recall that

$$e_t = \theta_t - \theta_*\,, \qquad \hat{e}_t = \hat{\theta}_t - \theta_*\,, \qquad \text{and} \qquad \zeta_t = b_t - b - (A_t - A)\theta_*\,.$$

Further, introduce

$$F_{i,j} = \begin{cases} (I-\alpha A_i)(I-\alpha A_{i-1})\dots(I-\alpha A_j)\,, & \text{if } i \geq j\,; \\ I\,, & \text{otherwise}\,. \end{cases}$$

Let σ_1^2 be so that $\mathbf{E}\left[\|\zeta_t\|^2\right] \leq \sigma_1^2$. Note that we can always choose $\sigma_1^2 = 2(\sigma_{A_P}^2 \|\theta_*\|^2 + \sigma_{b_P}^2)$.

Error Recursion Recall that $\theta_t = \theta_{t-1} + \alpha\big(b_t - A_t\theta_{t-1}\big)$. Subtracting θ_* from both sides, we get

$$\theta_t - \theta_* = \theta_{t-1} - \theta_* + \alpha\big(b_t - A_t(\theta_{t-1} - \theta_* + \theta_*)\big)\,,$$

and substituting e_t and using $A\theta_* - b = 0$ we get

$$\begin{aligned} e_t &= (I-\alpha A_t)e_{t-1} + \alpha(b_t - A_t\theta_*) = (I-\alpha A_t)e_{t-1} + \alpha(b_t - b - (A_t - A)\theta_*) \\ &= (I-\alpha A_t)e_{t-1} + \alpha\zeta_t\,. \end{aligned} \tag{5}$$

Unrolling this for $t-1, t-2, \ldots, 1$, we get

$$e_t = F_{t,1}e_0 + \alpha\sum_{i=1}^{t} F_{t,i+1}\zeta_i\,, \tag{6}$$

which matches (3). Now, $\hat{e}_t = \frac{1}{t+1}\sum_{i=0}^{t} e_i$ and hence

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right] = \frac{1}{(t+1)^2}\sum_{i,j=0}^{t}\mathbf{E}\left[\langle e_i, e_j\rangle\right]\,. \tag{7}$$

To bound the terms $\mathbf{E}\left[\langle e_i, e_j\rangle\right]$ the following lemma that uses the martingale structure of the noise will be useful:

Lemma 7. The following hold:

(i) Let $1 \le i < t$ and let $x, y \in \mathbb{R}^d$ be $\mathcal{F}_i$-measurable random vectors. Then,

$$\mathbf{E}[x^\top F_{t,i+1}y|\mathcal{F}_i] = x^\top(I-\alpha A)^{t-i}y\,.$$

(ii) Let $1 \le i < t$ and let $x \in \mathbb{R}^d$ be an $\mathcal{F}_{i-1}$-measurable random vector. Then, $\mathbf{E}[x^\top F_{t,i+1}\zeta_i] = 0$.

(iii) For all $0 \le i < t$, $\mathbf{E}[\langle e_i, F_{t,i+1}e_i\rangle] = \mathbf{E}[\langle e_i, (I-\alpha A)^{t-i}e_i\rangle]$.

(iv) Let $0 \le j < i$ and let $x \in \mathbb{R}^d$ be an $\mathcal{F}_j$-measurable random vector. Then,

$$\mathbf{E}[\langle F_{i,j+1}x, F_{i,j+1}x\rangle] \le (1-\alpha\rho_s(\alpha, P_A))^{i-j}\,\mathbf{E}\,\|x\|^2\,.$$

Proof. We start by noting that for arbitrary t, i, $F_{t,i}$ is $\mathcal{F}_t$-measurable. Indeed, this holds because $F_{t,i}$ is a function of $\{A_s\}_{\min(t,i)\le s\le t}$ only.

Part *(i)*: By definition $F_{t,i+1} = (I-\alpha A_t)F_{t-1,i+1}$. By our note above, $F_{t-1,i+1}$ is $\mathcal{F}_{t-1}$-measurable. Since x and y are also $\mathcal{F}_{t-1}$-measurable, we get

$$\mathbf{E}\left[x^\top F_{t,i+1}y|\mathcal{F}_{t-1}\right] = x^\top\mathbf{E}\left[(I-\alpha A_t)|\mathcal{F}_{t-1}\right]F_{t-1,i+1}y = x^\top(I-\alpha A)F_{t-1,i+1}y\,.$$

If $t-1 = i$, the proof is done. When $t-1 \ge i+1$, by the tower-rule for conditional expectations and our measurability assumptions,

$$\mathbf{E}\left[x^\top F_{t,i+1}y|\mathcal{F}_{t-2}\right] = x^\top(I-\alpha A)\mathbf{E}\left[F_{t-1,i+1}|\mathcal{F}_{t-2}\right]y = x^\top(I-\alpha A)^2F_{t-2,i+1}y\,.$$

Continuing this way we get

$$\mathbf{E}\left[x^\top F_{t,i+1}y|\mathcal{F}_{t-j}\right] = x^\top(I-\alpha A)^jF_{t-j,i+1}y\,,\quad j = 1, 2, \ldots, t-i\,.$$

Specifically, for $j = t-i$ we get

$$\mathbf{E}\left[x^\top F_{t,i+1}y|\mathcal{F}_i\right] = x^\top(I-\alpha A)^{t-i}y\,.$$

Part *(ii)*: By part (i),

$$\mathbf{E}\left[x^\top F_{t,i+1}\zeta_i|\mathcal{F}_i\right] = x^\top(I-\alpha A)^{t-i}\zeta_i\,.$$

Using the tower rule,

$$\mathbf{E}\left[x^\top F_{t,i+1}\zeta_i|\mathcal{F}_{i-1}\right] = x^\top(I-\alpha A)^{t-i}\mathbf{E}\left[\zeta_i|\mathcal{F}_{i-1}\right] = 0\,.$$

Part (*iii*): The result follows directly from part (i). Indeed, θ_i depends only on $A_1,\dots,A_i,b_1,\dots,b_i$ and so e_i is $\mathcal{F}_i$-measurable. Hence, by part (i),

$$\mathbf{E}\left[\langle e_i, F_{t,i+1}e_i\rangle|\mathcal{F}_i\right] = \mathbf{E}\left[\langle e_i,(I-\alpha A)^{t-i}e_i\rangle|\mathcal{F}_i\right]\,.$$

Taking expectation of both sides gives the desired result.

