Exponential Parameter Convergence of Discrete-Time High-Order Tuners Under Persistent Excitation

Yingnan Cui† and Anuradha M. Annaswamy

Abstract

We consider two algorithms that have been shown to have accelerated performance, Polyak’s heavy ball method and Nesterov’s acceleration method. In the context of adaptive control, we show that both algorithms will present accelerated learning phenomenon under persistent excitation. Simulation results show that the two algorithms have very similar behavior and are both faster than normalized gradient descent.

1 Introduction

One critical part of any adaptive control system is the online parameter estimation algorithm. Beyond the standard gradient descent algorithm, many online parameter estimation schemes have been proposed and shown to have accelerated performance in both continuous and discrete time [1-4]. In order to obtain parameter convergence, a necessary and sufficient condition is persistent excitation [1, 4]. Two types of online optimization algorithms are of great interest to the adaptive control community: the first type focuses on adjustments of the learning rate matrix during the update and the second type adopts the idea of “momentum” during the update [5-8]. This paper focuses on the second type of algorithm in discrete-time and shows that under persistent excitation, accelerated learning will happen in such systems.

Specifically, we will look into a class of online parameter estimation methods as known as higher-order tuners. In his seminal work [6], Morse introduces a new parameter tuning method that generates the first $n$ time-derivatives of each parameter in the set of tuned parameters without employing a normalized tuning error. The high-order tuner in [6] is continuous, and is later proved in [5] to result in exponential parameter convergence. The same exponential convergence results have also been proved in [9] using a different approach from [5]. But a discrete-time parameter convergence analysis has been lacking in the literature.

The discrete-time counterpart of the same idea of continuous-time high-order tuner first originated from optimization methods for machine learning. Polyak first showed that the momentum method can significantly accelerate convergence to a local minimum [10]. Nesterov’s work in [11], which used estimate sequences for the verification of classical momentum methods, recently captured much of the attention in machine learning community. Inspired by Nesterov’s work, many efforts have been dedicated to the understanding of Nesterov’s acceleration through a continuous-time perspective [12, 13]. Based on continuous ODEs, various discretization techniques lead to new discrete-time algorithms that are shown to be stable and have accelerated performance [7].

---

*This work is supported by the Boeing Strategic University Initiative.
†Y. Cui and A.M. Annaswamy are with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139.
One notable difference of the focus on accelerated learning between the machine learning community and the adaptive control community is the constancy of input features. In machine learning, the inputs are constant and thus non-asymptotic convergence rates can be found. Typical adaptive systems have time-varying inputs for the parameter estimation algorithm and asymptotic convergence can be shown through Lyapunov arguments [7]. Though previous works have shown accelerated performance for the high-order tuners, to our best knowledge no one has shown accelerated parameter learning for these algorithm, which is the focus of this paper [7].

Section 2 presents the definitions. Section 3 presents problem statement. Section 4 shows the asymptotic convergence of Polyak’s heavy-ball method. Section 5 shows the asymptotic convergence of Nesterov’s accelerated learning algorithm. We show simulation results in Section 6 and conclude in Section 7.

2 Preliminaries

Definition 2.1. The regressor \( \phi_k \) satisfies the following persistent excitation (PE) condition, for all unit vectors \( \mathbf{w} \in \mathbb{R}^n \):

\[
\frac{1}{\Delta T} \sum_{i=k-\Delta T}^{k-1} \| \mathbf{w}^\top \phi_i \| \geq \epsilon.
\]

Definition 2.2. For any fixed \( p \in [1, \infty) \), a sequence of scalars \( \xi = \{\xi_0, \xi_1, \ldots\} \) is defined to belong to \( \ell_p \) if

\[
\|\xi\|_{\ell_p} \equiv \left( \lim_{k \to \infty} \sum_{i=0}^{k} \|\xi_i\|^p \right)^{1/p} < \infty.
\]

When \( p = \infty \), \( \xi \in \ell_\infty \) if

\[
\|\xi\|_{\ell_\infty} \equiv \sup_{i \geq 0} \|\xi_i\| < \infty
\]

3 Problem Statement

We consider the class of discrete-time models of the form

\[
y_k = -\sum_{i=1}^{n} a_i^* y_{k-i} + \sum_{j=1}^{m} b_j^* u_{k-j-d} + \sum_{\ell=1}^{p} c_{\ell}^* f_{\ell}(y_{k-1}, \ldots, y_{k-n}, u_{k-1-d}, \ldots, u_{k-m-d}),
\]

where \( a_i^* \), \( b_j^* \) and \( c_{\ell}^* \) are unknown parameters that are constant and need to be identified, and \( d \) is a known time-delay. The function \( f_{\ell} \) is an analytic function of its arguments and is assumed to be such that the system in (4) is bounded-input-bounded-output (BIBO) stable. Denote \( z_{k-1} = [y_{k-1}, \ldots, y_{k-n}]^\top \) and \( v_{k-d-1} = [u_{k-1-d}, \ldots, u_{k-m-d}]^\top \). We rewrite (4) in the form of a linear regression

\[
y_k = \phi_k^\top \theta^*.
\]

where \( \phi_k = [z_{k-1}^\top, v_{k-d-1}^\top, f_1(z_{k-1}, v_{k-d-1}), \ldots, f_p(z_{k-1}, v_{k-d-1})]^\top \) is a regressor determined by exogenous signals and \( \theta^* = [a_1^*, \ldots, a_n^*, b_1^*, \ldots, b_m^*, c_1^*, \ldots, c_p^*]^\top \) is the underlying unknown parameter vector. We propose to identify the parameter \( \theta^* \) as \( \theta_k \) using an estimator

\[
\hat{y}_k = \phi_k^\top \hat{\theta}_k.
\]

