# Accelerated Concurrent Learning Algorithms via Data-Driven Hybrid Dynamics and Nonsmooth ODEs

# (Extended Version with Proofs)

**Daniel E. Ochoa*** DANIEL.OCHOA@COLORADO.EDU

**Jorge I. Poveda*** JORGE.POVEDA@COLORADO.EDU

**Anantharaman Subbaraman**† ANANTHARAMAN.SUBBARAMAN@US.BOSCH.COM

**Gerd S. Schmidt**† GERDSIMON.SCHMIDT@DE.BOSCH.COM

**Farshad R. Pour Safaei**† FARSHAD.RAMEZAN@US.BOSCH.COM

* *Department of Electrical, Computer and Energy Engineering, University of Colorado, Boulder.*

† *Bosch Research and Technology Center, Sunnyvale, CA.*

## Abstract

We introduce a novel class of accelerated data-driven concurrent learning algorithms. These algorithms are suitable for the solution of high-performance system identification and parameter estimation problems with *convergence certificates* in settings where the standard persistence of excitation condition is difficult to guarantee or verify *a priori*. To achieve (uniform) fast and fixed-time convergence, the proposed algorithms exploit the existence of information-rich data sets, as well as certain non-smooth regularizations of dynamical systems that generate a family of non-Lipschitz systems modeled as data-driven ordinary differential equations (DD-ODEs) and/or data-driven hybrid dynamical systems (DD-HDS). In each scenario, we provide stability and convergence certificates via Lyapunov theory. Moreover, to illustrate the practical advantages of the proposed algorithms, we consider an online estimation problem in Lithium-Ion batteries where the satisfaction of the persistence of excitation condition is in general difficult to guarantee.

**Keywords:** Concurrent learning, adaptive control, hybrid systems, Lyapunov theory.

## 1. Introduction

In this paper, we study efficient algorithms for online parameter estimation problems which can be cast as uncertain linear parametric models of the form

$$y(t) = \phi(t)^\top \theta^*, \tag{1}$$

where $y : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a measurable signal, $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is a uniformly bounded vector-valued regressor function, and $\theta^* \in \mathbb{R}^n$ is an unknown parameter that we want to estimate. This problem plays an important role in different areas, such as adaptive control Ioannou and Sun (2012), model-free optimization of dynamical systems Poveda et al. (2020), and reinforcement learning Kaelbling et al. (1996), to name just a few. To achieve online parameter estimation with convergence and robustness certificates, different feedback-based algorithms have been proposed during the last three decades; see Boyd and Sastry (1986); Narendra and Annaswamy (1987, 2012). It is well-known that most of the adaptive estimation dynamics that achieve uniform convergence to the true parameter $\theta^*$ require a persistence of excitation (PE) condition in the regressor $\phi$, of the form

$$\int_t^{t+T} \phi(s)\phi(s)^\top ds \succ \varepsilon I, \quad \forall\, t \geq t_0, \text{ where } \; T, \varepsilon > 0. \tag{2}$$

Indeed, in several adaptive estimation dynamics the PE condition has been shown to be *sufficient* and *necessary* to achieve (uniform) exponential convergence. This includes, the so-called gradient method Praly (2017):

$$\dot{\hat{\theta}} = -\sigma\phi(t)\left(\phi(t)^\top\hat{\theta} - y(t)\right), \qquad \sigma \in \mathbb{R}_{>0}, \tag{3}$$

which has been widely used in academic and industrial applications. To relax the PE condition, the works Chowdhary and Johnson (2010), Chowdhary et al. (2012b), and Chowdhary et al. (2012a) introduced a class of concurrent learning (CL) adaptive dynamics that incorporate a sequence of recorded data $\{\phi(t_k)\}_{k=1}^N$ that is "sufficiently rich", resulting in a data-driven ordinary differential equation (DD-ODE) of the form

$$\dot{\hat{\theta}} = -\sigma\phi(t)\left(\phi(t)^\top\hat{\theta} - y(t)\right) - \rho\sum_{k=1}^{N}\phi(t_k)\left(\phi(t_k)^\top\hat{\theta} - y(t_k)\right), \tag{4}$$

where $\sigma \in \mathbb{R}_{\geq 0}$ and $\rho \in \mathbb{R}_{>0}$ are tunable gains[1]. These types of algorithms have been extended in several directions to develop PE-free adaptive dynamics in the context of model-reference adaptive control Chowdhary and Johnson (2010), reinforcement learning Kamalapurkar et al. (2014), extremum seeking control Poveda et al. (2020), and general networked estimation problems Javed et al. (2021), to name just a few examples. However, by removing (or relaxing) the PE condition, these types of data-driven algorithms can also suffer from poor transient performance in terms of slow rates of convergence, especially when the matrix of recorded data is ill-conditioned. This behavior stems from the fact that systems of the form (3) or (4) can be cast as *time-varying* gradient flows for which the Hessian matrix might be degenerate whenever the PE condition is relaxed. Indeed, the slow learning rates that may emerge in CL have limited its application in practical engineering problems that require *fast* adaptation and/or estimation.

Motivated by this background, in this paper we introduce a novel class of concurrent learning algorithms able to achieve acceleration and/or fixed-time convergence properties. The dynamics make use of different types of regularization mechanisms that have been explored during the last years to design optimization algorithms and feedback controllers with high-transient performance, but which have never been studied in the context of CL. Given that the proposed dynamics are non-smooth, they are modeled either as non-Lipschitz ODEs Khalil (2002) or as hybrid dynamical systems Goebel et al. (2012). For these systems, we exploit hybrid Lyapunov-based methods to establish suitable stability and convergence properties. Moreover, to illustrate the performance of our algorithms, we study a battery

1. In this conference paper we focus exclusively on the case where $\sigma = 0$.

estimation problem where the satisfaction of the PE condition is usually difficult to verify *a priori*, and where the standard concurrent learning algorithm of Chowdhary and Johnson (2010) can lead to prohibitively slow convergence rates. As evidenced by the numerical experiments, the proposed fast algorithms significantly outperform the standard CL dynamics in terms of transient performance and steady state error.

The rest of this paper is organized as follows: Section 2 introduces the notation used in the paper, as well as some preliminaries on hybrid dynamical systems. Section 3 presents the main results, Section 4 presents the numerical experiments, and Section 5 ends with some conclusions. The proofs of the results are presented in the Appendix.

## 2. Notation and Preliminaries on Hybrid Dynamical Systems

We define $\mathbf{c}_n \in \mathbb{R}^n$ as the vector with all entries equal to $c \in \mathbb{R}$, and use $|\cdot|$ as the Euclidean norm. We use $|x|_{\mathcal{A}} := \inf_{y\in\mathcal{A}} |x-y|$ to denote the distance of a vector $x \in \mathbb{R}^n$ with respect to a closed set $\mathcal{A}$. To simplify notation, and given vectors $x \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$, and $z \in \mathbb{R}^{n_3}$, we use $(x, y, z)$ to denote the column vector $[x^\top, y^\top, z^\top]^\top$. We use $I_n \in \mathbb{R}^{n\times n}$ to denote the identity matrix, and $S^n$ to denote the n$^{th}$-Cartesian product of the set $S$.

### 2.1. Hybrid Dynamical Systems

In this paper, we will model our algorithms as Hybrid Dynamical Systems with time-varying flows, of the form (see Goebel et al. (2012)):

$$(x, s) \in C \times \mathbb{R}_{\geq 0}, \qquad \dot{x} = F(x, s), \qquad \dot{s} = 1, \tag{5a}$$

$$(x, s) \in D \times \mathbb{R}_{\geq 0}, \qquad x^+ = G(x), \qquad s^+ = s, \tag{5b}$$

where $x \in \mathbb{R}^m$ is the main state of the system, $s$ is an auxiliary state used to model the evolution of the continuous time, $F : \mathbb{R}_{\geq 0} \times \mathbb{R}^m \to \mathbb{R}^m$ is called the flow map, and $G : \mathbb{R}^m \to \mathbb{R}^m$ is called the jump map. The sets $C$ and $D$, called the flow set and the jump set, respectively, condition the points in $\mathbb{R}^m$ where the system can *flow* or *jump* via equations (5a) or (5b), respectively. In this way, the HDS can be represented by the notation $\mathcal{H} := \{C, F, D, G\}$. Systems of the form (5) can be seen as generalizations of purely continuous-time systems ($D = \emptyset$) and purely discrete-time systems ($C = \emptyset$). Solutions to HDS of the form (5) are parameterized by both a continuous-time index $t \in \mathbb{R}_{\geq 0}$, which increases continuously during flows, and a discrete-time index $j \in \mathbb{Z}_{\geq 0}$, which increments by one during jumps. Thus, the notation $\dot{x}$ in (5a) represents the derivative $\frac{dx(t,j)}{dt}$; and $x^+$ in (5b) represents the value of $x$ after an instantaneous jump, i.e., $x(t, j + 1)$. Naturally, solutions $(x, s) : \mathrm{dom}(x, s) \to \mathbb{R}^m$ to (5) are defined on hybrid time domains. For a precise definition of hybrid time domains and solutions to HDS (5) we refer the reader to (Goebel et al., 2012, Ch.2). In some cases, our models will not depend on the auxiliary state $s$.

### 2.2. Stability and Convergence Notions

To model the different convergence properties of our algorithms, we make use of class $\mathcal{KL}$ functions $\beta$, which are continuous functions that satisfy $\lim_{r\to 0^+} \beta(r, \nu) = 0$ for each fixed $\nu \in \mathbb{R}_{\geq 0}$, $\lim_{\nu\to\infty} \beta(r, \nu) = 0$ for each fixed $r \in \mathbb{R}_{\geq 0}$, and which are non-decreasing in its first argument, and non-increasing in the second argument. Class $\mathcal{KL}$ functions are standard in the feedback control; see Khalil (2002); Goebel et al. (2012). Moreover, these functions can model different types of convergence properties depending on the structure of $\beta$.