Part (*iv*): Note that $S_t \doteq \mathbf{E}\left[(I-\alpha A_t)^\top(I-\alpha A_t)|\mathcal{F}_{t-1}\right] = I-\alpha(A^\top+A)+\alpha^2\mathbf{E}\left[A_t^\top A_t|\mathcal{F}_{t-1}\right]$. Since $\{A_t\}_{t\ge 1}$ is an i.i.d. sequence, $\mathbf{E}\left[A_t^\top A_t|\mathcal{F}_{t-1}\right] = \mathbf{E}\left[A_1^\top A_1\right]$. Now, using the definition of $\rho_s(\alpha,P_A)$ (cf. Equation (4)), $\sup_{x:\|x\|=1} x^\top S_t x = 1-\alpha\inf_{x:\|x\|=1}x^\top(A^\top+A-\alpha\mathbf{E}\left[A_1^\top A_1\right])x = 1-\alpha\rho_s(\alpha,P_A)$. Hence,

$$\begin{aligned}\mathbf{E}\left[\langle F_{i,j+1}x, F_{i,j+1}x\rangle|\mathcal{F}_{i-1}\right] &= \mathbf{E}\left[x^\top F_{i-1,j+1}^\top(I-\alpha A_i)^\top(I-\alpha A_i)F_{i-1,j+1}x\,|\,\mathcal{F}_{i-1}\right]\\ &= (xF_{i-1,j+1})^\top\, S_i\, F_{i-1,j+1}x\\ &\le (1-\alpha\rho_s(\alpha,P_A))\,\langle F_{i-1,j+1}x, F_{i-1,j+1}x\rangle\end{aligned}$$

When $i=j-1$, $F_{i-1,j+1}=I$ and the proof is done by taking expectations of both sides. Otherwise, $i>j-1$ and repeating the same calculation as above and using the tower rule for conditional expectations we get

$$\begin{aligned}\mathbf{E}\left[\langle F_{i,j+1}x, F_{i,j+1}x\rangle|\mathcal{F}_{i-2}\right] &= \mathbf{E}\left[\mathbf{E}\left[\langle F_{i,j+1}x, F_{i,j+1}x\rangle|\mathcal{F}_{i-1}\right]|\mathcal{F}_{i-2}\right]\\ &\le (1-\alpha\rho_s(\alpha,P_A))\,\mathbf{E}\left[\langle F_{i-1,j+1}x, F_{i-1,j+1}x\rangle|\mathcal{F}_{i-2}\right]\\ &\le (1-\alpha\rho_s(\alpha,P_A))^2\,\langle F_{i-2,j+1}x, F_{i-2,j+1}x\rangle\,.\end{aligned}$$

If $i=j-2$, $F_{i-2,j+1}=I$ and taking expectations of both sides the proof is finished. Continuing the same way we get

$$\mathbf{E}\left[\langle F_{i,j+1}x, F_{i,j+1}x\rangle|\mathcal{F}_j\right] \le (1-\alpha\rho_s(\alpha,P_A))^{j-i}\,\|x\|^2\,.$$

Taking expectations of both sides finishes the proof. □

Returning to bounding $\mathbf{E}[\|\hat{e}_t\|^2]$ based on (7). We now show that the contribution of the "cross-terms", $\sum_{i=0}^t\sum_{j=i+1}^t \mathbf{E}\left[\langle e_i,e_j\rangle\right]$, on the right-hand side of (7) is proportional to the contribution of the "diagonal terms", $\sum_{i=0}^t\mathbf{E}\left[\langle e_i,e_i\rangle\right]$. For the bound, recall that $\rho_d(\alpha,P_A) = \lambda_{\min}\left(A+A^\top-\alpha A^\top A\right)$.

Lemma 8. Assume that $\rho_d(\alpha,P_A)\ge 0$. Then it holds that

$$\sum_{i=0}^{t-1}\sum_{j=i+1}^{t}\mathbf{E}\left[\langle e_i,e_j\rangle\right] \le \frac{2}{\alpha\rho_d(\alpha,P_A)}\sum\nolimits_{i=0}^t\mathbf{E}[\|e_i\|^2]\,.$$

Proof. Let $j>i$. Then,

$$\begin{aligned}\mathbf{E}\left[\langle e_i,e_j\rangle\right] &= \mathbf{E}\langle e_i, F_{j,i+1}e_i+\alpha\textstyle\sum_{k=i+1}^j F_{j,k+1}\zeta_k\rangle\\ &= \mathbf{E}\langle e_i, F_{j,i+1}e_i\rangle && \text{(from Lemma 7(}ii\text{))}\\ &= \mathbf{E}\langle e_i,(I-\alpha A)^{j-i}e_i\rangle\,. && \text{(from Lemma 7(}iii\text{))}\end{aligned}$$

Now, from the definition of spectral norms, $\langle e_i,(I-\alpha A)^{j-i}e_i\rangle \le \|e_i\|^2\left\|(I-\alpha A)^{j-i}\right\| \le \|e_i\|^2\|I-\alpha A\|^{j-i}$. By definition, $\rho_d(\alpha,P_A) = \lambda_{\min}\left(A+A^\top-\alpha A^\top A\right)$. Using this, and the definition of spectral norms,

$$\|I-\alpha A\|^2 = \sup_{x:\|x\|=1}\|(I-\alpha A)x\|^2 = \sup_{x:\|x\|=1}x^\top(I-\alpha A)^\top(I-\alpha A)x = 1-\alpha\rho_d(\alpha,P_A)\,.$$

Note that $1-\alpha\rho_d(\alpha,P_A)\geq 0$ as the left-hand side of the above equality is obviously nonnegative. Now, by the elementary inequality $\sqrt{1-x}\leq 1-\frac{x}{2}$ which holds as long as $1-x\geq 0$, $\|I-\alpha A\|\leq 1-\frac{\alpha\rho_d(\alpha,P_A)}{2}$. Putting things together,

$$\langle e_i,(I-\alpha A)^{j-i}e_i\rangle \leq \left(1-\tfrac{\alpha\rho_d(\alpha,P_A)}{2}\right)^{j-i}\|e_i\|^2 .$$