This leads to a prediction error

\[
e_{y,k} = \phi_k^\top \delta_k.
\]
where $e_{y,k} = \hat{y}_k - y_k$ is the output prediction error and $\hat{\theta}_k = \theta_k - \theta^*$ is the parameter error. The goal of parameter identification is to design an iterative procedure such that given regressor $\phi_k$ is persistently exciting, parameter error $\|\hat{\theta}_k\|$ goes to zero exponentially fast. Throughout the paper, we consider the squared loss function using (7),

$$L_k(\theta_k) = \frac{1}{2} e^2_{y,k} = \frac{1}{2} \hat{y}_k^T \phi_k \phi_k^T \hat{y}_k,$$

where the subscript $k$ in $L_k$ denotes $k$th iteration. As a starting point, normalized gradient descent algorithm has been shown to be stable although having a slow convergence rate [4]

$$\theta_{k+1} = \theta_k - \alpha \frac{\nabla L_k(\theta_k)}{N_k}, \quad 0 < \alpha < 2,$$

where $N_k$ is a normalizing signal and is defined as $N_k = 1 + \|\phi_k\|^2$.

4 Accelerated Learning with Heavy-Ball Method

The Heavy Ball method of Polyak [10] has the following form

$$\vartheta_{k+1} = \vartheta_k - \gamma \frac{\nabla L_k(\vartheta_{k+1})}{N_k},$$

$$\theta_{k+1} = \theta_k - \beta(\theta_k - \vartheta_k),$$

where $\beta$ and $\gamma$ are the hyperparameters. We consider the following Lyapunov candidate in the proof of parameter convergence

$$V_k = \frac{1}{\gamma} \|\vartheta_k - \theta^*\|^2 + \frac{1}{\gamma} \|\theta_k - \vartheta_k\|^2.$$

Let

$$c_1 = \frac{11}{8}, \quad c_2 = \frac{21}{32},$$

$$\epsilon_1 = \frac{\epsilon}{\max_k \{N_k\}}$$

$$0 < \lambda < \min \left\{ 1, \frac{1}{\gamma} \right\},$$

$$0 < \eta < \min \left\{ 1, \frac{\Delta T}{c_1 \lambda \gamma}, \epsilon_1 \right\},$$

and

$$0 < \zeta < \min \left\{ 1, \sqrt{\frac{\Delta T}{c_2 \gamma (1 - \lambda)^2}}, \epsilon_1 \right\}.$$

We define

$$\mu = \min \{\mu_1, \mu_2, \mu_3, \mu_4\},$$

where

$$\mu_1 = \frac{c_1 \lambda \gamma \eta^2}{\Delta T},$$

$$\mu_2 = \frac{c_2 \Delta T (\epsilon_1 - \gamma \eta)^2 \lambda \gamma}{(1 + \gamma \Delta T)^2},$$

$$\mu_3 = \frac{c_2 \zeta^2 (1 - \lambda) \gamma}{\Delta T}.$$
\[
\mu_4 = \frac{c_1 \Delta T (c_1 - \gamma \zeta)^2 \gamma (1 - \lambda)}{[2 + \gamma \Delta T + \Delta T \sqrt{1 - 8 \gamma}]^2}.
\]

The following Theorem states that exponential convergence of the parameter error towards zero will happen provided that \(\phi_k\) is persistently exciting. Note that since PE is a property that pertains to a period, we have to examine the Lyapunov candidate defined in (11) across a period. We will discuss in two cases based on how close the two states, \(\theta_k - \Delta T\) and \(\theta_k - \Delta T\), are at time \(k - \Delta T\), for all \(k\).

**Theorem 4.1.** If the regressor \(\phi_k\) satisfies the definition in (1), with \(0 < \beta < 2\) and \(0 < \gamma < \frac{\beta (2 - \beta)}{8}\), the update law in (10) will result in \(\theta_k - \theta^* \in \ell_\infty\), \(\theta_k - \theta_k \in \ell_\infty\), and \(V_k \leq \exp \left(-\mu \left[ \frac{k}{\Delta T} \right] \right) V_0\), where \(\mu\) is defined in (13).