**Definition 1** *Let $\mathcal{A} \subset \mathbb{R}^m$ be a closed set, and suppose every solution of (5) satisfies $|x(t,j)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t+j), \quad \forall \quad (t,j) \in dom(x)$. Then, the set $\mathcal{A}$ is said to be:*

*(a) Uniformly Globally Asymptotically Stable (UGAS) if $\beta$ is of class $\mathcal{KL}$.*

*(b) Uniformly Globally Exponentially Stable (UGES) if $\beta(r,s) = c_1 r e^{-c_2 s}$, $c_1, c_2 > 0$.*

*(c) Uniformly Globally Finite-time Stable (UGFS) if it is UGAS and there exists a continuous function $T : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\lim_{s \to T(r)} \beta(r,s) = 0$.*

*(d) Uniformly Globally Fixed-time Stable (UGFXS) if it is UGFS and there exists $T^* > 0$ such that $T(r) < T^*$ for all $r \in \mathbb{R}_{\geq 0}$.*

Note that all the properties listed in Definition 1 are stronger than standard convergence notions used in offline optimization or estimation algorithms. In particular, UGAS implies not only convergence in the standard limiting sense, but also *uniform* global stability (in the sense of Lyapunov) *and uniform* global attractivity.

## 3. Accelerated Adaptive Concurrent Learning Dynamics

To describe the dynamics considered in this paper for the estimation of $\theta^*$ in (1), let the mappings $\Psi : \mathbb{R} \to \mathbb{R}^n$ and $B : \mathbb{R}^n \to \mathbb{R}^n$ be defined as

$$\Psi(t) = \frac{\phi(t)}{\left(1 + \phi(t)^\top \phi(t)\right)^2}, \quad \text{and} \quad B(\hat{\theta}) \coloneqq \sum_{k=1}^{N} \Psi(t_k)\Big(\phi(t_k)^\top \hat{\theta} - y(t_k)\Big). \tag{6}$$

Using (6) and (1), the DD-ODE (4) can be written as a time-invariant dynamical system of the form

$$(\hat{\theta}, s) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}, \qquad \dot{\hat{\theta}} = -\sigma A(s, \hat{\theta}) - \rho B(\hat{\theta}), \quad \dot{s} = 1, \tag{7}$$

where $A(s,\hat{\theta}) \coloneqq \Psi(s)(\phi(s)^\top \hat{\theta} - y(s))$. Taking system (7) as a benchmark, we will construct four different data-driven CL dynamics that will achieve (uniform) global asymptotic convergence, exponential convergence, finite-time convergence, and fixed-time convergence, respectively, to the true parameter $\mathcal{A}_0 \coloneqq \{\theta^*\}$. The convergence of these dynamics will depend on the "richness" properties of the available recorded data, a notion that is captured by a finite-time version of persistence of excitation; see Astrom and Wittenmark (1989).

**Assumption 3.1** *Let $\{\phi(t_k)\}_{k=1}^N$ be a sequence of recorded data. Then, the matrix $D \coloneqq [\phi(t_1),\ \phi(t_2),\ \cdots,\ \phi(t_N)] \in \mathbb{R}^{n \times N}$ satisfies $rank(D) = n$.*

Sequences of data satisfying Assumption 3.1 are said to be *sufficiently-rich* (SR). The following lemma provides an equivalent (and instrumental) characterization of SR data.

**Lemma 2** *Let $\{\phi(t_k)\}_{k=1}^N$ be a sequence of recorded data, and let $P \coloneqq \sum_{k=1}^N \frac{\phi(t_k)\phi(t_k)^\top}{\left(1+\phi(t_k)^\top \phi(t_k)\right)^2}$. Then $\{\phi(t_k)\}_{k=1}^N$ is SR if and only if there exists $\gamma \in \mathbb{R}_{>0}$ such that $P \succeq \gamma I_n$.*

We call the constant $\gamma$ the *level of richness* of the data $\{\phi(t_k)\}_{k=1}^N$.

### 3.1. Data-Driven Accelerated Hybrid Dynamics with Periodic Restarting

The first dynamical system that we consider is inspired by Nesterov's ODEs studied in the context of accelerated optimization; see Su et al. (2016) and Wibisono et al. (2016). Such algorithms can induce suitable acceleration properties by incorporating dynamic momentum, emulating in continuous time the acceleration properties of Nesterov's accelerated optimization algorithm; see Nesterov (2004). However, unlike the results of Su et al. (2016) and Wibisono et al. (2016), in the setting of CL we are also interested in establishing suitable robustness properties that are relevant in applications where noisy measurements are unavoidable. Such robustness properties can be obtained by endowing the dynamics with discrete-time restarting mechanisms that persistently reset the momentum coefficient/state of the dynamics. The combination of continuous-time and discrete-time dynamics leads to a hybrid regularization of the Nesterov's discrete-time algorithm which, for time-invariant problems, has been modeled as a HDS of the form (5) in Poveda and Li (2019), Ochoa et al. (2019), and Ochoa et al. (2021). Based on this setting, the *hybrid accelerated concurrent learning* (HACL) dynamics that we consider in this paper are modeled by a HDS with state $x := (\hat{\theta}, p, \tau)$, where $\hat{\theta}$ is the estimation state, $p \in \mathbb{R}^n$ is the momentum state, and $\tau \in \mathbb{R}_{>0}$ is a resetting state. The dynamics are given by

$$x \in C := \left\{x \in \mathbb{R}^{2n+2} : \tau \in [\delta, \Delta]\right\}, \ \begin{pmatrix} \dot{x} \\ \dot{s} \end{pmatrix} = \begin{pmatrix} \dot{\hat{\theta}} \\ \dot{p} \\ \dot{\tau} \\ \dot{s} \end{pmatrix} = F(x,s) := \begin{pmatrix} \frac{2}{\tau}\left(p - \hat{\theta}\right) \\ -2k\tau\left(\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})\right) \\ \frac{1}{2} \\ 1 \end{pmatrix}, \tag{8a}$$

$$x \in D := \left\{x \in \mathbb{R}^{2n+2} : \tau = \Delta\right\}, \qquad \begin{pmatrix} x^+ \\ s^+ \end{pmatrix} = \begin{pmatrix} \hat{\theta}^+ \\ p^+ \\ \tau^+ \\ s^+ \end{pmatrix} = G(x) := \begin{pmatrix} \hat{\theta} \\ (1-q)p + q\hat{\theta} \\ \delta \\ s \end{pmatrix}, \tag{8b}$$

where $k \in \mathbb{R}_{>0}$ is a tunable gain, $\infty > \Delta > \delta > 0$ are tunable parameters that describe how frequently the algorithm resets, and $q \in \{0, 1\}$ is a Boolean variable that characterizes the resetting policy of the algorithm. In particular, when $q = 0$ the HACL only resets the coefficient $\tau$, whereas when $q = 1$ the algorithm also resets the momentum state $p$. By construction, the discrete-time updates of the system are periodic and separated by intervals of flow of duration $2(\Delta - \delta)$. To guarantee suitable convergence properties, we will impose the following "data-driven" condition on the parameters $(\delta, \Delta)$ and the gains $(k, \rho)$.

**Assumption 3.2** *The tunable parameters $(\delta, \Delta, k)$ satisfy $2k\rho(\Delta^2 - \delta^2) > \frac{1}{\gamma}$, where $\gamma > 0$ is given by Lemma 2.*

The following theorem, which is the first main result of this paper, characterizes the convergence properties of the HACL dynamics. Throughout the paper, we use $\tilde{\theta} := \hat{\theta} - \theta^*$ to denote the estimation error, and for simplicity we consider $\sigma = 0$. All the proofs are presented in the Appendix.

**Theorem 3.1** *Suppose that Assumptions 3.1 and 3.2 hold. Then, every maximal solution of system (8) has an unbounded time domain, and the closed set $\mathcal{A} := \mathcal{A}_0 \times \mathcal{A}_0 \times [\delta, \Delta] \times \mathbb{R}_{\geq 0}$ is UGAS. Moreover, for each compact set of initial conditions $K \subset C \cup D$, the following convergence properties hold for all $(t, j) \in dom(x, s)$:*

*(a) If $q = 0$, then for each $j \in \mathbb{Z}_{\geq 0}$ there exists $\beta_j \in \mathbb{R}_{>0}$ such that each trajectory of the system satisfies the bound*

$$|\tilde{\theta}(t,j)|^2 \leq \frac{\beta_j}{k\rho\tau(t,j)^2}. \tag{9}$$

*for all $(t,j) \in dom(x)$.*

*(b) If $q = 1$, the set $\mathcal{A}$ is UGES, and each trajectory of the system satisfies the bound*

$$|\tilde{\theta}(t,j)| \leq k_0\tilde{\gamma}^j|\tilde{\theta}(0,0)|, \quad \text{where} \quad \tilde{\gamma} = \sqrt{\frac{1}{k\rho\Delta^2}\left(\frac{1}{2\gamma} + k\rho\delta^2\right)} \in (0,1), \quad k_0 = \frac{\Delta}{\delta}. \tag{10}$$

*for all $(t,j) \in dom(x)$.*

The result of Theorem 3.1 establishes two main convergence properties: item (a) establishes that the estimation error decreases at a rate of approximately $O(1/\tau^2)$ during flows, where $\tau$ increases linearly with time. On the other hand, item (b) establishes exponential convergence with a convergence rate adjustable via the values of $(\delta, \Delta, k, \rho)$. In this case, information-rich data sets ($\gamma \gg 1$) lead to faster rates of convergence. Optimal restarting periods, similar to those studied in O'Donoghue and Candes (2013) and Poveda and Li (2021), can also be derived for system (8).