Thus, using $\sum_{i=1}^{\infty}\gamma^i=\gamma/(1-\gamma)\leq 1/(1-\gamma)$ which holds for any $\gamma\in(0,1)$ and since by assumption $\rho_d(\alpha,P_A)\geq 0$,

$$\begin{aligned}\sum_{i=0}^{t-1}\sum_{j=i+1}^{t}\mathbf{E}\left[\langle e_i,e_j\rangle\right] &= \sum_{i=0}^{t-1}\sum_{j=i+1}^{t}\left(1-\frac{\alpha\rho_d(\alpha,P_A)}{2}\right)^{j-i}\mathbf{E}[\|e_i\|^2] \leq \sum_{i=0}^{t-1}\mathbf{E}[\|e_i\|^2]\sum_{j=i+1}^{\infty}\left(1-\frac{\alpha\rho_d(\alpha,P_A)}{2}\right)^{j-i}\\ &\leq \sum_{i=0}^{t-1}\mathbf{E}[\|e_i\|^2]\frac{2}{\alpha\rho_d(\alpha,P_A)} = \frac{2}{\alpha\rho_d(\alpha,P_A)}\sum_{i=0}^{t-1}\mathbf{E}[\|e_i\|^2] \leq \frac{2}{\alpha\rho_d(\alpha,P_A)}\sum_{i=0}^{t}\mathbf{E}[\|e_i\|^2],\end{aligned}$$

finishing the proof. □

Theorem 9. Assume that $\rho_s(\alpha,P_A)\geq 0$. Then,

$$\mathbf{E}[\|\hat{e}_t\|^2]\leq\left(1+\frac{4}{\alpha\rho_d(\alpha,P_A)}\right)\frac{2}{\alpha\rho_s(\alpha,P_A)}\left(\frac{\|e_0\|^2}{(t+1)^2}+\frac{\alpha^2\,\sigma_1^2}{t+1}\right). \tag{8}$$

Proof. As it was noted earlier, $\rho_d(\alpha,P_A)\geq\rho_s(\alpha,P_A)$. Hence, it follows that $\rho_d(\alpha,P_A)\geq 0$ holds, too and we can use Lemma 8. Combining (7) and Lemma 8 we get

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right]\leq\left(1+\frac{4}{\alpha\rho_d(\alpha,P_A)}\right)\frac{1}{(t+1)^2}\sum_{i=0}^{t}\mathbf{E}[\|e_i\|^2].$$

Expanding $\|e_i\|^2$ using (6), using $(a+b)^2\leq 2a^2+2b^2$ which holds for any $a,b\in\mathbb{R}$,

$$\begin{aligned}\mathbf{E}[\|e_i\|^2] &\leq 2\mathbf{E}\left[\langle F_{i,1}e_0,F_{i,1}e_0\rangle\right]+2\alpha^2\sum_{j=1}^{i}\mathbf{E}\left[\langle F_{i,j+1}\zeta_j,F_{i,j+1}\zeta_j\rangle\right]\\ &\leq 2(1-\alpha\rho_s(\alpha,P_A))^i\|e_0\|^2+2\alpha^2\frac{\sigma_1^2}{\alpha\rho_s(\alpha,P_A)},\end{aligned}$$

where the second inequality follows from Lemma 7(iv), $\mathbf{E}[\|\zeta_j\|^2]\leq\sigma_1^2$ and from $\sum_{j=0}^{i-1}(1-\alpha\rho_s(\alpha,P_A))^j\leq\frac{1}{\alpha\rho_s(\alpha,P_A)}$. Hence, using again the sum of geometric series,

$$\sum_{i=0}^{t}\mathbf{E}[\|e_i\|^2]\leq\frac{2\alpha^2\sigma_1^2}{\alpha\rho_s(\alpha,P_A)}(t+1)+2\|e_0\|^2\sum_{i=0}^{t}(1-\alpha\rho_s(\alpha,P_A))^i\leq\frac{2\alpha^2\sigma_1^2}{\alpha\rho_s(\alpha,P_A)}(t+1)+\frac{2\|e_0\|^2}{\alpha\rho_s(\alpha,P_A)}.$$

Putting things together,

$$\mathbf{E}[\|\hat{e}_t\|^2]\leq\left(1+\frac{4}{\alpha\rho_d(\alpha,P_A)}\right)\frac{2}{\alpha\rho_s(\alpha,P_A)}\left(\frac{\|e_0\|^2}{(t+1)^2}+\frac{\alpha^2(\sigma_1^2)}{t+1}\right).$$

□

A.3 The case of Hurwitz A

To prove Theorem 1 we first show the existence of a nonsingular matrix $U\in\mathbb{R}^{d\times d}$ and a constant α_{P_U} which depends only on the distribution P_U of $U^{-1}A_tU$ such that for all $\alpha\in(0,\alpha_{P_U})$, $\alpha\rho_s(\alpha,P_U)>0$. This will be used together with a change of basis to finish the proof.

We start with a lemma that shows that if $A+A^\top$ is PD then $U=I$ works with a suitable constant $\alpha_P>0$:

Lemma 10. Assume that $A + A^\top$ is PD. Then there exists a constant $\alpha_P > 0$ such that $\rho_s(\alpha, P_A) > 0$ holds for all $\alpha \in (0, \alpha_P)$.