**Proof.** Expanding \(\Delta V_k := V_{k+1} - V_k\), we have

\[
\Delta V_k \leq \frac{1}{N_k} \left\{ - \left[ \|\tilde{\theta}_{k+1} \phi_k\| - 2 \|\phi_k\| \|\theta_k - \theta_k\| \right]^2 - \frac{7}{8} \left(2 \|(1 - \beta) (\theta_k - \theta_k) \phi_k\| - \frac{1}{2} \|(\theta_k - \theta^*) \phi_k\| \right)^2 - \frac{21}{32} \left(\|(\theta_k - \theta^*) \phi_k\| - 4 \|\theta_k - \vartheta_k\|^2 \right) - \frac{11}{8} \|\theta_k - \vartheta_k\|^2 \right\}
\]

Summing \(V_k\) from \(V_k - \Delta T\) to \(V_k\), we have

\[
V_k - V_{k-DT} = V_k - V_{k-1} + \cdots + V_{k-\Delta T+1} - V_{k-\Delta T} \\
\leq -c_1 \sum_{i=k-\Delta T}^{k-1} \|\theta_i - \theta_i\|^2 - c_2 \sum_{i=k-\Delta T}^{k-1} \frac{1}{N_i} \|(\theta_i - \theta^*) \phi_i\| \\\n\leq -c_1 \frac{1}{\Delta T} \left( \sum_{i=k-\Delta T}^{k-1} \|\theta_i - \theta_i\| \right)^2 \tag{16} \]

\[
- c_2 \frac{1}{\Delta T} \left( \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \|(\theta_i - \theta^*) \phi_i\| \right)^2 \]

Based on how large \(\|\theta_k - \Delta T - \theta^*\|^2\) is compared with \(V_k - \Delta T\), we consider two cases: \(\|\theta_k - \Delta T - \theta^*\|^2 \geq \lambda \gamma V_k - \Delta T\) and \(\|\theta_k - \Delta T - \theta^*\|^2 \leq \lambda \gamma V_k - \Delta T\), where \(\lambda\) satisfies (12).

**Case 1:** \(\|\theta_k - \Delta T - \theta^*\|^2 \geq \lambda \gamma V_k - \Delta T\)

(a) If \(W_1 \geq \eta \|\theta_k - \Delta T - \theta^*\|\), where \(\eta\) satisfies (13), then

\[
V_k - V_{k-DT} \leq -c_1 \frac{1}{\Delta T} \eta^2 \|\theta_k - \Delta T - \theta^*\|^2 \leq -c_1 \frac{1}{\Delta T} \lambda \gamma^2 V_k - \Delta T
\]

and

\[
V_k \leq \left(1 - c_1 \frac{1}{\Delta T} \lambda \gamma^2 \right) V_k - \Delta T. \tag{17}
\]
Since from (13) $0 < c_1 \frac{1}{\Delta T} \lambda \gamma \eta^2 < 1$, asymptotic convergence follows.

(b) If $W_1 \leq \eta \|\theta_{k-\Delta T} - \theta^*\|$, then

$$W_2 = \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta_{k-\Delta T} + \theta_{k-\Delta T} - \theta^*)^\top \phi_i \right\|$$

$$\geq \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_{k-\Delta T} - \theta^*)^\top \phi_i \right\|$$

$$- \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta_{k-\Delta T})^\top \phi_i \right\|$$

$$\geq \Delta T \epsilon_1 \|\theta_{k-\Delta T} - \theta^*\| - \Delta T \sup_{i \in[k-\Delta T,k-1]} \|\theta_i - \theta_{k-\Delta T}\|$$

$$\geq \Delta T \epsilon_1 \|\theta_{k-\Delta T} - \theta^*\| - \Delta T \sum_{i=k-\Delta T}^{k-2} \|\theta_{i+1} - \theta_i\|$$

$$\geq \Delta T \epsilon_1 \|\theta_{k-\Delta T} - \theta^*\|$$

$$- \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \left\| \frac{\phi_i \phi_i^\top \theta_{i+1}}{N_i} \right\|$$

$$\geq \Delta T \epsilon_1 \|\theta_{k-\Delta T} - \theta^*\|$$

$$- \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_{i+1} - \theta_i + \theta_i - \theta^*)^\top \phi_i \right\|$$

$$\geq \Delta T \epsilon_1 \|\theta_{k-\Delta T} - \theta^*\| - \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta^*)^\top \phi_i \right\|$$

$$- \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta_i)^\top \phi_i \right\|$$

$$\geq \Delta T \epsilon_1 \|\theta_{k-\Delta T} - \theta^*\| - \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta^*)^\top \phi_i \right\|$$

$$- \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \|\theta_i - \theta_i\|$$

Therefore

$$W_2 \geq \frac{\Delta T}{1 + \gamma \Delta T} (\epsilon_1 - \gamma \eta) \|\theta_{k-\Delta T} - \theta^*\|$$

From (13), $\epsilon_1 > \gamma \eta$. Substitute (18) into (16), we obtain

$$V_k - V_{k-\Delta T} \leq -c_2 \frac{1}{\Delta T} \frac{W_2^2}{2}$$

$$\leq -c_2 \frac{\Delta T}{(1 + \gamma \Delta T)^2} (\epsilon_1 - \gamma \eta)^2 \|\theta_{k-\Delta T} - \theta^*\|^2$$
\[ V_k \leq 1 - c_2 \frac{\Delta T}{(1 + \gamma \Delta T)^2} (\epsilon_1 - \gamma \eta)^2 \lambda \gamma V_{k-\Delta T} \]

