



**HAL**  
open science

# Dynamical Sparse Recovery with Finite-time Convergence

Lei Yu, Gang Zheng, Jean-Pierre Barbot

► **To cite this version:**

Lei Yu, Gang Zheng, Jean-Pierre Barbot. Dynamical Sparse Recovery with Finite-time Convergence. IEEE Transactions on Signal Processing, 2017, 65 (23), pp.6146-6157. 10.1109/TSP.2017.2745468 . hal-01649419

**HAL Id: hal-01649419**

**<https://inria.hal.science/hal-01649419>**

Submitted on 27 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dynamical Sparse Recovery with Finite-time Convergence

Lei Yu, Gang Zheng, Jean-Pierre Barbot

**Abstract**—Even though Sparse Recovery (SR) has been successfully applied in a wide range of research communities, there still exists a barrier to real applications because of the inefficiency of the state-of-the-art algorithms. In this paper, we propose a dynamical approach to SR which is highly efficient and with finite-time convergence property. Firstly, instead of solving the  $\ell_1$  regularized optimization programs that requires exhausting iterations, which is computer-oriented, the solution to SR problem in this work is resolved through the evolution of a continuous dynamical system which can be realized by analog circuits. Moreover, the proposed dynamical system is proved to have the finite-time convergence property, and thus more efficient than LCA (the recently developed dynamical system to solve SR) with exponential convergence property. Consequently, our proposed dynamical system is more appropriate than LCA to deal with the time-varying situations. Simulations are carried out to demonstrate the superior properties of our proposed system.

**Index Terms**—Sparse Recovery,  $\ell_1$ -minimization, Dynamical System, Finite-time Convergence

## I. INTRODUCTION

As a fundamental of Compressive Sensing (CS) theory [1], Sparse Recovery (SR), or sparse representation, has been substantially investigated in the last two decades. As a powerful tool, it has also been successfully applied in a wide range of research communities and obtained compelling results, including signal processing [1]–[5], medical imaging [6], [7], machine learning [8], [9], and computer vision [10]. In particular, the objective of SR is to find a concise representation of a signal using a few atoms from some specified (over-complete) dictionary,

$$y = \Phi x + \varepsilon$$

with  $y \in \mathbb{R}^M$  the observed measurements corrupted by some noises  $\varepsilon$ ,  $x \in \mathbb{R}^N$  the sparse representation with no more than  $s$  nonzero entries ( $s$ -sparsity) and  $\Phi \in \mathbb{R}^{M \times N}$  the dictionary (normally  $M \ll N$ ). Thus it always involves an underdetermined linear inverse problem. Providing that the Restricted Isometry Property (RIP) of dictionary is fulfilled, the unique solution is guaranteed [11].

The problem of SR is often casted as an optimization program that minimizes a cost function constructed by leveraging

the observation error term and the sparsity-inducing term [12]–[14], i.e.,

$$x^* = \arg \min_{x \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \psi(x) \quad (1)$$

and typically, the sparsity-inducing term  $\psi(x) = \|x\|_1 \triangleq \sum_i |x_i|$  and  $\lambda > 0$  is the balancing parameter. We call  $x^*$  as the *critical point*, i.e., the solution of (1). And typically, for sparse vectors  $x$  with  $s$ -sparsity, the solution will be unique providing that RIP condition for  $\Phi$  with order of  $2s$  is verified [11]. On the other hand, exploiting hierarchical Bayesian model built on the sparse signals [8], [15]–[18] results in compelling algorithms inherently with different sparsity-inducing term [16]. Moreover, the greedy algorithms are also favorable for SR due to the theoretical guarantees and high efficiency when the considered signal is highly sparse [9], [19]–[21].

Although greedy algorithms are efficient, the condition for the stably recovery of  $s$ -sparse  $x$  is generally very strong. In particular, it is showed in [22] that to guarantee a stably recovery of any  $s$ -sparse  $x$  with the orthogonal matching pursuit algorithm [20] in  $s$  iterations, the dictionary  $\Phi$  should satisfy the RIP with the restrict isometry constant  $\delta_s < 1/\sqrt{s+1}$ . Although it has been shown in [23] that stably recovery of any  $s$ -sparse  $x$  with the orthogonal matching pursuit algorithm [20] is also possible if  $\Phi$  satisfies the RIP with the restrict isometry constant  $\delta_{31s} < 1/3$ , the required number of iterations is  $30s$  which is computational expensive. Besides, other aforementioned algorithms are all *batch-based* which normally require a large number of iterations to guarantee the convergence (most of them with sublinear convergence rate) and thus with high computational complexity. It is thus implausible for the real applications where the signals are usually time-varying, such as radar imaging [24], face recognition [10], DOA estimation [5] and so on. Regarding the real applications, many “online” algorithms have been proposed recently either by generalizing  $\ell_1$  regularized LS in the manner of LMS (Least Mean Square) [25], [26] and RLS (Recursive Least Square) [27], or extending the Bayesian approaches following an adaptive framework [28], [29]. On the other hand, instead of the online algorithms, the Locally Competitive Algorithm (LCA) [30] has been proposed to solve the SR problem by exploiting the continuous dynamical systems. And recent advances in very-large-scale integration (VLSI) enables the realization of LCA with analog chips [31]. Consequently, instead of numerically calculating the matrix multiplications in the digital approaches, LCA can obtain the computation result from analog circuits which will be very efficient.

This work is supported by NSFC Grant 61401315, and the Project-sponsored by SRF for ROCS, SEM, under Grant 230303.

Lei Yu is with School of Electronic and Information, Wuhan University, Wuhan Hubei, China (email: ly.wd@whu.edu.cn).

Gang Zheng is with Non-A, INRIA Lille, France (email: gang.zheng@inria.fr).