Proof. We compute

$$\begin{aligned}
\rho_s(\alpha, P_A) &= \inf_{x:\|x\|=1} x^\top(A^\top + A)x - \alpha x^\top \mathbf{E}\left[A_t^\top A_t\right] x && \text{(by Equation (4))}\\
&= \inf_{x:\|x\|=1} x^\top(A^\top + A)x - \alpha x^\top A^\top A\, x - \alpha x^\top \mathbf{E}\left[M_t^\top M_t\right] x && \text{(using } A_t = A + M_t, \mathbf{E}\left[M_t\right] = 0)\\
&\geq \lambda_{\min}(A^\top + A) - \alpha(\|A\|^2 + \sigma_{A_P}^2)\,,
\end{aligned}$$

where the last inequality follows from optimizing x separately for all the three term, $\lambda_{\min}(M) = \inf_{x:\|x\|=1} x^\top M x$ which holds for any real symmetric matrix M, the definition of the spectral norm and that $\sup_{x:\|x\|=1} x^\top \mathbf{E}\left[M_t^\top M_t\right] x \leq \mathbf{E}\left[\sup_{x:\|x\|=1} x^\top M_t^\top M_t x\right] = \mathbf{E}\left[\sup_{x:\|x\|=1} \|M_t x\|^2\right] = \mathbf{E}\left[\|M_t\|^2\right] \leq \sigma_{A_P}^2$. The proof is complete by choosing $\alpha_P = \frac{\lambda_{\min}(A^\top + A)}{\|A\|^2 + \sigma_A^2}$ and noting that α_P is positive because by assumption $\lambda_{\min}(A^\top + A) > 0$. ☐

Now assume that A is Hurwitz. By Lemma 6 there exists a nonsingular matrix $U \in \mathbb{R}^{d\times d}$ such that $B = U^{-1}AU$ is so that $B + B^\top$ is PD. Choose a matrix U with this properties. Let $B_t = U^{-1}A_t U$ and let P_U be the common distribution of $\{B_t\}_{t\geq 1}$. Note that $\mathbf{E}\left[B_t\right] = B$. Premultiplying both equations in (2) by U^{-1}, we get

$$\text{LSA:} \qquad U^{-1}\theta_t = U^{-1}\theta_{t-1} + \alpha(U^{-1}b_t - B_t U^{-1}\theta_{t-1})\,, \tag{9a}$$

$$\text{Average:} \qquad U^{-1}\hat{\theta}_t = \frac{1}{t+1}\sum\nolimits_{s=0}^{t} U^{-1}\theta_s\,. \tag{9b}$$

Defining $\gamma_t = U^{-1}\theta_t$ and $\hat{\gamma}_t = U^{-1}\hat{\theta}_t$, we see that

$$\gamma_t = \gamma_{t-1} + \alpha(U^{-1}b_t - B_t\gamma_{t-1})\,, \qquad\qquad \hat{\gamma}_t = \frac{1}{t+1}\sum\nolimits_{s=0}^{t} \gamma_s\,,$$

which is a CALSA update applied to the data $\{(B_t, U^{-1}b_t)\}_{t\geq 1}$. Since $B + B^\top$ is PD, by Lemma 10 there exist a constant $\alpha_{P_U} > 0$ such that for any $\alpha \in (0, \alpha_{P_U})$, $\rho_s(\alpha, P_U) > 0$. Take any such α.

Defining $\gamma_* = U^{-1}\theta_*$, we see that $B\gamma^* = U^{-1}b$. Let $z_t = \gamma_t - \gamma_*$ and $\hat{z}_t = \hat{\gamma}_t - \gamma_*$. Define σ_z^2 to be an upper bound on $\mathbf{E}\left[\left\|U^{-1}(b_t - b) + (B_t - B)\gamma_*\right\|^2\right]$. Note that $U^{-1}(b_t - b) + (B_t - B)\gamma_* = U^{-1}\left\{(b_t - b) + (A_t - A)\theta_*\right\}$, hence, we can choose $\sigma_z^2 = \left\|U^{-1}\right\|^2 \sigma_1^2$, where σ_1^2 is an upper bound on $\mathbf{E}\left[\|(b_t - b) + (A_t - A)\theta_*\|^2\right]$. Since $\rho_s(\alpha, P_U) > 0$, by Theorem 9 it holds that

$$\mathbf{E}[\|\hat{z}_t\|^2] \leq \left(1 + \frac{4}{\alpha\rho_d(\alpha, P_U)}\right) \frac{2}{\alpha\rho_s(\alpha, P_U)} \left(\frac{\|z_0\|^2}{(t+1)^2} + \frac{\alpha^2\,\sigma_z^2}{t+1}\right).$$

Furthermore,

$$\mathbf{E}[\|\hat{e}_t\|^2] = \mathbf{E}[\|U\hat{z}_t\|^2] \leq \|U\|^2\,\mathbf{E}[\|\hat{z}_t\|^2]\,, \qquad \text{and} \qquad \|z_0\|^2 \leq \left\|U^{-1}\right\|^2 \|e_0\|^2\,,$$

Putting things together we get

$$\mathbf{E}[\|\hat{e}_t\|^2] \leq \kappa(U)^2 \left(1 + \frac{4}{\alpha\rho_d(\alpha, P_U)}\right) \frac{2}{\alpha\rho_s(\alpha, P_U)} \left(\frac{\|e_0\|^2}{(t+1)^2} + \frac{\alpha^2\,\sigma_1^2}{t+1}\right),$$

which gives the desired result:

Theorem 1. Let P be a distribution over $\mathbb{R}^{d\times d} \times \mathbb{R}^d$ satisfying Assumption 1. Then, for any $U \in \mathbb{R}^{d\times d}$ and P_U as in the previous paragraph there exists $\alpha_{P_U} > 0$ such that for all $\alpha \in (0, \alpha_{P_U})$ and for all $t \geq 0$,

$$\mathbf{E}\left[\|\hat{\theta}_t - \theta_*\|^2\right] \leq \nu \left\{\frac{\|\theta_0 - \theta_*\|^2}{(t+1)^2} + \frac{v^2}{t+1}\right\},$$

where $\nu = \left(1 + \frac{4}{\alpha\rho_d(\alpha, P_U)}\right) \frac{2\kappa(U)^2}{\alpha\rho_s(\alpha, P_U)}$ and $v^2 = 2\alpha^2(\sigma_{A_P}^2\,\|\theta_*\|^2 + \sigma_{b_P}^2)$.

Theorem 2 (Lower Bound)**.** There exists a distribution P over $\mathbb{R}^{d\times d}\times\mathbb{R}^d$ satisfying Assumption 1 and a constant $\alpha_P>0$ so that $(\rho_d(\alpha,P)\geq)\ \rho_s(\alpha,P)>0$ holds for all $0<\alpha<\alpha_P$ and for any $t\geq 0$, $\mathbf{E}\left[\|\hat{\theta}_t-\theta_*\|^2\right]\geq \frac{1}{\alpha^2\,\rho_d(\alpha,P)\rho_s(\alpha,P)}\left\{\frac{\beta_{t+1}^2\|\theta_0-\theta_*\|^2}{(t+1)^2}+\frac{\alpha^2\sigma_{b_P}^2\sum_{s=1}^t\beta_s^2}{(t+1)^2}\right\}$, where $\beta_t=1-(1-\frac{1}{2}\alpha\rho_s(\alpha,P))^t$.