And thus
\[ V_k \leq \left[ 1 - c_2 \frac{\Delta T}{(1 + \gamma \Delta T)^2} (\epsilon_1 - \gamma \eta)^2 \lambda \gamma \right] V_{k-\Delta T}. \quad (19) \]

**Case 2:** \( \| \theta_{k-\Delta T} - \theta^* \|^2 \leq \lambda \gamma V_{k-\Delta T} \)

From
\[ V_{k-\Delta T} = \frac{1}{\gamma} \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \|^2 + \frac{1}{\gamma} \| \theta_{k-\Delta T} - \theta^* \|^2, \]
we immediately get
\[ \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \|^2 \geq (1 - \lambda) \gamma V_{k-\Delta T}. \quad (20) \]

(a) If \( W_2 \geq \zeta \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \| \), where \( \zeta \) satisfies (14), then we have
\[ V_k - V_{k-\Delta T} \leq -c_2 \frac{1}{\Delta T} \zeta^2 \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \|^2 \]
\[ \leq -c_2 \frac{1}{\Delta T} \zeta^2 (1 - \lambda) \gamma V_{k-\Delta T} \]

Thus
\[ V_k \leq \left[ 1 - c_2 \frac{1}{\Delta T} \gamma \zeta^2 (1 - \lambda) \right] V_{k-\Delta T}. \quad (21) \]

Since from (14), \( 0 < c_2 \frac{1}{\Delta T} \gamma \zeta^2 (1 - \lambda) < 1 \), asymptotic convergence follows.

(b) If \( W_2 \leq \zeta \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \| \), then
\[ W_1 \]
\[ \geq \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\theta_i - \vartheta_i) \phi_i \| \]
\[ = \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\theta_{k-\Delta T} - \vartheta_{k-\Delta T} + \theta_i - \theta_{k-\Delta T} \]
\[ - (\vartheta_i - \vartheta_{k-\Delta T}) \phi_i \| \]
\[ \geq \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\theta_{k-\Delta T} - \vartheta_{k-\Delta T}) \phi_i \| \]
\[ - \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\theta_i - \vartheta_{k-\Delta T}) \phi_i \| \]
\[ - \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\theta_i - \theta_{k-\Delta T}) \phi_i \| \]
\[ \geq \Delta T \epsilon_1 \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \|
\[ - \Delta T \sup_{i \in [k-\Delta T, k-1]} \| \theta_i - \vartheta_{k-\Delta T} \|
\[ - \Delta T \sup_{i \in [k-\Delta T, k-1]} \| \theta_i - \theta_{k-\Delta T} \|
\[ \geq \Delta T \epsilon_1 \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \|
\[-\Delta T \sum_{i=k-\Delta T}^{k-2} ||\vartheta_{i+1} - \vartheta_i|| - \Delta T \sum_{i=k-\Delta T}^{k-2} ||\theta_{i+1} - \theta_i|| \geq \Delta T \epsilon_1 ||\theta_{k-\Delta T} - \vartheta_{k-\Delta T}||
\]
\[-\gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \|(\vartheta_i - \theta^*)^T \phi_i\| - \beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i|| \geq \Delta T \epsilon_1 ||\theta_{k-\Delta T} - \vartheta_{k-\Delta T}||
\]
\[-\gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \|(\vartheta_i - \theta^*)^T \phi_i\|\]
\[-\gamma \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i|| - \beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i|| \geq \Delta T \epsilon_1 ||\theta_{k-\Delta T} - \vartheta_{k-\Delta T}||
\]
\[-\gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \|(\vartheta_i - \theta^*)^T \phi_i\| - \beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i|| \geq \Delta T \epsilon_1 ||\theta_{k-\Delta T} - \vartheta_{k-\Delta T}||
\]
\[-\gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \|(\vartheta_i - \theta^*)^T \phi_i\|\]
\[-\beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i||\]
\[-\beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i||\]
\[-\beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i||\]
\[-\beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i||\]
\[-\beta \Delta T \sum_{i=k-\Delta T}^{k-1} ||\theta_i - \vartheta_i||\]

Finally we get
\[W_1 \geq \frac{1}{1 + \gamma \Delta T + \beta \Delta T} \Delta T (\epsilon_1 - \gamma \zeta) ||\theta_{k-\Delta T} - \vartheta_{k-\Delta T}||
\]

From (16),
\[V_k - V_{k-\Delta T} \leq -c_1 \frac{1}{\Delta T} W_1^2 \leq -c_1 \frac{\Delta T (\epsilon_1 - \gamma \zeta)^2}{(1 + \gamma \Delta T + \beta \Delta T)^2} ||\theta_{k-\Delta T} - \vartheta_{k-\Delta T}||^2 \leq -c_1 \frac{\Delta T (\epsilon_1 - \gamma \zeta)^2}{(1 + \gamma \Delta T + \beta \Delta T)^2} \gamma (1 - \lambda) V_{k-\Delta T}.
\]

Thus we have
\[V_k \leq \left[1 - c_1 \frac{\Delta T (\epsilon_1 - \gamma \zeta)^2}{(1 + \gamma \Delta T + \beta \Delta T)^2} \gamma (1 - \lambda) \right] V_{k-\Delta T} \quad (22)
\]