Jean-Pierre Barbot is with Quartz EA 7393, ENSEA, Cergy Pontoise and Non-A Inria Lille, (email: barbot@ensea.fr)

Mathematically, LCA is in fact a continuous version of the iterative soft-thresholding algorithm [13], [32]. Moreover, providing that  $\Phi$  satisfying RIP, LCA guarantees an exponential convergence rate [31]. Even though armed with analog circuits, the LCA is much more efficient than its discrete version [32], the exponential rate is not enough to ensure the convergence of SR during the evolution of the LCA dynamics especially when signals varying rapidly. Consequently, the main objective of this paper is to redesign the dynamics of LCA to increase the convergence rate. As we can see that sparse recovery problem (1) is an optimization problem. Note that, except the numerical method, continuous method can be also used to solve the optimization problem, which historically has a strong link to control theory [33], [34]. In fact, in [35], the proposed LCA method exactly used control theory to solve the optimization problem (1). In order to clarify the motivation, let us firstly recall some basic backgrounds of control theory.

### A. Recall to System Stabilities

Researchers in control community are interested in stabilizing different types of dynamical systems with some proper control laws. Consider the following system:

$$\dot{u} = f(u) \quad (2)$$

with  $u \in \mathbb{R}^N$  the system state with respect to time  $t$  and denote  $u(t)$  the value of state at time instant  $t$ . For this system, we call the point  $u^* \in \mathbb{R}^N$  as an equilibrium point if  $f(u^*) = 0$ . Note that the linear time-invariant system has only one isolated equilibrium point, but nonlinear system and switched system may have more than one isolated equilibrium points. Therefore, only local stability around each equilibrium point can be analyzed. Concerning the concept of stability, different definitions are given in the literature.

**Definition 1.** System (2) is said to be:

- 1) **locally Lyapunov stable** around  $u^*$ , if for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that, if  $\|u(0) - u^*\| < \delta$ , then  $\|u(t) - u^*\| < \epsilon$ , for all  $t > 0$ ;
- 2) **locally asymptotically stable** around  $u^*$ , if there exists  $\delta > 0$  such that, if  $\|u(0) - u^*\| < \delta$ , then  $\lim_{t \rightarrow \infty} \|u(t) - u^*\| = 0$ ;
- 3) **locally finite-time stable** around  $u^*$ , if there exist  $\delta > 0$  and  $T > 0$  such that, if  $\|u(0) - u^*\| < \delta$ , then  $\|u(t) - u^*\| = 0$  for all  $t > T$ .

Lyapunov stability only requires the solution  $u(t)$  starting from the neighborhood of the equilibrium point  $u^*$  staying inside its neighborhood. Asymptotical stability needs that the trajectory of the system should converge to  $u^*$  as  $t$  tends to  $\infty$ . The strongest definition is the finite-time, which furthermore imposes that  $u(t)$  should exactly equal to  $u^*$  after a finite-time  $T$ . Moreover, the extension of local stability to global one needs just to relax the neighborhood of the equilibrium point  $u^*$  ( $\|u(0) - u^*\| < \delta$ ) by allowing all  $u(0) \in \mathbb{R}^N$ . If the system globally converges to  $u^*$ , it implies as well that  $u^*$  is the unique equilibrium point.

Without solving the differential equation, in control theory, the Lyapunov function method is widely used to determine

the types of stability for the studied system. Suppose  $u^*$  is the equilibrium point, and denote by  $e(t) = u(t) - u^*$ , the basic idea is to choose a Lyapunov function  $V(e)$  which should be locally positive definite for all  $e \neq 0$  and  $V(0) = 0$ , then system (2) is:

- 1) locally Lyapunov stable around  $u^*$ , if

$$\dot{V}(e) \leq 0, \forall e \neq 0$$

- 2) locally asymptotically stable with rate  $k$  around  $u^*$ , if

$$\dot{V}(e) \leq -kV(e), \forall e \neq 0$$

with  $k > 0$ ;

- 3) locally finite-time stable around  $u^*$ , if

$$\dot{V}(e) \leq -kV^\alpha(e), \forall e \neq 0$$

with  $k > 0$  and  $\alpha \in (0, 1)$ .

Similarly, **global stability** can be proved by choosing a globally positive definite and radically unbounded Lyapunov function, i.e.,  $V(e) \rightarrow \infty$  if  $\|e\| \rightarrow \infty$ . Besides, with the chosen  $V(e)$ , if one can only prove  $\dot{V}(e) \leq 0$ , then LaSalle Theorem can be still used to prove the asymptotical stability. It states that if the set  $\dot{V}(e) = 0$  contains only  $e = 0$ , then it is asymptotical stable.

### B. Motivations

In this paper, a new dynamical system will be proposed, of which the equilibrium point is unique and yields the solution of the optimization problem (1). Therefore, the above basic results (Lyapunov and LaSalle Theorems) from control theory will be used to analyze the performance of convergence. In order to well explain how to design a dynamical system with non-asymptotical (finite-time) convergent performance, let us consider the following two simple systems:

$$\dot{u} = -u \quad (3)$$

and

$$\dot{u} = -|u|^\alpha \text{sgn}(u), \text{ with } \alpha \in (0, 1). \quad (4)$$

It is easy to see that  $u^* = 0$  is the only equilibrium point for both systems. For system (3), choose the Lyapunov function as  $V(u) = u^2$ , we have  $\dot{V} = -2u^2 = -2V$ , thus  $u$  of system (3) asymptotically converges to the equilibrium point 0. Concerning system (4), choose as well  $V = u^2$ , which gives  $\dot{V} = -2u|u|^\alpha \text{sgn}(u) = -2V^{\frac{1+\alpha}{2}}$ . Since  $\alpha \in (0, 1)$ ,  $u$  of system (4) converges to the equilibrium point 0 after a finite-time  $T$ . Particularly, when  $\alpha = 1$ , system (4) is exactly system (3), and the finite-time convergence property is degraded to asymptotical one. In other words, by introducing the sign function (which is called as sliding mode technique in control theory), the convergence performance of the studied system can be improved.