Proof. Let $d=2$ and let $A=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}$, for some $\lambda_2>\lambda_1>0$ and let $\{N_t\}_{t\geq 1}\subset\mathbb{R}$ be an i.i.d. sequence of zero-mean random variables with $\sigma^2=\mathbf{E}\left[N_t^2\right]<+\infty$. Consider the LSA with $(A_t,b_t)\sim P$ such that $A_t=A$ for all $t\geq 1$ and $b_t=(N_t,0)^\top$. Note that in this example $\theta_*=0$, $\sigma_{A_P}^2=0$, $A_P=A$ and $\sigma_{b_P}^2=\sigma^2$. We calculate $\hat{e}_t$ to be

$$\begin{aligned}\hat{e}_t&=\frac{1}{t+1}\sum_{s=0}^t e_t=\frac{1}{t+1}\sum_{s=0}^t(I-\alpha A)^{t-s}e_0+\alpha\sum_{s=1}^t\sum_{i=s}^t(I-\alpha A)^{i-s}b_s\\&=\frac{1}{t+1}(\alpha A)^{-1}\left[\left(I-(I-\alpha A)^{t+1}\right)e_0+\alpha\sum_{s=1}^t\left(I-(I-\alpha A)^{t+1-s}\right)b_s\right].\end{aligned}$$

Denote by x_i the ith entry of a vector i. Owning to the diagonal structure of A and that $b_{s,2}=0$, the two components, $\hat{e}_{t,1}$ and $\hat{e}_{t,2}$ of the vector $\hat{e}_t$ are calculated to satisfy

$$\begin{aligned}\hat{e}_{t,1}&=\frac{1}{t+1}\frac{1-(1-\alpha\lambda_1)^{t+1}}{\alpha\lambda_1}e_{0,1}+\frac{1}{t+1}\sum_{s=1}^t(1-(1-\alpha\lambda_1)^{t+1-s})N_s\,,\\\hat{e}_{t,2}&=\frac{1}{t+1}\frac{1-(1-\alpha\lambda_2)^{t+1}}{\alpha\lambda_2}e_{0,2}\,.\end{aligned}$$

Now, choosing $U=I$, $\rho_s(\alpha,P_U)=\rho_s(\alpha,P_A)=\min(\lambda_1(2-\alpha\lambda_1),\lambda_2(2-\alpha\lambda_2))$ which is positive as long as $0<\alpha\lambda_2<2$ (which, in this case, also happens to be the necessary condition for the errors to decay). Setting $\lambda\doteq\lambda_1=\lambda_2$, $\alpha_P=1/\lambda$ we see that $\rho_s(\alpha,P_A)=\rho_d(\alpha,P_A)=\lambda(2-\alpha\lambda)$ (the reason for choosing α_P this way will become clear later). Thanks to $\mathbf{E}\left[N_s\right]=0$,

$$\mathbf{E}\left[\hat{e}_{t,1}^2\right]=\frac{1}{(t+1)^2}\left(\frac{1-(1-\alpha\lambda)^{t+1}}{\alpha\lambda}\right)^2e_{0,1}^2+\frac{\sigma^2\sum_{s=1}^t(1-(1-\alpha\lambda)^{t+1-s})^2}{(t+1)^2}\,,$$

while

$$\mathbf{E}\left[\hat{e}_{t,2}^2\right]=\hat{e}_{t,2}^2=\frac{1}{(t+1)^2}\left(\frac{1-(1-\alpha\lambda)^{t+1}}{\alpha\lambda}\right)^2e_{0,2}^2\,.$$

For $0<\alpha<\alpha_P$, $\alpha\lambda\leq\alpha\rho_s(\alpha,P_A)\leq 2\alpha\lambda$, hence

$$\begin{aligned}1-(1-\alpha\lambda)^i&\geq 1-\left(1-\tfrac{1}{2}\alpha\rho_s(\alpha,P_A)\right)^i\,,\quad\alpha\in(0,\alpha_P),\ i\geq 0\\(\alpha\lambda)^{-1}&\geq(\alpha\rho_s(\alpha,P_A))^{-1}\,.\end{aligned}$$

Thus, defining $\beta_i=1-\left(1-\frac{1}{2}\alpha\rho_s(\alpha,P_A)\right)^i$,

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right]=\mathbf{E}\left[\hat{e}_{t,1}^2\right]+\mathbf{E}\left[\hat{e}_{t,2}^2\right]\geq\frac{1}{(t+1)^2}\left(\frac{\beta_{t+1}}{\alpha\rho_s(\alpha,P_A)}\right)^2\|e_0\|^2+\frac{\sigma^2\sum_{s=1}^t\beta_s^2}{(t+1)^2}\,,$$

finishing the proof. □

Note that it may seem silly to consider $d=2$ and use a diagonal matrix with identical elements. The reason we considered $d=2$ above is because the proof also leads to the stronger conclusion that

$$\inf_P\sup_{\alpha:\alpha\rho_s(\alpha,P_A)>0}\alpha\rho_s(\alpha,P_A)=0\,,\tag{10}$$

where the infimum over P is for those instances where λ_2 is fixed and $0<\lambda_1<\lambda_2$ is allowed to change. Indeed, when $\lambda_1<\lambda_2$, $\alpha_P=2/\lambda_2$ and (10) holds (choosing $\lambda_1\to 0$). When (10) holds we see that the lower bound blows up for any fixed choice of α in a worst-case sense. Thus, already in the two-dimensional, diagonal case, there is no “universal” stepsize.