Summary
Considering (17), (19), (21) and (22), we have
\[V_k \leq (1 - \mu) V_{k-\Delta T},\]
where $\mu$ is defined in (15). Collecting the terms, we obtain

$$V_k \leq \exp \left( -\mu \frac{k}{\Delta T} \right) V_0.$$

### 5 Accelerated Learning with High-Order Tuner

From [7], the high-order tuner has the form of

$$\vartheta_{k+1} = \vartheta_k - \gamma \nabla L_k(\theta_{k+1}) N_k,$$

$$\theta_{k+1} = \bar{\theta}_k - \beta(\bar{\theta}_k - \vartheta_k),$$

$$\bar{\theta}_k = \theta_k - \gamma \beta \nabla L_k(\theta_k) N_k,$$

where $\beta$ and $\gamma$ are hyperparameters. We consider the same Lyapunov function as defined in (11).

Let

$$c_3 = \frac{7}{4}, \quad c_4 = \frac{9}{16},$$

$$\epsilon_2 = \frac{\epsilon}{\max_k \{N_k\}},$$

$$0 < \lambda < \min \left\{ 1, \frac{1}{\lambda} \right\},$$

$$0 < \eta < \min \left\{ 1, \sqrt{\frac{c_3 \Delta T}{\lambda \gamma}}, \frac{\epsilon_2}{\gamma} \right\},$$

$$0 < \zeta < \min \left\{ 1, \sqrt{\frac{c_4 \Delta T}{\gamma(1-\lambda)^2}}, \frac{\epsilon_2}{\gamma}, \frac{1}{\gamma}, \frac{1}{(1-\gamma \beta) \Delta T} \right\},$$

and define

$$\mu = \min \{ \mu_1, \mu_2, \mu_3, \mu_4 \},$$

where

$$\mu_1 = \frac{c_3 \lambda \gamma \eta^2}{\Delta T},$$

$$\mu_2 = \frac{c_4 \Delta T (\epsilon_2 - \gamma \eta^2) \lambda \gamma}{(1 + \gamma \Delta T)^2},$$

$$\mu_3 = \frac{c_4 \zeta^2 (1-\lambda)}{\Delta T},$$

$$\mu_4 = \frac{c_3 \xi^2 (1-\lambda) \gamma}{\Delta T},$$

and

$$\xi = \frac{\Delta T - \gamma \zeta \Delta T - \frac{\gamma \beta (1 + \beta \Delta T)}{1 - \gamma \beta}}{1 + \gamma \Delta T + \beta^2 + \gamma \beta \frac{1 + \beta \Delta T}{1 - \gamma \beta}}.$$

Theorem 5.1 states that under PE, parameter error goes to zero with the update law in (23).
Theorem 5.1. If the regressor $\phi_k$ satisfies the PE definition in [1], with $0 < \beta < 1$ and $0 < \gamma < \frac{\beta(2-\beta)}{8+\beta^2}$, the update law in (23) will result in $\bar{\vartheta}_k - \theta^* \in \ell_\infty$, $\bar{\theta}_k - \theta^* \in \ell_\infty$, and $V_k \leq \exp \left( -\mu \left[ \frac{1}{\Delta T} \right] \right) V_0$, where $\mu$ is defined in (27).

Proof. Expanding $\Delta V_k := V_{k+1} - V_k$, we have

$$\Delta V_k \leq -\frac{7}{4} \|\bar{\vartheta}_k - \vartheta_k\|^2 + \frac{1}{N_k} \left\{ -\left[ \|\bar{\vartheta}_{k+1}\| \phi_k \|\bar{\vartheta}_k - \vartheta_k\| \right]^2 - \left[ \frac{\sqrt{\beta}}{\sqrt{N_k}} \|\nabla L_k(\theta_k)\| - \beta \|\phi_k\| \|\bar{\vartheta}_k - \vartheta_k\| \right]^2 - \left( \frac{25}{4} + \beta^2 \right) \|\bar{\vartheta}_k - \vartheta_k\|^2 - \beta^2 \|\bar{\theta}_k - \vartheta_k\| \right\} \}

$$

where we applied Cauchy-Schwartz inequality for the second inequality.

Case 1: $\|\bar{\vartheta}_k - \Delta T - \theta^*\|^2 \geq \lambda \gamma V_k - \Delta T$, where $\lambda$ satisfies (24).

(a) If $W_1 \geq \eta \|\bar{\vartheta}_k - \Delta T - \theta^*\|$, where $\eta$ satisfies (25), then

$$V_k - V_{k-\Delta T} \leq c_3 \sum_{i=k-\Delta T}^{k-1} \|\bar{\vartheta}_i - \vartheta_i\|^2 - c_4 \sum_{i=k-\Delta T}^{k-1} \frac{1}{N_i} \|\vartheta_i - \theta^*\| \phi_i \|\|^2 \leq -c_3 \frac{1}{\Delta T} \left( \sum_{i=k-\Delta T}^{k-1} \|\bar{\vartheta}_i - \vartheta_i\| \right)^2 - c_4 \frac{1}{\Delta T} \left( \sum_{i=k-\Delta T}^{k-1} \frac{1}{N_i} \|\vartheta_i - \theta^*\| \phi_i \|\|^2 \right)^2,$$  \hspace{1cm} (28)