Let us then turn to the problem (1). Motivated by the above example, the finite-time convergence can also be fulfilled by exploiting the sliding mode technique in LCA [35], which has an asymptotical (exponential) convergence property to solve the optimization problem (1). The rest of this paper is organized as follows. The new dynamical system is built in

Section II and its finite-time convergence property is proved in Section III. Relationships between our proposed method and the related works are discussed in Section IV. Simulations are implemented to verify the theorems and demonstrate the superior of our proposed system to LCA in Section V and extensions to recover time-varying sparse signals are empirically presented in Section VI. Conclusions are made in Section VII.

## II. SPARSE RECOVERY VIA DYNAMICAL SYSTEM

### A. Preliminary of LCA

Let us now first take a look at the LCA method proposed in [35] to solve the optimization problem (1):

$$\begin{cases} \tau \dot{u}(t) = -u(t) - (\Phi^T \Phi - I)a(t) + \Phi^T y \\ \hat{x}(t) = a(t) \end{cases} \quad (5)$$

where  $u \in \mathbb{R}^N$  is the state vector,  $\hat{x}$  represents the estimation of the sparse signal  $x$  of (1),  $\tau > 0$  is a time-constant determined by the physical properties of the implementing system. Since  $\tau$  always exists as a company of the derivative with respect to the time  $t$ , it can be simply set to  $\tau = 1$  for mathematical analysis and then added to the final result if time derivative exists.  $a(t) = T_\lambda(u(t))$  with  $T_\lambda(\cdot)$  is a *continuous* soft thresholding function

$$T_\lambda(u) = \max(|u| - \lambda, 0) \cdot \text{sgn}(u) \quad (6)$$

with  $\lambda > 0$ .

Then defining by  $u_i$  the  $i$ -th element of state  $u$ , we can call  $u_i$  as an *active* node if the amplitude  $|a_i(t)|$  is different from zero, otherwise we call this node *inactive*. Then define by  $\Gamma$  the set of active nodes, i.e.,  $a_\Gamma \neq 0$ , and  $\Gamma^c$  the set of inactive nodes, i.e.,  $a_{\Gamma^c} = 0$ . In order to guarantee the existence of unique solution of optimization problem (1), assumptions on  $\Phi$  should be made before going deep into analysis, where the restricted isometry property (RIP) [11] is assumed.

**Assumption 1** (RIP [11]). *Matrix  $\Phi$  satisfies the  $s$ -order of RIP condition with constant  $\delta_s \in (0, 1)$ .*

The above assumption implies that for any  $s$  sparse signals  $x$ , i.e., vectors with at most  $s$  nonzero elements, the following condition is verified

$$(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2.$$

Denote  $\Gamma$  the index set of nonzero elements for  $x$ , it implies that

$$1 - \delta_s \leq \text{eig}(\Phi_\Gamma^T \Phi_\Gamma) \leq 1 + \delta_s$$

where  $\Phi_\Gamma$  denotes the submatrix of  $\Phi$  with active nodes. Explicitly, RIP condition guarantees that eigenvalues of any Gramm matrix  $\Phi_\Gamma^T \Phi_\Gamma$  for any index set  $\Gamma$  are bounded.

Suppose that RIP of  $\Phi$  is fulfilled with constant  $\delta_s$ , the LCA system (5) converges exponentially, which can be concluded in the following theorem.

**Theorem 1** (LCA Convergence Property [35]). *If Assumption 1 holds, then LCA system (5) converges to the equilibrium point  $u^*$  exponentially fast with convergence speed  $(1 - \delta_s)/\tau$ , i.e.,  $\exists \mathcal{K} > 0$ , such that  $\forall t \geq 0$*

$$\|u(t) - u^*\|_2 \leq \mathcal{K} e^{-(1 - \delta_s)t/\tau}$$

### B. The Proposed Dynamical System

In this paper, a new dynamical system is proposed to solve the  $\ell_1$ -minimization problems (1). As stated in the last section, motivated by the sliding mode technique, a new dynamical system is constructed by introducing the parameter  $\alpha \in (0, 1]$ , i.e.,

$$\begin{cases} \tau \dot{u}(t) = -[u(t) + (\Phi^T \Phi - I)a(t) - \Phi^T y]^\alpha \\ \hat{x}(t) = a(t) \end{cases} \quad (7)$$

with  $[\cdot]^\alpha$  being a function defined as

$$[\cdot]^\alpha = |\cdot|^\alpha \cdot \text{sgn}(\cdot)$$

where  $|\cdot|, \cdot, \text{sgn}$  are all element-wise operators,  $\alpha \in \mathbb{R}_+$  denotes an exponential coefficient and

$$\text{sgn}(\omega) \begin{cases} = 1, & \text{if } \omega > 0 \\ \in [-1, 1], & \text{if } \omega = 0. \\ = -1, & \text{if } \omega < 0 \end{cases}$$

In the following sections, we will demonstrate that the new designed system (7) resolves the optimization problem (1) and converges to the equilibrium point with finite-time convergence.