### 3.2. Data-Driven Accelerated Hybrid Dynamics with Adaptive Restarting

The HACL dynamics (8) implement a periodic restarting mechanism that is coordinated by the state $\tau$. In this subsection, we now consider an alternative approach based on *adaptive restarting*, where the momentum state is reset whenever a certain state-dependent condition is satisfied. Such type of mechanisms have been studied in the optimization literature; see Su et al. (2016), O'Donoghue and Candes (2013), Teel et al. (2019). However, in the context of CL, these types of mechanisms have remained mostly unexplored. To study this case, we introduce the function $H(\hat{\theta}, p) = \frac{|\tilde{\theta}|_P^2}{2} + \frac{1}{2}|p|^2$, where $\tilde{\theta} := \hat{\theta} - \theta^*$, and where $|\cdot|_P : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a data-induced norm defined as $|u|_P^2 := u^\top Pu$ for all $u \in \mathbb{R}^n$, with $P$ as defined in Lemma 2. We then consider the *Hybrid Hamiltonian Concurrent Learning* (HHCL) algorithm, with $x = (\hat{\theta}, p, \tau) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_{\geq 0}$, and dynamics given by

$$x \in C := C_0 \times [0,\ \Delta], \qquad \begin{pmatrix} \dot{\hat{\theta}} \\ \dot{p} \\ \dot{\tau} \end{pmatrix} = \begin{pmatrix} 0 & k\rho I_p & 0 \\ -k\rho I_p & 0 & 0 \\ 0 & 0 & k\rho \end{pmatrix} \begin{pmatrix} \frac{\partial H}{\partial p} \\ \frac{\partial H}{\partial \hat{\theta}} \\ 1 \end{pmatrix}, \tag{11a}$$

$$x \in D := (C_0 \times \{\Delta\}) \cup (D_0 \times [0, \Delta]), \quad \begin{pmatrix} \hat{\theta}^+ \\ p^+ \\ \tau^+ \end{pmatrix} = \begin{pmatrix} I_p & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\theta} \\ p \\ \tau \end{pmatrix}, \tag{11b}$$

where $\Delta := \frac{n\pi}{2k\rho\sqrt{\gamma}}$, $\gamma$ is the level of richness of the data, and where

$$C_0 := \left\{(\hat{\theta}, p) \ : \ \langle B(\hat{\theta}), p\rangle \leq 0\right\}, \quad D_0 := \left\{(\hat{\theta}, p) \ : \ \langle B(\hat{\theta}), p\rangle = 0 \ \&\ |p|^2 \geq \left|B(\hat{\theta})\right|^2/\overline{\lambda}\right\},$$

with $\overline{\lambda} \geq \lambda_{\max}(P)$. Given that $\frac{\partial H}{\partial \hat{\theta}} = B(\hat{\theta})$ and $\frac{\partial H}{\partial p} = p$, the construction of the sets $C_0$ and $D_0$ indicate that system (11) is allowed to flow whenever there is no increase in the *potential* energy of the data-induced Hamiltonian function $H$.

**Remark 3.1** *The role of the timer $\tau$ in system (11) is to guarantee the existence of an initial reset after an interval of flow of duration $\Delta > 0$. Once this reset has occurred, the update $p^+ = 0$ will guarantee that the next reset of the system will happen before $\tau = \Delta$, i.e., due to the condition $|p|^2 \geq \left|B(\hat{\theta})\right|^2/\overline{\lambda}$ for all $x \in D$. Such types of bounds on the reset times have also been used in Teel et al. (2019) for standard optimization problems. However, in the context of CL, their application is new. In particular, note that the parameter $\Delta$ in the jump set $D$ is now data-dependent.*

The following theorem is the second main result of this paper.

**Theorem 3.2** *Suppose that Assumption 3.1 holds. Then, system (11) renders the set $\mathcal{A}_H = \mathcal{A}_0 \times \{0\} \times [0, \Delta]$ UGES, and every solution has an unbounded time-domain satisfies the bound*

$$|\tilde{\theta}(t,j)| \leq \sqrt{\frac{2c_0}{\gamma}} \min\left\{1, e^{-\frac{\alpha}{2}(t-\tilde{\Delta})}\right\} |\tilde{\theta}(0,0)| \tag{12}$$

*for all $(t,j) \in dom(x)$, where $\alpha = \frac{1}{\Delta}\ln\left(1 + \frac{\gamma}{\overline{\lambda}}\right)$, $\tilde{\Delta} = 2\Delta$ and $c_0 > 0$.*

### 3.3. Finite-Time and Fixed-Time Concurrent Learning Dynamics

While the hybrid CL dynamics (8) and (11) can induce sublinear and linear convergence rates, the convergence properties of the algorithms are still of *asymptotic* nature, i.e., $\theta(t) \to \theta^*$ only as $t \to \infty$. In this subsection, we consider a different class of learning dynamics able to achieve exact convergence to the true parameter $\theta^*$ in a *finite* amount of time. Moreover, in some cases this finite time can be upper bounded by a constant independent of the initial conditions of the estimate $\hat{\theta}$, which leads to *fixed-time* convergence guarantees.

In particular, to achieve finite time convergence we consider the *Finite-Time Concurrent Learning* (FTCL) dynamics modeled by the following non-smooth DD-ODE with $s \in \mathbb{R}_{\geq 0}$:

$$\hat{\theta} \in C := \mathbb{R}^n, \qquad \dot{\hat{\theta}} = -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^{\frac{1}{2}}}, \qquad \dot{s} = 1, \tag{13}$$

where $(k, \sigma, \rho) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ are tunable gains, where the pair $(A, B)$ is defined in (6) and (7), and where we let $\dot{\hat{\theta}} := 0$ at the unique equilibrium point $\theta^*$. Note that this system is not Lipschitz continuous. However, under Assumption 3.1 and the uniform boundedness of the regressor $\phi$, the vector field is everywhere continuous in $\hat{\theta}$ due to the structure of $A$ and $B$. Thus, existence of solutions is guaranteed. For this system we establish the following result, which is the third main contribution of the paper.

**Theorem 3.3** *Suppose that Assumption 3.1 holds. Then, system (13) renders the set $\mathcal{A}_0 \times \mathbb{R}_{\geq 0}$ UGFTS, every solution has an unbounded time domain, and the settling time function satisfies $T(\hat{\theta}(0)) \leq \frac{2}{k\gamma\rho}|P|^{1/2}\sqrt{|\hat{\theta}(0) - \theta^*|}$.*

The result of Theorem 3.3 guarantees that $\hat{\theta}(t) = \theta^*$ for all $t \geq T(\hat{\theta}(0))$, where $T(\hat{\theta}(0))$ depends on the initial conditions of the estimate $\hat{\theta}(0)$, as well as the level of richness $\gamma$ of the data. To remove the dependence on $\hat{\theta}(0)$ we can further consider a class of *Fixed-Time Concurrent Learning* (FXCL) dynamics, modeled by the following nonsmooth DD-ODE:

$$\hat{\theta} \in C := \mathbb{R}^n, \quad \dot{\hat{\theta}} = -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^a} - k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^{-a}}, \quad \dot{s} = 1, \tag{14}$$

where $(k, \sigma, \rho) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ are tunable gains, $a \in (0, 1)$ is a tunable exponent, and where we let $\dot{\hat{\theta}} := 0$ at the unique equilibrium point $\theta^*$. System (14) is not Lipschitz continuous. However, under Assumption 3.1 and the uniform boundedness of the regressor, it is everywhere continuous in $\hat{\theta}$. The following theorem is the fourth main result of this paper.

**Theorem 3.4** *Suppose that Assumption 3.1 holds. Then, system (14) renders UGFXS the set $\mathcal{A}_0 \times \mathbb{R}_{\geq 0}$ with $T^* = \frac{\pi}{2a\gamma\rho k}$, and every solution has an unbounded time-domain.*

**Remark 3.2** *Note that the fixed-time $T^*$ is independent of the initial estimate $\hat{\theta}(0)$, but dependent on the level of richness of the data of the regressor, i.e., dependend on $\gamma > 0$.*

### 3.4. Discussion

The results of Theorems 3.1-3.4 establish new convergence bounds for CL algorithms that explicitly show the dependence on the richness of the data, i.e., the constant $\gamma$. In particular, while the standard CL dynamics of Chowdhary and Johnson (2010) achieve exponential convergence with rate of convergence proportional to the level or richness of the matrix $P$ (cf. Lemma 2), under suitable tuning of the restarting parameters the data-driven hybrid dynamics introduced in this paper can achieve rates of convergence proportional to the squared root of the level of richness of the matrix $P$, see Poveda and Li (2019), Teel et al. (2019) and Poveda and Li (2021). This acceleration property is induced by the addition of momentum to the dynamics, and the design of the flow set and the jump set. Similarly, for the non-smooth DD-ODES (13) and (14), our results establish finite and fixed-time convergence bounds that are similar to those obtained in Poveda and Krstic (2021) for adaptive model-free optimization, but which are new in the context of CL, with an explicit characterization of the convergence time in terms of the level of richness of the data. Finally, note that for applications where $\gamma$ is small the bound (9) establishes a desirable "semi-acceleration" property for estimation problems that lead to convex optimization problems that are not necessarily strongly convex.