Since from (25), $0 < \frac{c_3}{\Delta T} \lambda \gamma \eta^2 < 1$, asymptotic convergence follows.
(b) If $W_1 \leq \eta \| \vartheta_{k-\Delta T} - \theta^* \|$, then

$$W_2$$

$$= \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\vartheta_i - \vartheta_{k-\Delta T} + \vartheta_{k-\Delta T} - \theta^*)^T \phi_i \|$$

$$\geq \Delta T \epsilon_2 \| \vartheta_{k-\Delta T} - \theta^* \| - \Delta T \sum_{i=k-\Delta T}^{k-2} \| \vartheta_{i+1} - \vartheta_i \|$$

$$\geq \Delta T \epsilon_2 \| \vartheta_{k-\Delta T} - \theta^* \| - \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\vartheta_i - \theta^*)^T \phi_i \|$$

$$- \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\vartheta_i - \theta_i + \vartheta_i - \theta^*)^T \phi_i \|$$

$$\geq \Delta T \epsilon_2 \| \vartheta_{k-\Delta T} - \theta^* \| - \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\vartheta_i - \theta^*)^T \phi_i \|$$

$$- \frac{1}{\sqrt{N_i}} \| (\bar{\theta}_i - \theta_i)^T \phi_i \|$$

$$\geq \Delta T \epsilon_2 \| \vartheta_{k-\Delta T} - \theta^* \| - \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \| (\vartheta_i - \theta^*)^T \phi_i \|$$

$$- \gamma \Delta T \sum_{i=k-\Delta T}^{k-1} \| \bar{\theta}_i - \theta_i \|$$

Therefore

$$W_2 \geq \frac{\Delta T}{1 + \gamma \Delta T} (\epsilon_2 - \gamma \eta) \| \vartheta_{k-\Delta T} - \theta^* \| \quad (30)$$

Note that from $\epsilon_2 > \gamma \eta$. Substitute (30) into (16), we obtain

$$V_k - V_{k-\Delta T} \leq -c_4 \frac{1}{\Delta T} W_2^2$$

$$\leq -c_4 \frac{\Delta T}{(1 + \gamma \Delta T)^2} (\epsilon_2 - \gamma \eta)^2 \| \vartheta_{k-\Delta T} - \theta^* \|^2$$

$$\leq -c_4 \frac{\Delta T}{(1 + \gamma \Delta T)^2} (\epsilon_2 - \gamma \eta)^2 \lambda \gamma V_{k-\Delta T}$$
Thus
\[ V_k \leq \left[ 1 - c_4 \frac{\Delta T}{(1 + \gamma \Delta T)^2} (\varepsilon_2 - \gamma \eta)^2 \lambda \gamma \right] V_{k-\Delta T} \tag{31} \]

**Case 2:** \( \|\theta_{k-\Delta T} - \theta^*\|^2 \leq \lambda \gamma V_{k-\Delta T} \)

From
\[ V_{k-\Delta T} = \frac{1}{\gamma} \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\|^2 + \frac{1}{\gamma} \|\theta_{k-\Delta T} - \theta^*\|^2 \]
we immediately get
\[ \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\|^2 \geq (1 - \lambda) \gamma V_{k-\Delta T} \tag{32} \]

(a) If \( W_2 \geq \zeta \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\| \), where \( \zeta \) satisfies (26), then we have
\[
V_k - V_{k-\Delta T} \leq -c_4 \frac{1}{\Delta T} \zeta^2 \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\|^2 \\
\leq -c_4 \frac{1}{\Delta T} \zeta^2 (1 - \lambda) \gamma V_{k-\Delta T}
\]
Thus
\[ V_k \leq \left[ 1 - c_4 \frac{1}{\Delta T} \gamma \zeta^2 (1 - \lambda) \right] V_{k-\Delta T} \tag{33} \]

Since from (26), \( 0 < c_4 \frac{1}{\Delta T} \gamma \zeta^2 (1 - \lambda) < 1 \), asymptotic convergence follows.

(b) If \( W_2 \leq \zeta \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\| \), then