**Theorem 2.** *Under Assumption 1, the state  $u(t)$  of (7) converges in finite time to its equilibrium point  $u^*$ , and  $\hat{x}(t)$  of (7) converges in finite-time to  $x$  of (1).*

**Remark 1.** *Considering the dynamic (7), even if it is not a Lipschitz function at  $u = 0$  for  $\alpha \in (0, 1)$ , it still has a unique solution (Cauchy problem). This is due to the fact that dynamic (7) is at least locally asymptotically stable at  $u = 0$  and then the only one solution is clearly  $\forall t > 0$   $u(t) = 0$  if  $u(0) = 0$ . Moreover for  $\alpha = 0$  the solution must be considered in a Filippov meaning [36].*

**Remark 2.** *When  $\alpha = 1$ , the proposed dynamical system (7) becomes exactly the same as LCA proposed in [31]. For  $\alpha = 0$  the dynamic of neuron cell becomes a first order sliding mode dynamics and chattering phenomenon occurs at equilibrium point. This is not wished in the neural network and more particularly into our proposed optimization algorithm for the problem (1).*

## III. CONVERGENCE IN FINITE TIME

In this section, we will analyze the property of the proposed system (7) in the following four steps. At first, similar to LCA, we can also prove that the output of the proposed system (7) converges to the critical point of (1). After that, we will prove that the trajectory of (7) stays in a bounded space. Then, the attractive property of an invariant set is proved via LaSalle theorem [37], [38] by introducing a new semipositive definite function. At last, the finite time convergence of (7) is proved.

In the following, note that variables  $u, x, a$  are always the function of the time  $t$  which are sometimes neglected for simplicity and the derivative with point above always means the derivative with respect to the time  $t$ .  $u_i$  represents the  $i$ -th element of vector  $u$  and  $I$  as the identity matrix.  $u^*$  is a constant with respect to time  $t$  which represents the equilibrium point of the trajectory  $u(t)$ .

### A. Solution Equivalence

Considering the proposed dynamical system (7), the second claim of Theorem 2 can be easily proved by slightly modifying the result in many papers related to LCA, such as [31]. And we re-write the following lemma to make our proofs complete.

**Lemma 1.** *Equilibrium points of (7) are critical points of (1).*

*Proof.* The subgradient of (1) with respect to  $x$  in the set valued framework [36], [39], [40] gives

$$\frac{\partial \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \|x\|_1}{\partial x} = (\Phi^T(\Phi x - y) + \lambda \text{sgn}(x))^T \quad (8)$$

And define  $x = T_\lambda(u)$ , then  $x$  and  $u$  should have the same sign. By simple calculation,

$$u - x = (|u| - \max(|u| - \lambda, 0)) \cdot \text{sgn}(u) = \lambda \text{sgn}(x)$$

Then, substitute the sgn in (8)

$$\frac{\partial (\frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \|x\|_1)}{\partial x} = (\Phi^T(\Phi x - y) + u - x)^T$$

Consequently,  $\dot{u} = 0$  in (7) only when  $\frac{\partial (\frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \|x\|_1)}{\partial x} = 0$ , this completes the proof.  $\square$

The above lemma connects the dynamical system (7) and the optimization problem (1), and guarantees the equivalence of the output of (7) and the critical point of (1). Since Assumption 1 implies the uniqueness of critical point of (1), then Lemma 1 means that the system (7) has only one equilibrium point.

**Remark 3.** *The generalized active function  $T_\lambda$  is not the main contribution of this paper, thus only soft-thresholding function is addressed. Alternating of active function will get the same result as Lemma 1, and proofs with generalized active functions can be referred in the appendix of [35].*

Due to the sgn function, the resulted system (7) is actually a hybrid (switched) system and it might exist the Zeno phenomenon (Infinite transitions within finite time [41]), which makes the analysis very complicated. Consequently, it is necessary to verify whether Zeno exists, and the following lemma verifies this point.

**Lemma 2.** *The system (7) with continuous threshold (6) is everywhere integrable and has a unique solution, moreover Zeno behavior can't occur.*

*Proof.* According to control theory, the existence and uniqueness of the solution of dynamical systems is not guaranteed only at the state point where the system is not Lipschitz. For the proposed (7), its solution exists except when  $u(t) + (\Phi^T \Phi - I)a - \Phi^T y$  is equal to zero, i.e., the equilibrium point. Nevertheless, at this equilibrium point, Lemma 1 shows that it is the unique equilibrium point of (7) which coincides with the unique critical point of (1).

As we will prove in Theorem 2 that this unique equilibrium point is globally stable, therefore, the solution of (7) with continuous threshold (6) always has a unique solution.

Moreover, since  $T_\lambda(u)$  defined in (6) is a continuous threshold, i.e.,  $T_\lambda(u) \in \mathbb{C}^0$ , then according to the definition of the

proposed dynamics in (7), the trajectory  $u$  should belong to  $\mathbb{C}^1$ , which implies that Zeno behavior does not exist for the proposed system (7) with continuous threshold (6).  $\square$

In order to invoke the LaSalle theorem in the next subsection, we must prove first that the state behavior stays in a bounded space.

**Lemma 3.** *For all bounded initial states  $u$ , the trajectory of (7) stays in a bounded set.*

*Proof.* In order to prove that the state trajectory stays on a bounded set, we invoke again (1) but with respect to  $u$  (let  $x = T_\lambda(u)$ ),

$$V(u) = \frac{1}{2} \|y - \Phi T_\lambda(u)\|_2^2 + \lambda \|T_\lambda(u)\|_1$$

and the derivatives with respect to time  $t$  is

$$\dot{V}(u) = (u + (\Phi^T \Phi - I)T_\lambda(u) - \Phi^T y)^T F'_\lambda \dot{u}$$

with  $F'_\lambda = \frac{\partial T_\lambda(u)}{\partial u}$  being the Frechet derivative with respect to  $u$ , thus it leads to a diagonal matrix with 1 on the diagonal if the neuron is active and 0 if not.