## 4. Numerical Experiments: Li-ion battery parameter estimation

To illustrate our algorithms in a practical context, we apply the methods to solve an estimation problem related to the characterization of the impedance parameters of an equivalent circuit model of a Lithium-Ion (Li-Ion) battery. These types of batteries have strict requirements to achieve a reliable life-time operation, see (Reddy and Linden, 2011, Part 4) and (Plett, 2016, Volume II, Chapter 1). In this setting, it is customary to design real-time algorithms that control the charge and estimate internal states based on a dynamic

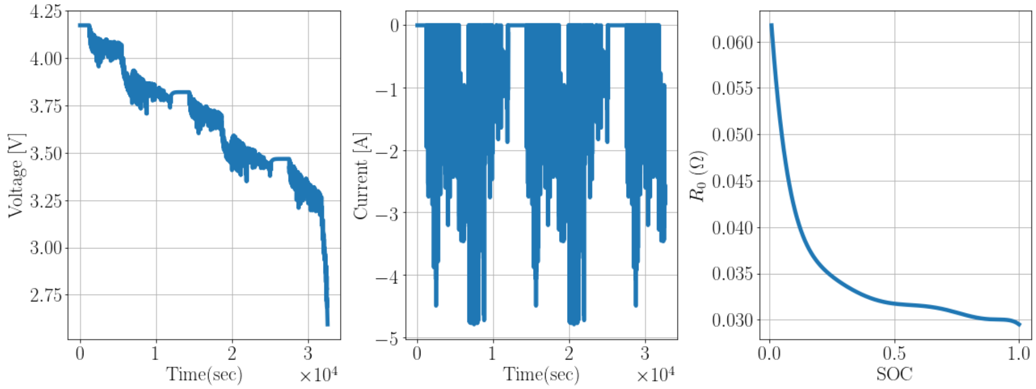


Figure 1: Battery model simulation with nominal value of $R_0$

model of the battery. Under limited computational capability, circuit based models are typically used in battery management systems. A first order equivalent circuit model of a Li-ion battery is described below Feng et al. (2015); Liaw et al. (2004): $\dot{z} = \frac{I}{C_0}$, $\dot{i}_1 = \frac{I-i_1}{R_1 C_1}$, $V = \Phi(z) + IR_0 + i_1 R_1$, where $C_0[As]$ denotes the battery capacity, $R_0[\Omega]$,$R_1[\Omega]$,$C_1[F]$ denote the battery impedance parameters, $z$ denotes the state of charge (SOC) of the battery, i.e. the relative capacity, $I[A]$ denotes the input current, $i_1[A]$ is the current through the parallel resistor $R_1$, and $\Phi$ is mapping from state of charge to open circuit voltage. The impedance parameters $R_0, R_1, C_1$ are typically functions of the state of charge Hannan et al. (2017). For ease of presentation, we make some simplifying assumptions regarding the estimation problem. We note that the methods described here can be extended to the case without such assumptions, whereas standard methods like Recursive least squares are not suitable due to the form of the parameter dependency on the battery state of charge. The capacity $C_0$ evolves on a slower time scale, and it is typically estimated using rest voltage measurements Subbaraman et al. (2019), which decouples it from the battery impedance parameter estimation which can change over a single charge/discharge cycle due to variations of state of charge, temperature, etc. If $C_0$, the initial SOC, and the OCV-SOC relation are known at least over a cycle, the battery SOC can be calculated accurately using Coulomb counting. Thus, we assume that the battery capacity and the initial SOC are known, and the estimation is restricted to battery impedance parameters, for which the functional dependency of SOC is restricted to $R_0$, whereas $R_1$ and $C_1$ are constant. Thus, we write $R_0$ as $R_0(z) = \sum_{i=0}^{N} \alpha_i z^i$, and we define the parameter vector $w = [\alpha_0, ..., \alpha_N, R_1]$. Then, the system's model can be rewritten as $(V - \Phi(z)) = [I, .., z^N I, i_1]w$. The initial SOC is 1.0, $C_0 = 3.4$Ah, and $R_1 C_1 = 100$. A sample drive-cycle behavior and the nominal value of $R_0$ as a function of the SOC are illustrated in Figure 1.

**Simulation Results:** We select the regressor vector as $\phi = \left[I, Iz, \cdots, Iz^N, i_1\right]^\top$, with $N = 7$, where $I$ and $i_1$ are the input-current and the current through resistor $R_1$, respectively. To achieve good numerical performance, in addition to being compliant with Assumption 3.1, we pay special attention to the convergence rates of the different dynamics and how they are closely related to the magnitude of the minimum eigenvalue of the matrix $P$. With this in mind, we aim to select data that is both SR, and which at the same time maximizes said eigenvalue. To do this, we follow the procedure presented in Chowdhary and Johnson

(2011), and we find a subset of 50 data points from the battery sample drive-cycle behavior. We highlight that the resulting $\lambda_{\min}(P) \approx 10^{-6}$, and even though this value can increase slightly when larger data-sets are considered, the size is kept small in order to facilitate faster computation, specially for the FTCL and FXCL algorithms which involve calculating the norm of $B$. With the selected data-set we implement the different methods presented in Section 3 and compare them with the standard concurrent learning dynamics (4). The results are presented in Figure 2. All the proposed algorithms outperform the standard CL approach for the number of iterations considered.

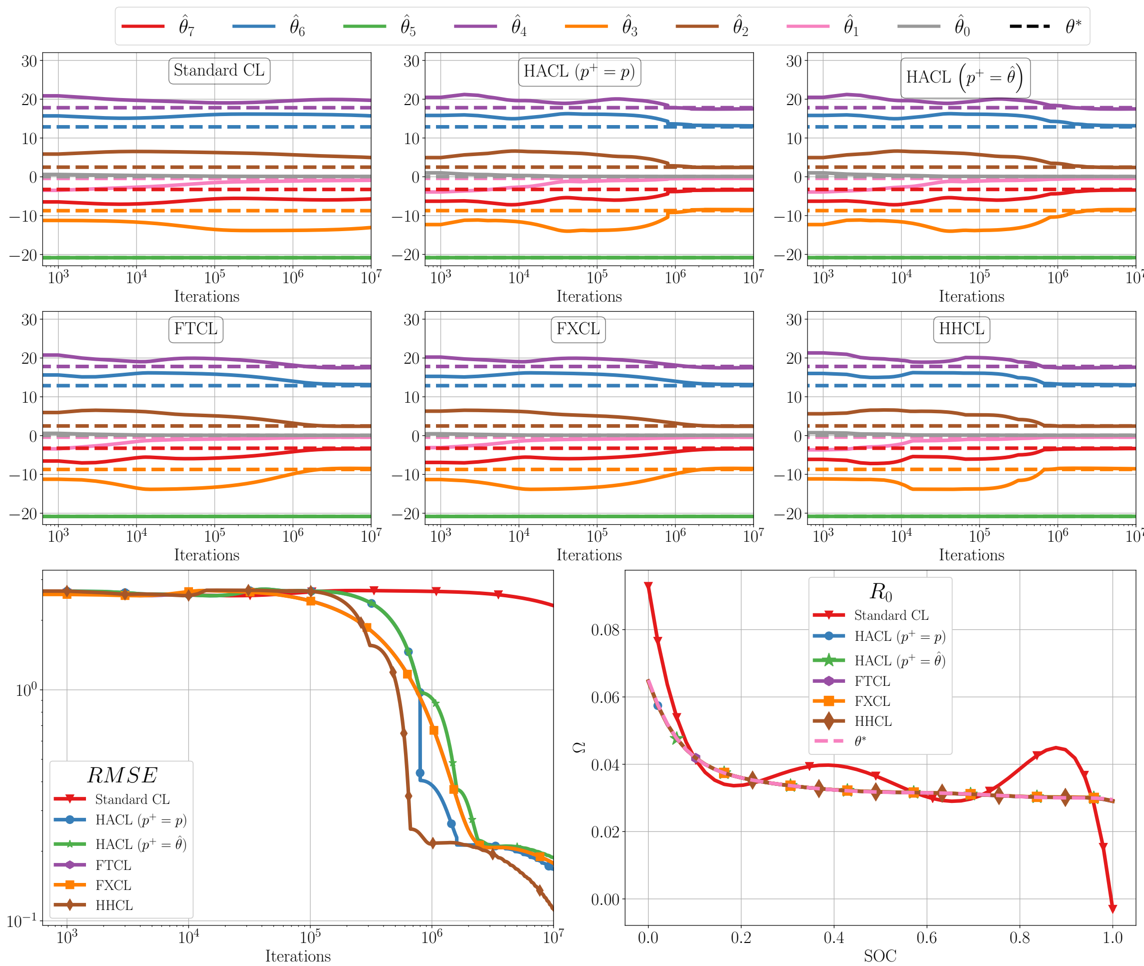


Figure 2: Estimation of parameter $R_0$ in a Li-Ion battery using the proposed algorithms.

## 5. Conclusions

In this paper, we introduced a new class of concurrent learning algorithms with acceleration and finite/fixed-time convergence properties. The algorithms are suitable for identification and parameter estimation problems that arise in the context of adaptive control, model-free optimization, and reinforcement learning. The proposed algorithms are modeled as non-smooth ODEs or hybrid dynamical systems, for which suitable stability, convergence, and robustness properties can be established via Lyapunov-based tools and invariance principles. We illustrated the advantages of the methods via numerical examples in a Lithium-Ion battery estimation problem where standard concurrent learning dynamics can generate prohibitively slow rates of convergence. Future research directions will consider data-driven accelerated adaptive controllers for dynamical systems, as well as extensions to *multi-agent* problems and algorithms based on *multi-time scale* stochastic hybrid dynamics.

## References

K. J. Astrom and Bjorn Wittenmark. *Adaptive Control*. Addison-Wesley Publishing Company, 1989.

S. P. Bhat and D. S. Bernstein. Finite-time stability of continuous autonomous systems. *SIAM J. Control Optim.*, 38:751–766, 2000.

S. Boyd and S. S. Sastry. Necessary and sufficient conditions for parameter convergence in adaptive control. *Automatica*, 22(6):629–639, 1986.

G. Chowdhary and E. Johnson. Concurrent learning for convergence in adaptive control without persistence of excitation. *49th IEEE Conference on Decision and Control*, pages 3674–3679, 2010.

G. Chowdhary and E. Johnson. A singular value maximizing data recording algorithm for concurrent learning. In *Proceedings of the 2011 American Control Conference*, pages 3547–3552. IEEE, 2011.

G. Chowdhary, T. Wu, M. Cutler, N. K. Ure, and J. P. How. Experimental results of concurrent learning adaptive controllers. *AIAA Guidance, Navigation and Control Conference,*, pages 1–14, 2012a.

G. Chowdhary, T. Yucelen, M. Muhlegg, and E. N. Johnson. Concurrent learning adaptive control of linear systems with exponentially convergent bounds. *International Journal of Adaptive Control and Signal Processing*, 27(4):280–301, 2012b.