\[
W_1 = \sum_{i=k-\Delta T}^{k-1} \|\bar{\theta}_i - \tilde{\theta}_i\| \\
= \sum_{i=k-\Delta T}^{k-1} \|\theta_{k-\Delta T} - \theta_{k-\Delta T} + \tilde{\theta}_i - \theta_{k-\Delta T} \\
- (\bar{\theta}_i - \theta_{k-\Delta T})\| \]
\[
\geq \sum_{i=k-\Delta T}^{k-1} \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\| \\
- \sum_{i=k-\Delta T}^{k-1} \|\bar{\theta}_i - \theta_{k-\Delta T}\| - \sum_{i=k-\Delta T}^{k-1} \|\tilde{\theta}_i - \theta_{k-\Delta T}\| \]
\[
\geq \Delta T \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\| - \sum_{i=k-\Delta T}^{k-1} \|\bar{\theta}_i - \theta_{k-\Delta T}\| \\
- \sum_{i=k-\Delta T}^{k-1} \left\| \phi_i \phi_i^T \tilde{\theta}_i - \gamma \beta N_i \right\| \\
\geq \Delta T \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\| - \sum_{i=k-\Delta T}^{k-1} \|\bar{\theta}_i - \theta_{k-\Delta T}\| \\
- \sum_{i=k-\Delta T}^{k-1} \|\phi_i \phi_i^T \tilde{\theta}_i - \gamma \beta \sum_{i=k-\Delta T}^{k-1} N_i \right\| \\
\geq \Delta T \|\theta_{k-\Delta T} - \theta_{k-\Delta T}\|
\[- \Delta T \sup_{i \in [k - \Delta T, k - 1]} \| \theta_i - \theta_{k - \Delta T} \|
\]
\[- \Delta T \sup_{i \in [k - \Delta T, k - 1]} \| \theta_i - \theta_{k - \Delta T} \| \]
\[- \gamma \beta \sum_{i = k - \Delta T}^{k - 1} \left\| \frac{\phi_i \phi_i^\top \tilde{\theta}_i}{N_i} \right\| \]
\[\geq \Delta T \| \theta_{k - \Delta T} - \theta_{k - \Delta T} \| - \Delta T \sum_{i = k - \Delta T}^{k - 2} \| \theta_{i + 1} - \theta_i \| \]
\[- \Delta T \sum_{i = k - \Delta T}^{k - 2} \| \theta_{i + 1} - \theta_i \| - \gamma \beta \sum_{i = k - \Delta T}^{k - 1} \left\| \frac{\phi_i \phi_i^\top \tilde{\theta}_i}{N_i} \right\| \]
\[\geq \Delta T \| \theta_{k - \Delta T} - \theta_{k - \Delta T} \| - \gamma \Delta T \sum_{i = k - \Delta T}^{k - 1} \left \| \frac{\phi_i \phi_i^\top \tilde{\theta}_{i + 1}}{N_i} \right \| \]
\[- \beta \Delta T \sum_{i = k - \Delta T}^{k - 1} \left\| \beta (\tilde{\theta}_i - \theta_i) - \gamma \beta \frac{\phi_i \phi_i^\top \tilde{\theta}_i}{N_i} \right\| \]
\[- \gamma \beta \sum_{i = k - \Delta T}^{k - 1} \left\| \frac{\phi_i \phi_i^\top \tilde{\theta}_i}{N_i} \right\| \]
\[\geq \Delta T \| \theta_{k - \Delta T} - \theta_{k - \Delta T} \| \]
\[- \gamma \Delta T \sum_{i = k - \Delta T}^{k - 1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta^*)^\top \phi_i \right\| \]
\[- \gamma \Delta T \sum_{i = k - \Delta T}^{k - 1} \frac{1 - \beta}{\sqrt{N_i}} \left\| (\theta_i - \theta_i)^\top \phi_i \right\| \]
\[- \beta^2 \sum_{i = k - \Delta T}^{k - 1} \left\| \tilde{\theta}_i - \theta_i \right\| - \gamma \beta^2 \Delta T \sum_{i = k - \Delta T}^{k - 1} \left\| \frac{\phi_i \phi_i^\top \tilde{\theta}_i}{N_i} \right\| \]
\[- \gamma \beta \sum_{i = k - \Delta T}^{k - 1} \left\| \frac{\phi_i \phi_i^\top \tilde{\theta}_i}{N_i} \right\| \]
\[\geq \Delta T \| \theta_{k - \Delta T} - \theta_{k - \Delta T} \| - \gamma \Delta T W_2 - \gamma \Delta T W_1 \]
\[- \beta^2 W_1 - \gamma \beta (1 + \beta \Delta T) \left( \sum_{i = k - \Delta T}^{k - 1} \left\| \phi_i \phi_i^\top \tilde{\theta}_i \right\| \right) \]

The term $W_3$ can be upper-bounded by

\[W_3 \leq \sum_{i = k - \Delta T}^{k - 1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta^*)^\top \phi_i \right\| \]
\[+ \sum_{i = k - \Delta T}^{k - 1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta_i)^\top \phi_i \right\| \]
\[\leq \sum_{i = k - \Delta T}^{k - 1} \frac{1}{\sqrt{N_i}} \left\| (\theta_i - \theta^*)^\top \phi_i \right\| \]
+ \sum_{i=k-\Delta T}^{k-1} \frac{1}{\sqrt{N_i}} \left\| \phi_i^T \left( \tilde{\theta}_i - \vartheta_i + \gamma \beta \frac{\phi_i^T \hat{\theta}_i}{N_i} \right) \right\| \\
\leq W_2 + W_1 + \gamma \beta W_3

Therefore, we have

\[ W_3 \leq \frac{1}{1 - \gamma \beta} (W_1 + W_2). \]

Thus

\[ W_1 \geq \Delta T \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \| - \gamma \Delta T W_2 - \gamma \Delta T W_1 \]
\[ - \beta^2 W_1 - \gamma \beta (1 + \beta \Delta T) \frac{1}{1 - \gamma \beta} (W_1 + W_2) \]

and

\[ W_1 \geq \Delta T - \gamma \Delta T \left( \frac{\beta (1 + \beta \Delta T)}{1 - \gamma \beta} \right) \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \| \]
\[ \geq \xi \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \| \]