Now considering dynamical system (7), it gives

$$\begin{aligned} \dot{V} &= -(u + (\Phi^T \Phi - I)T_\lambda(u) - \Phi^T y)^T F'_\lambda \cdot \\ &\quad [u + (\Phi^T \Phi - I)T_\lambda(u) - \Phi^T y]^\alpha \\ &\leq 0 \end{aligned}$$

As  $\lim_{\|u\| \rightarrow \infty} V(u) = \infty$  and  $\dot{V}(u) \leq 0$ , one can conclude that  $u$  stays in a bounded set, i.e., (7) is Lyapunov stable.  $\square$

### B. Global Convergence

Even if LaSalle theorem requests that the state behavior must evolve in a bounded space, as this bounded space can be as wide as we want with respect to the initial state, then we consider this convergence as a global one.

On the other hand, it has been proved that under the Assumption 1, the uniqueness of the solution to (7) is guaranteed [11]. Thus it implies that there exists a solution  $u^*$  to dynamical system (7). In order to prove the convergence property of (7), the error term <sup>1</sup> is introduced.

$$\begin{aligned} \tilde{u}(t) &\triangleq u(t) - u^* \\ \tilde{a}(t) &\triangleq a(t) - a^* \end{aligned}$$

And then define the Lyapunov function with respect to  $\tilde{u}$ ,

$$E(\tilde{u}) = \frac{1}{2} \|\tilde{u}\|_2^2 + \mathbf{1}^T (\Phi^T \Phi - I)G(\tilde{u}) \quad (9)$$

with  $\mathbf{1} \in \mathbb{R}^N$  being the vector with all elements equal to 1 and  $G(\tilde{u}) = [G_1(\tilde{u}_1), G_2(\tilde{u}_2), \dots, G_N(\tilde{u}_N)]^T \in \mathbb{R}^N$ , where

$$G_i(\tilde{u}_i) = \int_0^{\tilde{u}_i} g_i(s) ds$$

<sup>1</sup>Variables  $\tilde{u}$  and  $\tilde{a}$  are always function of  $t$  which are neglected in the following sections for simplicity.

with  $g_i(s) = T_\lambda(s+u_i^*) - T_\lambda(u_i^*)$ . Then we have the following properties.

**Lemma 4.** *The function  $E$  defined in (9) satisfies the following properties:*

- 1) For all  $\tilde{u}_i \geq 0$ ,  $0 \leq G_i(\tilde{u}_i) \leq \frac{\tilde{u}_i^2}{2}$ ,
- 2)  $E$  is non-increasing, i.e.  $\dot{E} \leq 0$ ,
- 3) For dynamical system (7),  $E$  cannot be negative, i.e.  $E \geq 0$ ,
- 4) There exists a positive constant  $\nu > 0$  such that

$$E(\tilde{u}) \leq \nu \|\tilde{u}\|_2^2$$

*Proof.* 1) According to (6), the operator  $T_\lambda$  is non-decreasing, thus one can conclude that  $g_i(s) \geq 0, \forall s \geq 0$ , thus

$$G_i(\tilde{u}_i) \geq 0$$

For the second inequality, we first have

$$T_\lambda(x) - T_\lambda(y) \leq x - y, \forall x \geq y$$

which implies  $g_i(s) \leq s, \forall s \geq 0$ . Consequently,

$$G_i(\tilde{u}_i) = \int_0^{\tilde{u}_i} g_i(s) ds \leq \int_0^{\tilde{u}_i} s ds = \frac{\tilde{u}_i^2}{2}$$

where equality holds only if  $u_i^* \geq \lambda$  or  $u_i^* + \tilde{u}_i \leq -\lambda$ .

- 2) The time derivatives of  $E$  gives

$$\dot{E}(\tilde{u}) = (\tilde{u} + (\Phi^T \Phi - I)\tilde{a})^T \dot{\tilde{u}} \quad (10)$$

Then due to the fact that  $u^*$  is constant, thus

$$\dot{\tilde{u}} = \dot{u} = -[u + (\Phi^T \Phi - I)a - \Phi^T y]^\alpha \quad (11)$$

According to the definition,  $u^*$  and  $a^*$  are the equilibrium points of dynamical system (7), which concludes that

$$u^* + (\Phi^T \Phi - I)a^* - \Phi^T y = 0 \quad (12)$$

Then plug (12) into (7), we can get

$$\dot{\tilde{u}} = -[\tilde{u} + (\Phi^T \Phi - I)\tilde{a}]^\alpha \quad (13)$$

Consequently, combine equation (10) and (13), we have

$$\begin{aligned} \dot{E} &= -(\tilde{u} + (\Phi^T \Phi - I)\tilde{a})^T [\tilde{u} + (\Phi^T \Phi - I)\tilde{a}]^\alpha \\ &= -\|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_{1+\alpha}^{1+\alpha} \\ &\leq 0 \end{aligned} \quad (14)$$

3) From the result of Lemma 3, the proposed system (7) is Lyapunov stable for any initial condition, i.e.  $\tilde{u} = 0$ , which means  $E$  will converge to 0. Furthermore, we know that  $\dot{E} \leq 0$  for all  $\tilde{u}$ , so for any  $E < 0$ , it will be non-increasing all the time instead of converging to 0, i.e. the system will not converge, which is contraindicative. Thus for the proposed dynamic system (7),  $E$  is non-negative, i.e.,  $E \geq 0$ .