T. Feng, L. Yang, X. Zhao, H. Zhang, and J. Qiang. Online identification of lithium-ion battery parameters based on an improved equivalent-circuit model and its implementation on battery state-of-power prediction. *Journal of Power Sources*, 281:192–203, 2015.

K. Garg and D. Panagous. Fixed-time stable gradient flows: Applications to continuous-time optimization. *IEEE Transactions on Automatic Control*, 65(5):2002–2015, 2021.

R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012.

M. A Hannan, MS. H Lipu, A. Hussain, and Azah Mohamed. A review of lithium-ion battery state of charge estimation and management system in electric vehicle applications: Challenges and recommendations. *Renewable and Sustainable Energy Reviews*, 78:834–854, 2017.

P. A. Ioannou and J. Sun. *Robust Adaptive Control*. Dover Publications Inc., Mineola, NY., 2012.

M. U. Javed, J. I. Poveda, and X. Chen. Excitation conditions for uniform exponential stability of the cooperative gradient algorithm over weakly connected digraphs. *IEEE Control and Systems Letters, DOI:10.1109/LCSYS.2021.3049153*, pages 1–6, 2021.

L. Pack Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

R. Kamalapurkar, J. R. Klotz, and W. E. Dixon. Concurrent learning-based approximate feedback-nash equilibrium solution of n-player nonsero-sum differential games. *IEEE/CAA Journal of Automatica Sinica*, 1:239–247, 2014.

H. K. Khalil. *Nonlinear Systems*. Prentice Hall, Upper Saddle River, NJ, 2002.

B. Y. Liaw, G. Nagasubramanian, R. G Jungst, and D. H Doughty. Modeling of lithium ion cells—a simple equivalent-circuit model approach. *Solid state ionics*, 175(1-4):835–839, 2004.

K. S. Narendra and A. Annaswamy. Persistent excitation in adaptive systems. *International Journal of Control*, 45(1):127–160, 1987.

K. S. Narendra and A. M. Annaswamy. *Stable Adaptive Systems*. Courier Corporation, 2012.

Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic course*. Kluwer Academic Publishers, Boston, MA., 2004.

D. Ochoa, J. I. Poveda, C. Uribe, and N. Quijano. Optimal resource allocation with momentum. *IEEE Conference on Decision and Control*, pages 3954–3959, 2019.

D. E. Ochoa, J. I. Poveda, C. Uribe, and N. Quijano. Robust optimization over networks using distributed restarting of accelerated dynamics. *IEEE Control Systems Letters, DOI 10.1109/LCSYS.2020.3001632*, pages 1–6, 2021. Extended technical report available at: http://deot95.github.io/technicalreports/lcsys2021.

O'Donoghue and E. J. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2013.

S. Parsegov, A. Polyakov, and P. Shcherbakov. Nonlinear fixed-time control protocol for uniform allocation of agents on a segment. *IEEE Conference on Decision and Control*, pages 7732–7737, 2012.

G. L. Plett. *Battery Management Systems - Volume I - Battery Modeling & Volume II - Equivalent-Circuit Methods*. Artech House, 2016.

A. Polyakov. Nonlinear feedback design for fixed-time stabilization of linear control systems. *IEEE Transactions on Automatic and Control*, 57(8):2106–2110, 2012.

J. I. Poveda and M. Krstic. Non-smooth extremum seeking control with user-prescribed fixed-time convergence. *IEEE Transactions on Automatic Control, 10.1109/TAC.2021.3063700*, 2021.

J. I. Poveda and N. Li. Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization. *58th IEEE Conference on Decision and Control*, pages 3000–3005, 2019.

J. I. Poveda and N. Li. Robust hybrid zero-order optimization algorithms with acceleration via averaging in time. *Automatica*, 123, 2021.

J. I. Poveda, M. Benosman, and K. Vamvoudakis. Data-Enabled Extremum Seeking: A Cooperative Concurrent Learning-Based Approach. *International Journal of Adaptive Control and Signal Processing, DOI: 10.1002/acs.3189*, pages 1–29, 2020.

L. Praly. Convergence of the gradient algorithm for linear regression models in the continuous and discrete time cases. *Research Report, PSL Research Unversity, Mines ParisTech*, 2017.

T. B. Reddy and D. Linden, editors. *Linden's Handbook of Batteries*. McGraw Hill, 2011.

W. Su, S. Boyd, and E. Candes. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *J. of Machine Learning Research*, 17(153):1–43, 2016.

A. Subbaraman, N. Ravi, R. Klein, G. S. Schmidt, and C. Mayhew. Method and system for estimating battery open cell voltage, state of charge, and state of health during operation of the battery. *US Patent 10,312,699*, 2019.

A. R. Teel, J. I. Poveda, and J. Le. First-order optimization algorithms with resets and hamiltonian flows. *In proc. of IEEE Conference on Decision and Control*, pages 5838–5843, 2019.

Andrew R Teel, Fulvio Forni, and Luca Zaccarian. Lyapunov-based sufficient conditions for exponential stability in hybrid systems. *IEEE Transactions on Automatic Control*, 58(6):1591–1596, 2012.

A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47): E7351–E7358, 2016.

## Appendix A. Proofs

In this section, we present the main stability and convergence proofs of the algorithms.

### A.1. Accelerated Hybrid Dynamics with Periodic Restarting

#### A.1.1. Proof of Theorem 3.1

We will establish the UGAS result by using hybrid Lyapunov functions. In particular, we consider the function

$$V(x) = \frac{|p-\hat{\theta}|^2}{4} + \frac{|p-\theta^*|^2}{4} + \frac{k\rho\tau^2}{2}\left|\tilde{\theta}\right|_P^2. \tag{15}$$

Using the definition of $P$ and $|\cdot|_P$, we have

$$\frac{|\tilde{\theta}|_P^2}{2} = \frac{|\tilde{\theta}^\top P\tilde{\theta}|}{2} \leq \frac{\lambda_{\max}(P)|\tilde{\theta}|^2}{2}. \tag{16}$$

Moreover, using the fact that the data is SR with level of richness $\gamma > 0$, we have that:

$$\frac{|\tilde{\theta}|_P^2}{2} = \frac{\tilde{\theta}^\top P\tilde{\theta}}{2} \geq \gamma\frac{\tilde{\theta}^\top\tilde{\theta}}{2} \implies \frac{\gamma}{2}|\tilde{\theta}|^2 \leq \frac{|\tilde{\theta}|_P^2}{2}. \tag{17}$$

Therefore, for all $x \in C \cup D$, using the fact that $\tau(t,j) \geq \delta$ for all $(t,j) \in \mathrm{dom}(x)$, we obtain:

$$\begin{aligned} V(x) &\geq \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{|p-\theta^*|^2}{4} + \frac{k\rho\delta^2}{2}|\tilde{\theta}|_P^2 \geq \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{|p-\theta^*|^2}{4} + \gamma k\rho\delta^2\frac{\left|\hat{\theta}-\theta^*\right|^2}{2} \\ &\geq \frac{\left|p-\hat{\theta}\right|^2}{4} + \min\left\{\frac{1}{4},\ \frac{\gamma k\rho\delta^2}{2}\right\}\left(|p-\theta^*|^2 + \left|\hat{\theta}-\theta^*\right|^2\right) \geq \underline{c}|x|_{\mathcal{A}}^2, \end{aligned} \tag{18}$$

where $\underline{c} := \min\left\{\frac{1}{4},\ \frac{\gamma k\rho\delta^2}{2}\right\}$. On the other hand, using (16) and the fact that $\tau(t,j) \leq \Delta$ for all $(t,j) \in \mathrm{dom}(x)$, we obtain

$$V(x) \leq \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{|p-\theta^*|^2}{4} + \frac{k\rho\Delta^2}{2}\left|\tilde{\theta}\right|_P^2 \leq \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{|p-\theta^*|^2}{4} + k\rho\lambda_{\max}(P)\Delta^2\frac{\left|\hat{\theta}-\theta^*\right|^2}{2}. \tag{19}$$

Furthermore, using the inequality $\left|p-\hat{\theta}\right|^2 \leq 2\left(\left|\hat{\theta}-\theta^*\right|^2 + |p-\theta^*|^2\right)$, we have that

$$\begin{aligned} V(x) &\leq \frac{1}{2}\left(\left|\hat{\theta}-\theta^*\right|^2 + |p-\theta^*|^2\right) + \frac{|p-\theta^*|^2}{4} + k\rho\lambda_{\max}(P)\Delta^2\frac{\left|\hat{\theta}-\theta^*\right|^2}{2} \\ &= \frac{3}{4}|p-\theta^*|^2 + \frac{1}{2}\left(1 + k\rho\lambda_{\max}(P)\Delta^2\right)\left|\hat{\theta}-\theta^*\right|^2 \leq \bar{c}|x|_{\mathcal{A}}^2, \end{aligned} \tag{20}$$

where $\overline{c} \coloneqq \max\left\{\frac{3}{4}, \frac{1}{2}\left(1 + k\rho\lambda_{\max}(P)\Delta^2\right)\right\}$. Similarly, using $|p-\theta^*|^2 \le 2\left(\left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2\right)$, we obtain:

$$\begin{aligned}
V(x) &\le \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{1}{2}\left(\left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2\right) + k\rho\lambda_{\max}(P)\Delta^2\frac{\left|\hat{\theta}-\theta^*\right|^2}{2}, \\
&= \frac{3}{4}\left|p-\hat{\theta}\right|^2 + \frac{1}{2}\left(1 + k\rho\lambda_{\max}(P)\Delta^2\right)\left|\hat{\theta}-\theta^*\right|^2 \le \overline{c}\left(\left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2\right), \\
&\implies \left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2 \ge \frac{V(z)}{\overline{c}}. \qquad (21)
\end{aligned}$$