B Proof of Theorem 4

Theorem 4. Any $\alpha \in (0,1)$ is a universal stepsize for both SPD and SPDSN. Furthermore, for any fixed $\alpha \in (0,1)$, $\theta_0 \in \mathbb{R}^d$,

$$\varepsilon_t(\text{SPD}) \asymp \|e_0\|^2 + \alpha^2\sigma_b^2 t\,, \quad \varepsilon_t'(\text{SPD}) \asymp \frac{\|e_0\|^2}{\alpha t} + \sigma_b^2\alpha\,,$$
$$\text{and} \qquad \varepsilon_t'(\text{SPDSN}) \asymp \frac{\|e_0\|^2}{\alpha t} + \frac{d}{t}\,.$$

Proof. For any P in SPD or SPDSN, the corresponding A_P matrix is real symmetric and positive definite. Thus, for each problem instance there exists an orthogonal matrix $U \in \mathbb{R}^{d\times d}$ (i.e., $U^\top U = I$) such that $U^\top A_P U = \Lambda_P$. Define $\zeta_t \doteq U^\top b_t$, and $\gamma_t = U^\top \theta_t$. Now, we have

$$\theta_t = (I - \alpha A_P)\theta_{t-1} + \alpha b_t\,, \tag{11}$$
$$\gamma_t = (I - \alpha \Lambda_P)\gamma_{t-1} + \alpha \zeta_t\,. \tag{12}$$

Since U is an orthogonal matrix $\|\gamma\|_{\Lambda_P} = \gamma^\top \Lambda_P \gamma = \theta^\top U \Lambda_P U^\top \theta = \|\theta\|_{A_P}$ and similarly $\|\gamma\| = \|\theta\|$ and when $P \in \mathcal{P}_{SPDSN}$ we have $\mathbf{E}\left[\zeta_t\zeta_t^\top\right] \preceq \Lambda_P$. Thus, for any adversarial choice of A_P is equivalent to choosing a problem when the A_P is diagonal. Thus, in order to prove the bounds it is enough to consider such diagonal problems. In what follows, by slightly abusing our notation, we recycle e_t and let $e_t \doteq \gamma_t - \gamma_*$.

It is clear that (12), has d separate 1-dimensional equations. Further, the MSE is a summation of d separate terms (i.e. $\mathbf{E}\left[\|\hat{e}_t\|\right] = \sum_{i=1}^d \hat{e}_t(i)^2$, $\mathbf{E}\left[\|\hat{e}_t\|_{\Lambda_P}\right] = \sum_{i=1}^d \Lambda(i)\hat{e}_t(i)^2$, where $\Lambda(i), i = 1,\dots,d$ are the eigenvalues) and any adversarial choice will maximize each of the d terms. Thus, by symmetry it follows that under our boundedness assumption such an adversarial choice should have $\Lambda(i) = \lambda, i = 1,\dots,d$. In what follows, we deal with the 1-dimensional case, where $\lambda \in (0,1)$ is the eigenvalue, and $b_t \in \mathbb{R}$ is the noise sequence and the condition $\mathbf{E}\left[b_t b_t^\top\right] \le A_P$ translates to $\mathbf{E}\left[b_t^2\right] \le \lambda$. In the 1-dimensional case, the error $\hat{e}_t = 1/(t+1)\sum_{s=0}^t e_t$ satisfies

$$\hat{e}_t = \frac{1}{t+1}\left[\sum_{s=0}^t (1-\alpha\lambda)^s e_0 + \alpha \sum_{i=1}^t \sum_{s=0}^{t-i} (1-\alpha\lambda)^s b_i\right]\,. \tag{13}$$

We now look at the three claims and prove matching lower and upper bounds for the error. In what follows, we use the following facts:

1. For all $t \ge 0$, $(1-\frac{1}{t+1})^0 \ge (1-\frac{1}{t+1})^1 \ge \cdots \ge (1-\frac{1}{t+1})^t \ge 1/e$, where e is the base of the natural logarithm. This will be used in the lower bound proofs.

2. When $0 \le 1-\alpha\lambda \le 1$, $\sum_{s=0}^t (1-\alpha\lambda)^s \le \min\{(\alpha\lambda)^{-1}, (t+1)\}$. This will be used in the upper bound proofs.

Lower Bound for $\varepsilon_t(\text{SPD})$: For sufficiently large t, and any step-size$\alpha \in (0,1)$ pick $\lambda > 0$ such that $\alpha\lambda(t+1) = 1$. Taking expectation of both sides of (13) and letting $\mathcal{B} = e_0^2$, we have

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right] = \frac{1}{(t+1)^2}\left[\left\{\sum_{s=0}^t (1-\alpha\lambda)^s\right\}^2 \mathcal{B} + \alpha^2\sigma_{b_P}^2 \sum_{i=1}^t \left\{\sum_{s=0}^{t-i}(1-\alpha\lambda)^s\right\}^2\right]\,, \tag{14}$$

where we use the property of independence. Now for our choice of λ we have that for some universal constants $c, c' > 0$,

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right] \ge \frac{c}{(t+1)^2}\left[(t+1)^2\mathcal{B} + \alpha^2\sigma_{b_P}^2 \sum_{i=1}^t (i)^2\right] \ge c'\left[B + \alpha^2\sigma_{b_P}^2 t\right]\,, \tag{15}$$

which completes the lower bound part for this case.

Upper Bound for $\varepsilon_t(\text{SPD})$: Starting from (24), we have

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right] \leq \frac{1}{(t+1)^2}\left[\mathcal{B}(\min\{(\alpha\lambda)^{-1},(t+1)\})^2+\alpha^2\sigma_{b_P}^2(t+1)(\min\{(\alpha\lambda)^{-1},(t+1)\})^2\right] \tag{16}$$

$$=\left[\mathcal{B}(\min\{\frac{1}{\alpha\lambda(t+1)},1\})^2+\alpha^2\sigma_{b_P}^2(t+1)(\min\{\frac{1}{\alpha\lambda(t+1)},1\})^2\right] \tag{17}$$

$$\leq\left[\mathcal{B}+\alpha^2\sigma_{b_P}^2(t+1)\right], \tag{18}$$

which completes the upper bound part for this case.