- 4) By definition of  $E(\tilde{u})$  in (9), we have

$$\begin{aligned} E(\tilde{u}) &= \frac{1}{2} \|\tilde{u}\|_2^2 + \mathbf{1}^T (\Phi^T \Phi - I)G(\tilde{u}) \\ &= \frac{1}{2} \|\tilde{u}\|_2^2 + \mathbf{1}^T \Phi^T \Phi G(\tilde{u}) - \mathbf{1}^T G(\tilde{u}) \\ &\geq \mathbf{1}^T \Phi^T \Phi G(\tilde{u}) \geq 0 \end{aligned}$$

For the second inequality, by exploiting the first result of this lemma, one can have

$$E(\tilde{u}) \leq \frac{1}{2} \|\tilde{u}\|_2^2 + \mathbf{1}^T (\Phi^T \Phi)G(\tilde{u})$$

According to Lemma 6 in appendix, the eigenvalue of  $\Phi^T \Phi$  is upper bounded, and thus it has

$$\mathbf{1}^T (\Phi^T \Phi)G(\tilde{u}) \leq \frac{N(1 + \delta_s)}{2s} \|\tilde{u}\|_2^2$$

thus defining by  $\rho = N/s$  the signal to sparsity rate,

$$E(\tilde{u}) \leq \frac{\rho(1 + \delta_s) + 1}{2} \|\tilde{u}\|_2^2 \quad (15)$$

Then, by defining  $\nu = \frac{\rho(1 + \delta_s) + 1}{2}$ , one can conclude the second inequality.  $\square$

According to (9) and its third property stated in Lemma 4, one can deduce that  $E$  is a positive semi-definite and radically unbounded Lyapunov function. Then armed with the second property of Lyapunov function  $E$ , we have the following theorem.

**Theorem 3.** *Under Assumption 1, the dynamical system (7) globally converges to the critical point of (1).*

*Proof.* According to the LaSalle Theorem [38], we can conclude that  $\tilde{u}$  will converge to an invariant subset  $U_{inv}$  of  $M \triangleq \{\tilde{u} | \tilde{u} + (\Phi^T \Phi - I)\tilde{a} = 0\}$ . From (14) and (13), it is easy to conclude that  $\dot{E} = 0$  implies  $\dot{\tilde{u}} = 0$ , thus all state of  $U$  is invariant. Consequently  $U_{inv} = U$ . Finally, we can conclude that  $\tilde{u}$  converges to  $U$ , and then  $a$  converge to a set of critical points of (1) i.e.,  $a^*$ , and according to Assumption 1,  $a^*$  is unique, then  $U_{inv} = U$  is reduced to a singleton  $\{\tilde{u} = 0\}$  or equivalently  $\{a = a^*\}$ .  $\square$

### C. Finite Time Convergence Property

In this subsection, the convergence property of (7) will be considered. Hereafter, we will prove then, for  $\tilde{u}$  sufficiently close to 0,  $\|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_2^2$  is not singular with respect to  $\tilde{u}$ .

**Lemma 5.** *There exist a time  $t_e < \infty$  and a positive value  $\kappa > 0$ , such that when  $t > t_e$ , the following inequality is verified,*

$$\kappa \|\tilde{u}\|_2^2 \leq \|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_2^2 \quad (16)$$

*Proof.* In order to prove this result, we should first prove the relation between  $\tilde{u}$  and  $\tilde{a}$ . According to Lemma 1, and armed with the result from [35], one can also conclude that none of switching occurs after a finite time  $t_1 < \infty$ . It means that after time  $t_1$ , every nodes  $u_i(t)$  will be with the same sign as  $u_i^*$ . Then following cases are respectively considered. For the  $i$ -th element,

- 1) if  $|u_i^*(t)| < \lambda$  we have  $|\tilde{a}_i| = |T_\lambda(\tilde{u}_i + u_i^*) - T_\lambda(u_i^*)| = 0 \leq |\tilde{u}_i|$ .
- 2) If  $|u_i^*(t)| > \lambda$  we have  $\tilde{a}_i = \tilde{u}_i + u_i^* - \lambda \cdot \text{sgn}(\tilde{u}_i + u_i^*) - u_i^* + \lambda \cdot \text{sgn}(u_i^*)$ . According to Lemma 3, the proposed

system is globally convergent. It implies that  $|\tilde{u}_i|$  can be very small, i.e., for any small  $\epsilon > 0$ , there exists a time  $t(\eta) < \infty$ , such that  $|\tilde{u}_i(t)| < \eta, \forall t > t(\eta)$ . Thus, define  $t_2 = t(\lambda - \epsilon)$ , we have  $|\tilde{u}_i(t)| < \lambda - \epsilon, \forall i$  and  $t > t_2$ , then  $\tilde{u}_i + u_i^*$  and  $u_i^*$  are with same sign, and we have  $\tilde{a}_i = \tilde{u}_i$ .

Above all, one can conclude that there exists a time  $t_e = \max\{t_1, t_2\} < \infty$ , such that for all  $t > t_e$ ,

$$\begin{cases} \tilde{a}_i = \tilde{u}_i & \text{if } i \in \Gamma \\ \tilde{a}_i = 0 & \text{if } i \in \Gamma^c \end{cases}$$

Consequently, one have

$$\begin{aligned} \|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_2^2 &= \|\Phi_\Gamma^T \Phi_\Gamma \tilde{u}_\Gamma\|_2^2 + \|\tilde{u}_{\Gamma^c} - \Phi_{\Gamma^c}^T \Phi_\Gamma \tilde{u}_\Gamma\|_2^2 \\ &\geq \|\Phi_\Gamma^T \Phi_\Gamma \tilde{u}_\Gamma\|_2^2 \end{aligned}$$

Exploiting Assumption 1, one can conclude that the Gramm matrix  $\Phi_\Gamma^T \Phi_\Gamma$  is not singular, thus  $\|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_2^2 > 0$  as long as  $\|\tilde{u}\|_2^2 > 0$ . And furthermore, there exists a small value  $\kappa > 0$  such that

$$\kappa \|\tilde{u}\|_2^2 \leq \|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_2^2$$

□

Now consider the dynamical system (7) with soft thresholding function (6), then Theorem 2 can be proved as follows.