Now, let $\eta \coloneqq 1 - \frac{\delta}{\Delta^2} - \frac{1}{2k\gamma\rho\Delta^2}$, $\tilde{\rho} \coloneqq \min\left\{\frac{1}{\Delta}, \frac{k\gamma\rho\delta}{2}\right\}$ and $\lambda \coloneqq \min\left\{\frac{\tilde{\rho}}{\overline{c}}, -\log(1-\eta)\right\}$. Note that $\lambda \in \mathbb{R}_{>0}$, since $\frac{\rho}{\overline{c}} > 0$ and $\eta \in (0,1)$ by Assumption 3.2. Additionally, note that we can write $B$ as

$$B(\hat{\theta}) = \sum_{k=1}^{N} \frac{\phi(t_k)}{\left(1+\phi(t_k)^\top\phi(t_k)\right)^2}\left(\phi(t_k)^\top\hat{\theta} - \phi(t_k)\theta^*\right) = \sum_{k=1}^{N} \frac{\phi(t_k)\phi(t_k)^\top}{\left(1+\phi(t_k)^\top\phi(t_k)\right)^2}\tilde{\theta} = P\tilde{\theta}. \qquad (22)$$

We now proceed to compute the time-derivative of $V$ along the trajectories generated by the flows of the Accelerated Hybrid Dynamics with Periodic Restarting. Using $\sigma = 0$, we can write $\dot{V}$ as:

$$\begin{aligned}
\dot{V}(x) &= \nabla V(x)^\top \dot{x}, \\
&= \left[-\frac{(p-\hat{\theta})}{2} + k\rho\tau^2 P\tilde{\theta},\ \frac{(p-\hat{\theta})}{2} + \frac{(p-\theta^*)}{2},\ k\rho\tau\left|\tilde{\theta}\right|_P^2\right]^\top \dot{x}, \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} + 2k\rho\tau(p-\hat{\theta})^\top P\tilde{\theta} - k\rho\tau(p-\hat{\theta})^\top P\tilde{\theta} - k\rho\tau(p-\theta^*)^\top P\tilde{\theta} + \frac{k\rho\tau}{2}\left|\tilde{\theta}\right|_P^2 \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - k\rho\tau\left(-(p-\hat{\theta})^\top P + (p-\theta^*)^\top P + \frac{\tilde{\theta}^\top P}{2}\right)\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - k\rho\tau\left(\tilde{\theta}^\top P - \frac{\tilde{\theta}^\top P}{2}\right)\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \frac{1}{2}k\rho\tau\left|\tilde{\theta}\right|_P^2.
\end{aligned}$$

Therefore, by using (17) together with (21), we obtain

$$\dot{V}(x) \le -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \rho k\tau\frac{\left|\tilde{\theta}\right|_P^2}{2} \le -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \frac{k\rho\gamma\tau}{2}\left|\hat{\theta}-\theta^*\right|^2,$$

$$\leq -\frac{\left|p-\hat{\theta}\right|^2}{\Delta} - \frac{k\rho\gamma\delta}{2}\left|\hat{\theta}-\theta^*\right|^2 \leq -\min\left\{\frac{1}{\Delta},\ \frac{k\rho\gamma\delta}{2}\right\}\left(\left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2\right) \leq -\frac{\tilde{\rho}}{\bar{c}}V(x). \tag{23}$$

Since by definition we have $\lambda \leq \frac{\tilde{\rho}}{\bar{c}}$, it follows that

$$\dot{V}(x) \leq -\lambda V(x), \qquad \forall x \in C. \tag{24}$$

Note that this bound implies that $V$ does not increase during flows, and that it satisfies $V(x(t,j)) \leq V(x(s,j))$ for all $t > s$ and each fixed $j$ such that $(t,j) \in \mathrm{dom}(x)$. This generates the bound (9).
On the other hand, during jumps the restarting policy $q = 0$ generates changes in the Lyapunov function given by

$$\begin{aligned} V(x^+) - V(x) &= \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{|p-\theta^*|^2}{4} + \frac{k\rho\delta^2}{2}|\tilde{\theta}|_P^2 - \frac{\left|p-\hat{\theta}\right|^2}{4} - \frac{|p-\theta^*|^2}{4} - \frac{k\rho\Delta^2}{2}|\tilde{\theta}|_P^2 \\ &= -k\rho\frac{\Delta^2-\delta^2}{2}|\tilde{\theta}|_P^2 \leq 0, \end{aligned} \tag{25}$$

which implies that $V$ does not increase during jumps. Additionally, when $q = 1$, the change of the Lyapunov function during jumps is given by

$$\begin{aligned} V(x^+) - V(x) &= \frac{\left|\hat{\theta}-\theta^*\right|^2}{4} + \frac{k\rho\delta^2}{2}|\tilde{\theta}|_P^2 - \frac{\left|p-\hat{\theta}\right|^2}{4} - \frac{|p-\theta^*|^2}{4} - \frac{k\rho\tau^2}{2}|\tilde{\theta}|_P^2 \\ &\leq \frac{|\tilde{\theta}|_P^2}{4\gamma} + \frac{k\rho\delta^2}{2}|\tilde{\theta}|_P^2 - \frac{\left|p-\hat{\theta}\right|^2}{4} - \frac{|p-\theta^*|^2}{4} - \frac{k\rho\tau^2}{2}|\tilde{\theta}|_P^2 \\ &= -\frac{\left|p-\hat{\theta}\right|^2}{4} - \frac{|p-\theta^*|^2}{4} - \left(1 - \frac{\delta^2}{\tau^2} - \frac{1}{2k\rho\gamma\tau^2}\right)\frac{\rho k\tau^2|\tilde{\theta}|_P^2}{2} \\ &\leq -\eta V(x) \\ &\implies V(x^+) \leq e^{-\lambda}V(x) \qquad \forall x \in D,\ x^+ \in G(x), \end{aligned} \tag{26}$$

where we used $(1-\eta) \leq e^{-\lambda}$ which follows from the definition of $\eta$. Inequality (26) and the non-increment of $V$ during flows implies the bound (10). Inequalities (23) and (26), together with the quadratic bounds on the Lyapunov function (18) and (20), imply UGES of the set $\mathcal{A}$ ((Teel et al., 2012, Thm. 1)). Similarly, inequalities (23) and (25) imply UGAS via the hybrid invariance principle (Goebel et al., 2012, Thm. 8.8), which in turn implies that $V \to 0^+$, i.e., the sequence $\{\beta_j\}_{j=0}^{\infty}$ with $\beta_j := V(x(t_j,j))$ and $t_j := \min\{t \in \mathbb{R}_{\geq 0} : (t,j) \in \mathrm{dom}(x)\}$, is monotonically decreasing and converges to zero.

### A.2. Hybrid Hamiltonian Concurrent Learning

#### A.2.1. Proof of Theorem 3.2

First, notice that the Hamiltonian function $H$ satisfies:

$$\dot{H}(\hat{\theta},p) = \nabla H^\top \begin{pmatrix}\dot{\hat{\theta}} \\ \dot{p}\end{pmatrix} = k\rho\left(\frac{\partial H}{\partial\hat{\theta}}\right)^\top\left(\frac{\partial H}{\partial p}\right) - k\rho\left(\frac{\partial H}{\partial p}\right)^\top\left(\frac{\partial H}{\partial\hat{\theta}}\right) = 0, \tag{27}$$

where we used $\frac{\partial H}{\partial \hat{\theta}} = B(\hat{\theta})$. To analyze the behaviour of $H$ during jumps, we follow similar arguments as in Teel et al. (2019), and we first note that for any pair of vectors $u, v \in \mathbb{R}^n$ we have that:

$$\frac{|u|_P^2}{2} = \frac{|v|_P^2}{2} + v^\top P(u-v) + \frac{1}{2}(u-v)^\top P(u-v).$$

Therefore, since the data is SR with level of richness $\gamma > 0$, we obtain:

$$\frac{|u|_P^2}{2} \geq \frac{|v|_P^2}{2} + v^\top P(u-v) + \frac{\gamma}{2}|u-v|^2. \tag{28}$$

After minimizing with respect to $u$ at both sides of (28), we find that

$$0 = \min_{u \in \mathbb{R}^n} |u|_P^2 \geq \frac{|v|_P^2}{2} - \frac{1}{2\gamma}|Pv|^2 \implies \frac{|Pv|^2}{2\gamma} \geq \frac{|v|_P^2}{2}. \tag{29}$$

Now, let $\varepsilon := \frac{1}{\frac{\bar{\lambda}}{\gamma}+1}$, and note that $\varepsilon \in (0,1)$ since $\bar{\lambda} \geq \gamma$. Hence, using (29), we can write an upper bound for the value of the Hamiltonian after a jump as follows:

$$\begin{aligned}
H(\hat{\theta}^+, p^+) &= H(\hat{\theta}, 0) \\
&= \frac{|\tilde{\theta}|_P^2}{2} \\
&= (1-\varepsilon)\frac{|\tilde{\theta}|_P^2}{2} + \varepsilon\frac{|\tilde{\theta}|_P^2}{2} \\
&\leq (1-\varepsilon)\frac{|\tilde{\theta}|_P^2}{2} + \varepsilon\frac{\left|P\tilde{\theta}\right|^2}{2\gamma} \\
&= (1-\varepsilon)\frac{|\tilde{\theta}|_P^2}{2} + \varepsilon\frac{|B(\hat{\theta})|^2}{2\gamma}.
\end{aligned} \tag{30}$$