From (26), \( \xi > 0 \). Considering (28),

\[ V_k - V_{k-\Delta T} \leq -c_4 \frac{1}{\Delta T} W_1^2 \]
\[ \leq -c_4 \frac{1}{\Delta T} \xi^2 \| \theta_{k-\Delta T} - \vartheta_{k-\Delta T} \|^2 \]
\[ \leq -c_4 \frac{1}{\Delta T} \xi^2 (1 - \lambda) \gamma V_{k-\Delta T} \]

And therefore

\[ V_k \leq \left[ 1 - c_4 \frac{1}{\Delta T} \xi^2 (1 - \lambda) \gamma \right] V_{k-\Delta T} \]  \hspace{1cm} (34)

**Summary**

Considering all of the above cases, and specifically (29), (31), (33) and (34), we have

\[ V_k \leq (1 - \mu) V_{k-\Delta T}, \]

where \( \mu \) is defined in (27). Collecting terms, we obtain

\[ V_k \leq \exp \left( -\mu \left[ \frac{k}{\Delta T} \right] \right) V_0. \]

\[ \square \]

6 Numerical Simulations

6.1 Output Error Convergence with Constant Inputs

Consider a linear regression problem with parameter \( \theta^* = [2, -3, 1]^T \). In iteration 3, the regressor \( \phi_k \) becomes a constant value of \([1, -2, 1]^T\). In iteration 51, the regressor becomes a constant value of
Figure 1: Output errors of the three algorithms.

$[2, -1, -2]^T$. We compare the three algorithms in (9), (10) and (23). The hyperparameters are chosen as follows:

$$\beta = 0.9$$
$$\gamma = 0.1238$$
$$\alpha = \gamma \beta = 0.1114$$

From [1] we can see that both HB and HT are faster than normalized gradient descent algorithm.

6.2 Parameter Error Convergence with Persistent Excitation

Consider the same regression problem with parameter $\theta^* = [2, -3, 1]^T$ but now with persistently exciting regressors. The inputs for identifying parameters are defined as

$$\phi_k = [1, 2\sin(k), 2\cos(k)].$$

We adopt the same hyperparameters as defined in [35]. Figure 2 shows parameter error convergence for the three algorithms.

From the simulation, we can see that both HB and HT have very similar convergence behavior. The two algorithms are faster than GD in terms of parameter convergence.

7 Conclusion

Two algorithms, the heavy ball method and the high-order tuner have been discussed in this paper. Under persistent excitation, the parameters of the two algorithms are converging to zero exponentially fast. Compared to the gradient descent algorithm, the two algorithms have been shown to have accelerated learning rate. For future work, we will look into how a regularization term in the loss function can affect the overall performance of the mentioned algorithms.
Figure 2: Parameter errors of the three algorithms.

References

[1] K. S. Narendra and A. M. Annaswamy, Stable Adaptive Systems. Dover, 2005.
[2] P. A. Ioannou and J. Sun, Robust Adaptive Control. PTR Prentice-Hall, 1996.
[3] I. D. Landau, R. Lozano, M. M’Saad, and A. Karimi, Adaptive Control: Algorithms, Analysis and Applications. Springer Science & Business Media, 2011.
[4] G. C. Goodwin and K. S. Sin, Adaptive Filtering Prediction and Control. Prentice Hall, 1984.
[5] R. Ortega, “On morse’s new adaptive controller: parameter convergence and transient performance,” IEEE Transactions on automatic control, vol. 38, no. 8, pp. 1191–1202, 1993.
[6] A. S. Morse, “High-order parameter tuners for the adaptive control of linear and nonlinear systems,” in Systems, Models and Feedback: Theory and Applications. Birkhauser Boston, 1992, pp. 339–364.
[7] J. E. Gaudio, “Fast learning and adaptation in control and machine learning,” Ph.D. dissertation, Massachusetts Institute of Technology, 2020.
[8] Y. Cui, J. E. Gaudio, and A. M. Annaswamy, “New algorithms for discrete-time parameter estimation,” arXiv preprint arXiv:2103.16653, 2021.
[9] A. M. Annaswamy, A. Guha, Y. Cui, J. E. Gaudio, and J. M. Moreu, “Online algorithms and policies using adaptive and machine learning approaches,” arXiv preprint arXiv:2105.06577, 2021.
[10] B. T. Polyak, “Some methods of speeding up the convergence of iteration methods,” Ussr computational mathematics and mathematical physics, vol. 4, no. 5, pp. 1–17, 1964.
[11] Y. Nesterov, “A method of solving a convex programming problem with convergence rate $O(1/k^2)$,” Soviet Mathematics Doklady, vol. 27, pp. 372–376, 1983.
[12] W. Su, S. Boyd, and E. Candes, “A differential equation for modeling nesterov’s accelerated gradient method: theory and insights,” *Advances in neural information processing systems*, vol. 27, 2014.

[13] A. Wibisono, A. C. Wilson, and M. I. Jordan, “A variational perspective on accelerated methods in optimization,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. E7351–E7358, 2016.