*Proof of Theorem 2.* According to Lemmas 3, 4 and 5, the following result is straightforward, for  $t > t_e$ ,

$$\begin{aligned} \dot{E}(\tilde{u}) &= -\|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_{1+\alpha}^{1+\alpha} \\ &\leq -\|\tilde{u} + (\Phi^T \Phi - I)\tilde{a}\|_2^{1+\alpha} \\ &\leq -\kappa^{\frac{1+\alpha}{2}} \|\tilde{u}\|_2^{1+\alpha} \\ &\leq -(\kappa/\nu)^{\frac{1+\alpha}{2}} (E(\tilde{u}))^{\frac{1+\alpha}{2}} \end{aligned} \quad (17)$$

where the first inequality is due to the fact that  $\|x\|_{1+\alpha} \geq \|x\|_2$  as  $\alpha \in (0, 1]$ .

Then  $\forall t > t_e$   $E(\tilde{u})$  converges to zero in finite time denoted  $t_f > t_e$ . Finally, we have  $\forall t > t_f, u = u^*$  and this ends the proof. □

#### D. Convergence Time

According to (17), one can conclude that the trajectory of Lyapunov function is upper bounded,

$$E(\tilde{u}) \leq \left( E_0^{\frac{1-\alpha}{2}} - \frac{1-\alpha}{2} \theta^{\frac{1+\alpha}{2}} t \right)^{\frac{2}{1-\alpha}}, \quad t \leq t_f(E_0) \quad (18)$$

when  $t > t_f(E_0)$ , we have  $E(\tilde{u}) = 0$ .

Then it is not so difficult to analyze the convergence time. Particularly, according to Theorem 4.2 in [42], the settling time function  $t_f$  can be explicitly conducted by exploiting (17) (partial integration with respect to  $E$  and  $t$  on both sides),

$$t_f(E_0) = \frac{2}{\theta^{\frac{1+\alpha}{2}} (1-\alpha)} E_0^{\frac{1-\alpha}{2}} \quad (19)$$

with  $E_0 = E(\tilde{u}(0))$  being the initial condition of Lyapunov function and  $\theta = \kappa/\nu$ .

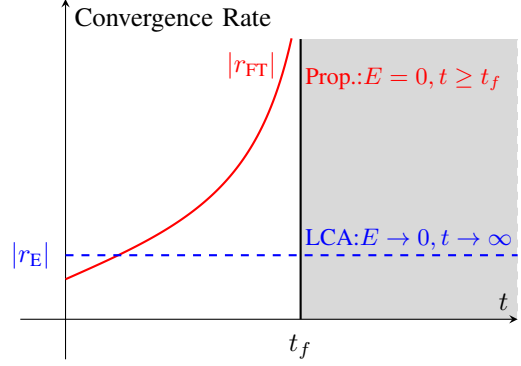


Fig. 1. The schematic diagram of convergence rate. The dashed line represents the convergence rate of LCA, i.e.  $|r_E|$ ; The solid curve represents the convergence rate of the proposed system, i.e.  $|r_{FT}|$ ; the dark shadowed area represents the equilibrium region of the proposed system, where  $E = 0$ .

It means that when  $t \geq t_f(E_0)$ , the Lyapunov function  $E(\tilde{u})$  exactly equals to 0, i.e. (7) is stable, as shown in Fig. 1. Note that the settling time is dependent to the initial value  $E_0$  and moreover, when  $\alpha \rightarrow 1$  the settling function  $t_f \rightarrow +\infty$ , which corresponds to the asymptotic convergence property. In the situation when parameter  $\alpha \in (0, 1)$ , we have to consider two different cases:

- when  $0 < E_0 \leq \exp(2)/\theta$ , the settling function  $t_f$  is monotone increasing with respect to  $\alpha$ ,
- when  $E_0 > \exp(2)/\theta$ , the settling function  $t_f$  has a minimum value at  $\alpha = 1 - \frac{2}{\ln(\theta E_0)}$

Consequently, when the state is close to the equilibrium point, smaller  $\alpha$  will lead to faster convergence.

On the other hand, regarding to the equation (15), the settling time is also dependent on the settings of the sparse recovery problem, i.e.  $(s, M, N)$ , which determine the RIP constant  $\delta_s$  and the signal to sparsity rate  $\rho$ . Apparently, the larger the number of measurements the smaller the RIP constant  $\delta_s$  which leads to the smaller settling time  $t_f$ . While the larger signal to sparsity rate  $\rho$  will result in a larger settling time  $t_f$ .

#### E. Convergence Rate

In this subsection, we will compare the convergence rate between finite-time and exponential convergence. In order to analyze the convergence property in counterpart of the exponential convergence rate, the logarithmic form of (18) is analyzed, i.e.

$$E(\tilde{u}) \leq e^{\frac{2}{1-\alpha} \log \left( E_0^{\frac{1-\alpha}{2}} - \frac{1-\alpha}{2} \theta^{\frac{1+\alpha}{2}} t \right)}$$

Then the convergence speed can be evaluated via the slope of the exponents with respect to time  $t$ , i.e.

$$r_{FT}(t, \alpha) = -\frac{1}{c_0 c_1^\alpha + (\alpha - 1) \frac{t}{2}}$$

with  $c_0 = \sqrt{\frac{E_0}{\theta}}$  and  $c_1 = \frac{1}{\sqrt{E_0 \theta}}$ . While the convergence speed of the corresponding exponential convergence rate can be direct obtained by setting  $\alpha = 1$ , then one can get

$$r_E = -\theta$$

Considering the convergence rate, apparently,  $r_{\text{FT}}$  is time varying, as shown by the red solid curve in Fig. 1. And moreover, when  $t \geq \frac{2}{1-\alpha} \left( \frac{E_0^{\frac{1-\alpha}{2}}}{\theta^{\frac{1-\alpha}{2}}} - \theta^{-1} \right)$ , we have  $r_{\text{FT}} \leq r_{\text{E}}$ , i.e.  $|r_{\text{FT}}| \geq |r_{\text{E}}|$ , namely, the proposed system (7) converges faster than LCA system.