Using the fact that

$$1 - \varepsilon = 1 - \frac{1}{\frac{\bar{\lambda}}{\gamma}+1} = \frac{\bar{\lambda}}{\gamma}\frac{1}{\frac{1}{\frac{\bar{\lambda}}{\gamma}+1}} = \varepsilon\frac{\bar{\lambda}}{\gamma}, \tag{31}$$

from (30), we obtain that

$$H(\hat{\theta}^+, p^+) \leq \frac{\varepsilon}{2\gamma}\left(\bar{\lambda}\left|\tilde{\theta}\right|_P^2 + \left|B(\hat{\theta})\right|^2\right),$$

which, using the definition of $H$, further reduces to

$$\begin{aligned}
H(\hat{\theta}^+, p^+) &\leq \frac{\varepsilon}{2\gamma}\left(\bar{\lambda}\left|\tilde{\theta}\right|_P^2 + \bar{\lambda}|p|^2\right) \\
&= \frac{\varepsilon\bar{\lambda}}{2\gamma}\left(\left|\tilde{\theta}\right|_P^2 + |p|^2\right)
\end{aligned}$$

$$= \frac{\varepsilon \bar{\lambda}}{\gamma} H(\hat{\theta}, p). \tag{32}$$

Therefore, given any arbitrary initial condition $(\hat{\theta}_0, p_0, \tau_0)$, and using (31) together with (32), after $j$ jumps we obtain

$$H(\hat{\theta}(t,j), p(t,j)) \leq \left(\varepsilon \frac{\bar{\lambda}}{\gamma}\right)^j H(\hat{\theta}_0, p_0) = (1-\varepsilon)^j H(\hat{\theta}_0, p_0) = e^{-j\Delta\alpha} H(\hat{\theta}_0, p_0), \tag{33}$$

where $\alpha \coloneqq \frac{1}{\Delta} \ln\left(\frac{1}{1-\varepsilon}\right) = \frac{1}{\Delta} \ln\left(1 + \frac{\gamma}{\bar{\lambda}}\right)$. As described in Remark 3.1 we have that $\tau(t,j) < \Delta$ for all $t + j > 0$, and thus $t < (j+1)\Delta \implies e^{-j\Delta\alpha} < e^{-\alpha(t-\Delta)}$. Using this inequality in (33), we obtain

$$H(\hat{\theta}(t,j), p(t,j)) \leq \min\left\{1, e^{-\alpha(t-\Delta)}\right\} H(\hat{\theta}_0, p_0). \tag{34}$$

Furthermore, note that, by the definition of the Hamiltonian, there exist $\kappa \coloneqq \min\{\gamma/2, 1/2\}$, and $K \coloneqq \max\left\{\lambda_{\max}(P)/2, \frac{1}{2}\right\}$, such that

$$\kappa |x|^2_{\mathcal{A}_H} \leq H(\hat{\theta}, p) \leq K |x|^2_{\mathcal{A}_H}$$

Hence, using (34) and (27), by (Teel et al., 2012, Thm. 1) the set $\mathcal{A}_H$ is UGES.
Now, we note that under the Hybrid Concurrent Learning dynamics defined in (11), by the definition of the jump map and set, for each initial condition there exists a time $t_1 \leq \Delta$ such that $x(t_1, 1) = (\theta(t_1, 1), 0, 0)$ for $\theta(t_1, 1) \in \mathbb{R}^n$. Thus, given the state value $(\theta(t_1, 1), 0, 0)$ at $(t_1, 1) \in \operatorname{dom}(x)$, and using (31) together with (32), after $j - 1$ additional jumps we have that

$$H(\hat{\theta}(t,j), p(t,j)), \leq e^{-(j-1)\Delta\alpha} H(\theta(t_1, 1), 0). \tag{35}$$

Moreover, by the definition of the Hamiltonian, we obtain that

$$\left|\tilde{\theta}(t,j)\right|^2 \leq \frac{2}{\gamma} H(\hat{\theta}, p) \qquad (t,j) \in \operatorname{dom}(x), \tag{36a}$$

$$H(\hat{\theta}(t_1, 1), 0) \leq \lambda_{\max}(P) \frac{\left|\tilde{\theta}(t_1, 1)\right|^2}{2}. \tag{36b}$$

Therefore, using (35) together with (36), we obtain

$$\left|\tilde{\theta}(t,j)\right|^2 \leq \frac{\lambda_{\max}(P)}{\gamma} \min\left\{1, e^{-\alpha(t-2\Delta)}\right\} \left|\tilde{\theta}(t_1, 1)\right|^2, \tag{37}$$

where we have used the fact that $e^{-(j-1)\Delta\alpha} < e^{-\alpha(t-2\Delta)}$. Since the flow map in (11) is globally Lipschitz, and $t_1 < \infty$, there exists $c_1 \in \mathbb{R}_{>0}$ such that

$$\left|\tilde{\theta}(t,0)\right| \leq c_1 \left|\tilde{\theta}(0,0)\right| \qquad \forall t \leq t_1. \tag{38}$$

Since $\tilde{\theta}(t,1) = \tilde{\theta}(t,0)$, from (37) and (38), we obtain that

$$\left|\tilde{\theta}(t,j)\right|^2 \leq \frac{\lambda_{\max}(P) c_1}{\gamma} \min\left\{1, e^{-\alpha(t-\tilde{\Delta})}\right\} \left|\tilde{\theta}(0,0)\right|^2,$$

which is equivalent to (12) with $c_0 \coloneqq \frac{\lambda_{\max}(P) c_1}{2}$ and $\tilde{\Delta} = 2\Delta$. This concludes the proof.

## A.3. Finite-Time Concurrent Learning

Uniqueness of the equilibrium point, and continuity of the Finite-Time Concurrent Learning dynamics defined in (13), follow similar arguments to the ones presented for the Fixed-Time Concurrent Learning dynamics, see Section A.4 for more details. We now proceed to prove the stability results for the Finite-Time dynamics.

### A.3.1. Proof of Theorem 3.3

We consider the Lyapunov function $V(\hat{\theta}) = \frac{1}{2}\left|B(\hat{\theta})\right|^2$. Using the facts that $B(\hat{\theta}) = P\tilde{\theta}$, $\sigma = 0$, and recalling that $|x|_P^2 \geq \gamma|x|^2 \quad \forall x \in \mathbb{R}^n$ whenever the data is SR with level of richness $\gamma$, we can write the time-derivative of $V$ along the trajectories of the Finite-Time Concurrent Learning dynamics (13) as follows:

$$\dot{V}(\hat{\theta}) = \nabla V(\hat{\theta})^\top \dot{\hat{\theta}} = -k\tilde{\theta}^\top P^\top P \frac{\rho P\tilde{\theta}}{|P\tilde{\theta}|^{\frac{1}{2}}} = -k\rho \frac{\left|P\tilde{\theta}\right|_P^2}{\left|\rho P\tilde{\theta}\right|^{1/2}} \leq -k\rho\gamma \frac{\left|P\tilde{\theta}\right|^2}{\left|P\tilde{\theta}\right|^{1/2}} = -k\rho\gamma\left|P\tilde{\theta}\right|^{\frac{3}{2}}. \tag{39}$$

Furthermore, by noting that $\left|P\tilde{\theta}\right| = \left(2V(\hat{\theta})\right)^{1/2}$, from (39), we obtain that

$$\dot{V}(\hat{\theta}) \leq -k\rho\gamma\left(2V(\hat{\theta})\right)^{\frac{3}{4}} = -2^{3/4}k\rho\gamma V(\hat{\theta})^{\frac{3}{4}}. \tag{40}$$

Hence, it follows by (Bhat and Bernstein, 2000, Thm. 4.2) that the set $\mathcal{A}_0$ is UGFTS with settling time function $T : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ bounded as

$$T(\hat{\theta}(0)) \leq \frac{2}{k\gamma\rho}|P|^{1/2}\sqrt{|\hat{\theta}(0) - \theta^*|}. \tag{41}$$

## A.4. Fixed-Time Concurrent Learning

### A.4.1. Uniqueness of the equilibrium point

In this section, we study the uniqueness of the equilibrium point for the Fixed-Time Concurrent Learning Dynamics. The next Lemma is instrumental for the proof.

**Lemma 3** *If the data is SR, then* $B(\hat{\theta}) = 0 \iff \hat{\theta} = \theta^*$.

**Proof** By the definition of $B(\hat{\theta})$ we have that $B(\theta^*) = 0$. Now, we prove that $B(\hat{\theta}) = 0 \implies \hat{\theta} = \theta^*$. By using (22), we have that

$$B(\hat{\theta}) = P\tilde{\theta}. \tag{42}$$

Since the data is SR with level of richness $\gamma > 0$, meaning that $P \succeq \gamma I_n$, we obtain that $\ker P = \{0_n\}$. Consequently, from (42), we obtain that

$$B(\hat{\theta}) = 0 \implies \tilde{\theta} = 0_n \implies \hat{\theta} = \theta^*.$$

This concludes the proof. ■

We now proceed to prove that the vector field describing the Finite-Time Concurrent Learning Dynamics only vanishes when the estimate is equal to the the true parameter vector.

**Lemma 4** *For any $\sigma \in \mathbb{R}_{\geq 0}$, if the data is SR, for the dynamics defined in (14) we have that $\dot{\hat{\theta}} = 0 \iff \hat{\theta} = \theta^*$.*

**Proof** The fact that the vector field vanishes at $\theta^*$ follows directly by construction. On the other hand, assume that $\dot{\hat{\theta}} = 0$ for $\hat{\theta} \neq \theta^*$. From Lemma 3 we have that $B(\hat{\theta}) = 0 \iff \hat{\theta} = \theta^*$ whenever the data is SR, and thus, that the the following implication holds:

$$0 = -k\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta}) \left( \frac{1}{|B(\hat{\theta})|^a} \right) \implies \sigma A(s,\hat{\theta}) + \rho B(\hat{\theta}), \tag{43}$$

Rewriting (43) by using (22) and the definition of $A$ we have

$$\begin{aligned} 0 &= \sigma \Psi(s) \left( \phi(s)^\top \hat{\theta} - y(s) \right) + \rho P \tilde{\theta} \\ &= \sigma \Psi(s) \left( \phi(s)^\top \hat{\theta} - \phi(s)^\top \theta^* \right) + \rho P \tilde{\theta} \\ &\implies \left( \sigma \Psi(s) \phi(s)^\top + \rho P \right) \tilde{\theta} = 0. \end{aligned} \tag{44}$$

Note that for all $x \in \mathbb{R}^n$ we have that

$$\begin{aligned} x^\top \left( \Psi(s)\phi(s)^\top \right) x &= x^\top \frac{\phi(s)\phi(s)^\top}{(1+\phi(s)^\top\phi(s))^2} x \\ &= \frac{\left|\phi(s)^\top x\right|^2}{(1+\phi(s)^\top\phi(s))^2} \\ &\geq 0 \qquad \forall s \in \mathbb{R}_{\geq 0} \end{aligned}$$

$$\implies \Psi(s)\phi(s)^\top \succeq 0, \tag{45}$$

i.e., that the matrix $\Psi(s)\phi(s)^\top$ is positive semidefinite. On the other hand, by assumption, the data is SR with level of richness $\gamma > 0$. Thus, we have that $P \succeq \gamma I$ for $\gamma \in \mathbb{R}_{>0}$, and hence that

$$\sigma \Psi(s)\phi(s)^\top + \rho P \succeq \rho \gamma I, \tag{46}$$

since $(\sigma, \rho) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$. By virtue of (46), $\ker\left(\Psi(s)\phi(s)^\top + \rho P\right) = \{0_n\}$, and thus, we obtain that (44) holds only if $\hat{\theta} = \theta^*$, which constitutes a contradiction. Hence, there is no value of $\hat{\theta} \neq \theta^*$ for which we have that $\dot{\hat{\theta}} = 0$. This concludes the proof. ■

### A.4.2. Continuity of Fixed-Time Concurrent Learning Dynamics

In this section, we prove that the vector field defined in (14) is continuous. To do so, we use the following Lemma.