Lower Bound for $\varepsilon'_t(\text{SPD})$: We will slightly abuse notation by writing $\|\hat{e}_t\|_\lambda^2=\lambda(\hat{e}_t)^2$ and writing $\|\hat{e}_t\|^2=(\hat{e}_t)^2$. As before, pick λ such that $\alpha\lambda(t+1)=1$. Taking expectation of both sides of (13), and letting $\mathcal{B}=e_0^2$, we have

$$\mathbf{E}\left[\|\hat{e}_t\|_\lambda^2\right]=\frac{1}{(t+1)^2}\left[\lambda(\sum_{s=0}^{t}(1-\alpha\lambda)^s)^2\mathcal{B}+\alpha^2\sigma_{b_P}^2\lambda\sum_{i=1}^{t}(\sum_{s=0}^{t-i}(1-\alpha\lambda)^s)^2\right], \tag{19}$$

where we use the property of independence. Now for our choice of λ, for some universal constants $c,c'>0$,

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right]\geq\frac{c}{(t+1)^2}\left[\lambda(t+1)^2\mathcal{B}+\lambda\alpha^2\sigma_{b_P}^2\sum_{i=1}^{t}(i)^2\right]=c'\left[\frac{\mathcal{B}}{\alpha(t+1)}+\alpha\sigma_{b_P}^2\right], \tag{20}$$

which completes the lower bound part for this case.

Upper Bound for $\varepsilon'_t(\text{SPD})$: Starting from (24), we have

$$\mathbf{E}\left[\|\hat{e}_t\|_\lambda^2\right]\leq\frac{1}{(t+1)^2}\left[\lambda\mathcal{B}(\min\{(\alpha\lambda)^{-1},(t+1)\})^2+\lambda\alpha^2\sigma_{b_P}^2(t+1)(\min\{(\alpha\lambda)^{-1},(t+1)\})^2\right] \tag{21}$$

$$=\left[\mathcal{B}(\min\{\frac{1}{\alpha\lambda(t+1)},1)^2+\lambda\alpha^2\sigma_{b_P}^2(t+1)(\min\{\frac{1}{\alpha\lambda(t+1)},1)^2\right]. \tag{22}$$

Here if $\min\{(\alpha\lambda)^{-1},(t+1)\}=\frac{1}{\alpha\lambda}$, then we have $\lambda(\min\{(\alpha\lambda)^{-1},(t+1)\})^2=\frac{1}{\alpha^2\lambda}\leq\frac{(t+1)}{\alpha}$, and hence $\frac{1}{(t+1)^2}\lambda(\min\{(\alpha\lambda)^{-1},(t+1)\})^2=\frac{1}{\alpha(t+1)}$. In the other case, when $\min\{(\alpha\lambda)^{-1},(t+1)\}=(t+1)$, we have $\frac{1}{(t+1)^2}\lambda(\min\{(\alpha\lambda)^{-1},(t+1)\})^2=\frac{1}{\alpha(t+1)}$. Thus

$$\mathbf{E}\left[\|\hat{e}_t\|_\lambda^2\right]\leq\left[\mathcal{B}\frac{1}{\alpha(t+1)}+\alpha\sigma_{b_P}^2\right]. \tag{23}$$

This completes the upper bound part for this case.

Lower Bound for $\varepsilon'_t(\text{SPDSN})$: Pick λ such that $\alpha\lambda(t+1)=1$. Taking expectations of both sides of (13), and letting $\mathcal{B}=e_0^2$, we have

$$\mathbf{E}\left[\|\hat{e}_t\|_\lambda^2\right]=\frac{1}{(t+1)^2}\left[\lambda(\sum_{s=0}^{t}(1-\alpha\lambda)^s)^2\mathcal{B}+\alpha^2\sigma_{b_P}^2\lambda\sum_{i=1}^{t}(\sum_{s=0}^{t-i}(1-\alpha\lambda)^s)^2\right], \tag{24}$$

where we use the property of independence. Now for our choice of λ we have, for some universal constants $c,c'>0$ that

$$\mathbf{E}\left[\|\hat{e}_t\|^2\right]\geq\frac{c}{(t+1)^2}\left[\lambda(t+1)^2\mathcal{B}+\lambda^2\alpha^2\sum_{i=1}^{t}(i)^2\right]=c'\left[\frac{\mathcal{B}}{\alpha(t+1)}+\frac{1}{t}\right], \tag{25}$$

where we used that $\sigma_{b_P}^2=\lambda$. This completes the lower bound part for this case.

Upper Bound for $\varepsilon'_t(\text{SPDSN})$: Starting from (24), we have

$$\mathbf{E}\left[\|\hat{e}_t\|_\lambda^2\right]\leq\frac{1}{(t+1)^2}\left[\lambda\mathcal{B}(\min\{(\alpha\lambda)^{-1},(t+1)\})^2+\lambda\alpha^2\sigma_{b_P}^2(t+1)(\min\{(\alpha\lambda)^{-1},(t+1)\})^2\right] \tag{26}$$

$$=\left[\mathcal{B}(\min\{\frac{1}{\alpha\lambda(t+1)},1\})^2+\lambda\alpha^2\lambda(t+1)(\min\{\frac{1}{\alpha\lambda(t+1)},1\})^2\right]. \tag{27}$$

Here if $\min\{(\alpha\lambda)^{-1},(t+1)\} = \frac{1}{\alpha\lambda}$, then we have $\lambda(\min\{(\alpha\lambda)^{-1},(t+1)\})^2 = \frac{1}{\alpha^2\lambda} \leq \frac{(t+1)}{\alpha}$, and hence $\frac{1}{(t+1)^2}\lambda(\min\{(\alpha\lambda)^{-1},(t+1)\})^2 = \frac{1}{\alpha(t+1)}$. In the other case, when $\min\{(\alpha\lambda)^{-1},(t+1)\} = (t+1)$, we have $\frac{1}{(t+1)^2}\lambda(\min\{(\alpha\lambda)^{-1},(t+1)\})^2 = \frac{1}{\alpha(t+1)}$. Thus,

$$\mathbf{E}\left[\|\hat{e}_t\|_\lambda^2\right] \leq \left[\mathcal{B}\frac{1}{\alpha(t+1)} + \alpha\frac{1}{\alpha(t+1)}\right], \tag{28}$$

where we uses that $\sigma_{b_P}^2 = \lambda$. This completes the upper bound part for this case. □

C Proofs for Section 5

Lemma 11. Let A be a $d \times d$ matrix with $B \geq \max_{ij}|A_{ij}|$. It follows that $\max_{x\in\mathbb{R}^d:\|x\|\leq 1} x^\top A^\top Ax \leq B^2 d$.