**Lemma 5** *If the data is SR, $\lim_{\hat{\theta} \to \theta^*} \frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0$ and $\lim_{\hat{\theta} \to \theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0 \quad \forall a \in (0,1)$.*

**Proof** We follow similar arguments as in Garg and Panagous (2021). First, by continuity of the $\ell_2$-norm we note that

$$\begin{aligned}\left|\lim_{\hat{\theta}\to\theta^*}\frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a}\right| &= \lim_{\hat{\theta}\to\theta^*}\left|\frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a}\right|,\\ &= \lim_{\hat{\theta}\to\theta^*}\frac{1}{\left|B(\hat{\theta})\right|^a}\left|B(\hat{\theta})\right|,\\ &= \lim_{\hat{\theta}\to\theta^*}\left|B(\hat{\theta})\right|^{1-a},\\ &= \left|\lim_{\hat{\theta}\to\theta^*}B(\hat{\theta})\right|^{1-a} = 0, \qquad \forall a\in(0,1).\end{aligned}$$

Hence, by positive-definiteness of the norm, we obtain that

$$\lim_{\hat{\theta}\to\theta^*}\frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0, \qquad \forall a\in(0,1). \tag{47}$$

On the other hand, by assumption, we have that $\phi$ is uniformly bounded, and thus, there exists $M\in\mathbb{R}_{\geq 0}$ such that $\left|\Psi(s)\phi(s)^\top\right|\leq M$. With this in mind, using (22) and rewriting $A$ in terms of its definition we find

$$\left|B(\hat{\theta})\right|^a = \left|P\tilde{\theta}\right|^a \geq \gamma^a\left|\tilde{\theta}\right|^a, \tag{48}$$

$$A(s,\hat{\theta}) = \Psi(s)\left(\phi(s)^\top\hat{\theta}-y(s)\right) = \Psi(s)\phi(s)^\top\tilde{\theta} \implies \left|A(s,\hat{\theta})\right|\leq M\left|\tilde{\theta}\right|. \tag{49}$$

Hence, using (49) and (48), we have that

$$0\leq\frac{\left|A(s,\hat{\theta})\right|}{\left|B(\hat{\theta})\right|^a}\leq\frac{M}{\gamma^a}\left|\hat{\theta}-\theta^*\right|^{1-a} \qquad \forall\hat{\theta}\in\mathbb{R}^n,\ a\in(0,1). \tag{50}$$

Moreover, using continuity of the $\ell_2$-norm together with (50), we obtain that

$$\left|\lim_{\hat{\theta}\to\theta^*}\frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a}\right| = \lim_{\hat{\theta}\to\theta^*}\frac{\left|A(s,\hat{\theta})\right|}{\left|B(\hat{\theta})\right|^a}\leq\lim_{\hat{\theta}\to\theta^*}\frac{M}{\gamma^a}\left|\hat{\theta}-\theta^*\right|^{1-a} = \frac{M}{\gamma^a}\lim_{\hat{\theta}-\theta^*}\left|\hat{\theta}-\theta^*\right|^{1-a} = 0 \qquad \forall a\in(0,1).$$

Therefore, by positive-definitness of the norm, we obtain

$$\lim_{\hat{\theta}\to\theta^*}\frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0 \qquad \forall a\in(0,1). \tag{51}$$

This concludes the proof. ■

**Lemma 6** *If the data is SR, the vector field defined in* (14) *is continuous.*

**Proof** Continuity for every $\hat{\theta} \in \mathbb{R}^n$ such that $B(\hat{\theta}) \neq 0$, follows directly from the continuity of $A$ and $B$. Hence, in order to show continuity we only need to analyze the cases where $B$ vanishes. As shown by Lemma 3, $B$ vanishes only when $\hat{\theta} = \theta^*$. Hence, using Lemma 5, we have that

$$\lim_{\hat{\theta}\to\theta^*} -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = -k\sigma \lim_{\hat{\theta}\to\theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} - k\rho \lim_{\hat{\theta}\to\theta^*} \frac{B(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0 \qquad \forall a \in (0,1) \quad (52)$$

$$\begin{aligned}
\lim_{\hat{\theta}\to\theta^*} -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{\left|B(\hat{\theta})\right|^{-a}} &= \lim_{\hat{\theta}\to\theta^*} -k\left(\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})\right)\left|B(\hat{\theta})\right|^a \\
&= -k\left(\sigma A(s,\theta^*) + \rho B(\theta^*)\right)\left|B(\theta^*)\right|^{-a} \\
&= 0, \qquad (53)
\end{aligned}$$

where in the second to last step we have used the continuity of $A$ and $B$, together with the fact that $A$ and $B$ vanish at $\hat{\theta} = \theta^*$. Equations (52) and (53), and the fact that (14) is defined to be 0 when $\hat{\theta} = \theta^*$, shows that the vector field describing the Fixed-Time Concurrent Learning dynamics is continuous at $\theta^*$, and hence, by the aforementioned arguments, that it is continuous everywhere. ■

### A.4.3. Proof of Theorem 3.4

We consider the Lyapunov function $V(\hat{\theta}) = \frac{\tilde{V}(\hat{\theta})^2}{2}$, where $\tilde{V}$ is an auxiliary function defined as $\tilde{V}(\hat{\theta}) := \frac{1}{2}\left|B(\hat{\theta})\right|^2$. Using (22) together with (17) and $\sigma = 0$, we can write the time-derivative of the Lyapunov function along the trajectories generated by the Fixed-Time Concurrent Learning dynamics as follows:

$$\begin{aligned}
\dot{V}(\hat{\theta}) &= \tilde{V}(\hat{\theta})\dot{\tilde{V}}(\hat{\theta}), \\
&= \tilde{V}(\hat{\theta})\nabla\tilde{V}(\hat{\theta})^\top\dot{\hat{\theta}}, \\
&= \tilde{V}(\hat{\theta})(P^\top P\tilde{\theta})^\top\dot{\hat{\theta}}, \\
&= -k\tilde{V}(\hat{\theta})(P^\top P\tilde{\theta})^\top\left(\frac{\rho P\tilde{\theta}}{\left|B(\hat{\theta})\right|^a} + \frac{\rho P\tilde{\theta}}{\left|B(\hat{\theta})\right|^{-a}}\right), \\
&= -k\rho\tilde{V}(\hat{\theta})\left|P\tilde{\theta}\right|_P^2\left(\frac{1}{\left|P\tilde{\theta}\right|^a} + \frac{1}{\left|P\tilde{\theta}\right|^{-a}}\right), \\
&\leq -k\gamma\rho\tilde{V}(\hat{\theta})\left|P\tilde{\theta}\right|^2\left(\frac{1}{\left|P\tilde{\theta}\right|^a} + \frac{1}{\left|P\tilde{\theta}\right|^{-a}}\right),
\end{aligned}$$

$$
\begin{aligned}
&= -k\gamma\rho\tilde{V}(\hat{\theta})\left(|P\tilde{\theta}|^{2-a} + |P\tilde{\theta}|^{2+a}\right), \\
&= -k\gamma\rho\tilde{V}(\hat{\theta})\left((2\tilde{V}(\hat{\theta}))^{1-\frac{a}{2}} + ((2\tilde{V}(\hat{\theta}))^{1+\frac{a}{2}}\right), \\
&= -k\gamma\rho(2^{1-\frac{a}{2}})\tilde{V}(\hat{\theta})^{2-\frac{a}{2}} - k\gamma\rho(2^{1+\frac{a}{2}})\tilde{V}(\hat{\theta})^{2+\frac{a}{2}}.
\end{aligned} \tag{54}
$$

By using the fact that $\tilde{V}(\hat{\theta}) = (2V(\hat{\theta}))^{1/2}$, (54) reduces to

$$
\begin{aligned}
\dot{V}(\hat{\theta}) &\leq -k\gamma\rho(2^{1-\frac{a}{2}})(2V(\hat{\theta}))^{1-\frac{a}{4}} - k\gamma\rho(2^{1+\frac{a}{2}})(2V(\hat{\theta}))^{1+\frac{a}{4}} \\
&= -k\gamma\rho(2^{2-3\frac{a}{4}})(V(\hat{\theta}))^{1-\frac{a}{4}} - k\gamma\rho(2^{2+3\frac{a}{4}})(V(\hat{\theta}))^{1+\frac{a}{4}}.
\end{aligned}
$$

The last inequality implies UGFXS via (Polyakov, 2012, Lemma 1). Moreover, by (Parsegov et al., 2012, Lemma 2), a sharp bound $T^*$ on the settling time function $T$ can be computed as

$$
T^* = \frac{\pi}{2a\gamma\rho k}. \tag{55}
$$