We now resort to an the following notation for real symmetric positive definite $d \times d$ matrices C and D: $C \succeq D$ if $C - D$ is positive definite. We cite the following lemma without proof (see [10]). The lemma is often attributed to Schur:

Lemma 12. Let A, B, C be $d \times d$ real matrices. Given a symmetric $2d \times 2d$ matrix $M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$, it follows that $M \succeq 0$ if $A \succeq 0$ and $C - B^\top A^{-1}B \succeq 0$.

The matrix $C - B^\top A^{-1}B$ appearing in this result is called the Schur complement of M.

Theorem 5. Let $B \geq 1$. The following hold: *i)* CATD(0) has a universal stepsize of $\alpha_{td} = \frac{1}{B^2}$ for the class SOFS(B). *ii)* CAGTD has a universal stepsize of $\alpha_{gtd} = \frac{1}{2B^2 d}$ for the class GTDOFF(B).

Proof. It is sufficient to show in all three cases that $\rho_s(\alpha, P_A) > 0$ holds for the said values of α.

We introduce the relations $=_q$, $\leq_q$ and $\geq_q$ to compare (not necessarily) symmetric $d \times d$ matrices over the reals when the matrices are identified with the *quadratic form* that they define. That is, for matrices $A, B \in \mathbb{R}^{d\times d}$, we write $A \leq_q B$ to denote that $x^\top(A - B)x \leq 0$ holds for all $x \in \mathbb{R}^d$. Then, we write $A \geq_q B$ when $B \leq_q A$ and we write $A =_q B$ when both $A \leq_q B$ and $B \geq_q A$ hold. Specifically, note that for any matrix A, $A =_q A^\top$. Further, for symmetric matrices $A \succeq B$ holds if and only if $A \geq_q B$ holds. Also, $A = B$ implies $A =_q B$, though the converse does not hold. In particular, for a symmetric matrix A, $\lambda_{\min}(A) > 0$ is equivalent to $A \geq_q 0$, where 0 stands for the all-zero matrix.

Part (*i*)**:** Using $\alpha = 1/B^2$ and $\|\phi_t\|^2 \leq B^2$, we have $\alpha\phi^\top\phi = \frac{1}{B^2}\phi_t^\top\phi_t \leq 1$, and hence

$$(A_t + A_t^\top) - \alpha A_t^\top A_t \geq_q 2\phi_t(\phi_t - \gamma\phi_t')^\top - (\phi_t - \gamma\phi_t')(\phi_t - \gamma\phi_t')^\top = (\phi_t + \gamma\phi_t')(\phi - \gamma\phi_t')^\top = \phi_t\phi_t^\top - \gamma^2\phi_t'\phi_t'^\top .$$

Taking expectations, and using $\mathbf{E}[\phi_t\phi_t^\top] = \mathbf{E}[\phi_t'(\phi_t')^\top]$, we get

$$\mathbf{E}\left[(A_t + A_t^\top) - \alpha A_t^\top A_t\right] \geq_q \mathbf{E}\left[\phi_t\phi_t^\top - \gamma^2\phi_t'\phi_t'^\top\right] = (1-\gamma^2)\mathbf{E}\left[\phi_t\phi_t^\top\right] \succeq 0 .$$

Part (*ii*)**:** When $\alpha = \beta$, the GTD update in Table 1 can be expressed as $x_{t+1} = x_t + \alpha(g_t - H_t x_t)$, where $x_t = \begin{bmatrix} y_t \\ \theta_t \end{bmatrix}$, $H_t = \begin{bmatrix} I & A_t \\ -(1-\alpha)A_t^\top & \alpha A_t^\top A_t \end{bmatrix}$, and $g_t = \begin{bmatrix} b_t \\ \alpha A_t^\top b_t \end{bmatrix}$, where $A_t = \phi_t(\phi_t - \gamma\phi_t')^\top$. To show that $\alpha = \frac{1}{2B^2 d}$ is a universal step-size choice, we need to show that for this value of α, $\lambda_{\min}\left(\mathbf{E}\left[H_t + H_t^\top - \alpha H_t^\top H_t\right]\right) > 0$.

Now $H_t + H_t^\top = \begin{bmatrix} 2I & \alpha A_t \\ \alpha A_t^\top & 2\alpha A_t^\top A_t \end{bmatrix}$ and $H_t^\top H_t = \begin{bmatrix} I + (1-\alpha)^2 A_t A_t^\top & A_t - \alpha(1-\alpha)A_t A_t^\top A_t \\ A_t^\top - \alpha(1-\alpha)A_t^\top A_t A_t^\top & A_t^\top A_t + \alpha^2 A_t^\top A_t A_t^\top A_t \end{bmatrix}$. From Lemma 11 and the choice of α, $\alpha A_t A_t^\top \preceq I$. Hence, using that $A \succeq B$ for symmetric A, B is equivalent to $A \geq_q B$, we get that $(H_t + H_t^\top) - \alpha(H_t^\top H_t) \succeq S_t$, where $S_t = \begin{bmatrix} I & \alpha^2(1-\alpha)A_t A_t^\top A_t \\ \alpha^2(1-\alpha)A_t^\top A_t A_t^\top & \alpha A_t^\top A_t - \alpha^3 A_t^\top A_t A_t^\top A_t \end{bmatrix}$. Now from Lemma 12, to show that $S_t \succeq 0$ we need to prove that the Schur complement of S_t is nonnegative.

For this we have

$$
\begin{aligned}
&\alpha A_t^\top A_t - \alpha^3 A_t^\top A_t A_t^\top A_t - \alpha^4 (1-\alpha)^2 A_t^\top A_t A_t^\top A_t A_t^\top A_t \\
&\quad \succeq \alpha A_t^\top A_t - \alpha^3 A_t^\top A_t A_t^\top A_t - \alpha^4 A_t^\top A_t A_t^\top A_t A_t^\top A_t \\
&\quad = \alpha A_t^\top \big(I - \alpha^2 A_t A_t^\top - \alpha^3 A_t A_t^\top A_t A_t^\top \big) A_t \\
&\quad \succeq \alpha A_t^\top \big(I - \alpha^2 A_t (I + \alpha A_t^\top A_t) A_t^\top \big) A_t \\
&\quad \succeq \frac{\alpha}{2} A_t^\top A_t \,,
\end{aligned}
$$

where the last inequality follows from the choice of α, bound B on the features and Lemma 11. $\square$