Moreover, as the evolution time  $t$  approaching to the settling time  $t_f$ , the denominator of  $r_{\text{FT}}$  will go to zero and then leads to infinite value of  $|r_{\text{FT}}|$ , i.e. system (7) will converge super fast to the equilibrium point, as shown in Fig. 1. In this case, the proposed system (7) is more appropriate to solve the dynamic sparse signals, where the consecutive data are close enough such that the initial  $E_0$  is sufficiently small, which makes the settling time  $t_f$  small enough to guarantee the real time sparse recovery.

On the other hand, the convergence rate is also related to the parameter  $\alpha$ . Thus the influence of  $\alpha$  can be explicitly analyzed in the following cases. When  $c_1 > 1$ , increasing  $\alpha$  will decrease the convergence rate. When  $0 < c_1 < 1$ , it can be divided into two case,

- when  $t > 2c_0 \log(c_1)c_1^\alpha$ , increasing  $\alpha$  will decrease the convergence rate;
- when  $t < 2c_0 \log(c_1)c_1^\alpha$ , increasing  $\alpha$  will increase the convergence rate;

#### IV. DISCUSSIONS

The proposed model in this paper is an extension to LCA proposed in [30], where the ODE of the dynamic system of LCA is essentially the same form as the well-known continuous Hopfield neural network (HNN) [43] and Lyapunov functions [38] plays a very important role in convergence analysis. However, the difference between LCA and HNN is also very essential. In particular, active function is continuous and smooth for HNN, however, it is not necessarily to be smooth for LCA and our proposed system. On the other hand, the previous researches have rarely been focused on the finite-time stabilities of the networks for autonomous systems (LCA is with exponential stability). In this paper, we modified the ODE of the LCA system to introduce the sliding mode technique, and proposed a completely different Lyapunov function (9) to implicitly prove our results.

Similarly to the seminal work of LCA, the proposed method in this paper, possesses of solving the sparse representation via the dynamical system composed of many neuron-like elements operating in parallel analog architectures [30]. It is worth to remark that comparing to the computational oriented algorithms, the computational complexity of the proposed method is actually not reduced. Alternatively, the complexity of the proposed method (as well as LCA) is transferred to the implementation of analog architectures realized by analog chips. While the algorithm is very efficient as long as the analog architectures are implemented, e.g., matrix multiplication result can be obtained in real time, the computer-oriented algorithms require tens or hundreds of operations to get the result. Consequently, LCA-like approaches would be more appropriate to real time applications. On the other hand, our proposed system is with finite-time convergence, instead, LCA

is with exponential convergence, consequently, our proposed system can cope with signals with higher varying speed than LCA, which can be illustrated by Example 1.

Comparing to LCA, the complexity to implement the analog architectures of our proposed dynamical system will be slightly increased due to the fractional exponent and sign function. In fact, those terms can be easily realized even by using simple operational amplifiers, with which the basic functionalities such as multiplication, division,  $\log$ ,  $\exp$ ,  $\text{abs}$  already exist. For example, the fractional exponential operator (such as  $x^\alpha$  with  $0 < \alpha < 1$ ) can be realized via cascading a logarithm operator and an exponential operator ( $x^\alpha = \exp(\alpha \ln x)$ ) [44].

On the other hand, besides the soft-thresholding activation function, other type of active functions introduced in [35] can also be exploited in the proposed system. And the analysis for alternatives can be addressed by analogy, where one only has to reformulate the Lemma 4 according to Appendix of [35] and the relationship between  $\tilde{u}$  and  $\tilde{a}$  used in Lemma 5 can also be derived.

#### V. SIMULATIONS

In this section, we present several simulations to illustrate the theoretical results presented in this paper. Simulations will be carried out in four aspects. At first, the global convergence property of the proposed system is illustrated. Afterwards, we will analyze the number of switches before the convergence for the proposed system. Then, the property of finite time convergence is addressed. At last, the effects of  $\alpha$  on the convergence rate is also analyzed.

In the following, we will respectively exploit the proposed dynamical system and LCA to solve the canonical sparse representation problems. Without special explanation, the simulations are carried out with the following setting. The original sparse signals  $x \in \mathbb{R}^N$  with  $N = 200$  and sparsity  $s = 10$  are randomly generated, of which the nonzero entries are drawn from a normal Gaussian distribution. Afterwards, measurements  $y \in \mathbb{R}^M$  with  $M = 100$  are collected via random projections,  $y = \Phi x + \varepsilon$ , where measurement matrix  $\Phi \in \mathbb{R}^{M \times N}$  is drawn from a normal Gaussian distribution ( $\Phi$  is normalized to make every column with unit norm) and  $\varepsilon$  is the Gaussian random noise with standard derivation  $\sigma = 0.016$ . Dynamical equations of LCA and our proposed system are simulated through a discrete approximation in Matlab with a step size of 0.001 and a solver time constant is chosen to be equal to  $\tau = 0.1$ . The initial states is also set to  $u(0) = 0$  and the threshold value  $\lambda = 0.05$  for both systems.

##### A. Global Convergence

In this subsection, global convergence property of our proposed system is evaluated. Theorem 3 states that the proposed should converge and recover the solution to the sparse representation problem (1), which has a unique minimizer. As shown in Fig. 2, we plot the output  $a^*$  of our proposed dynamical system (7) after convergence. The comparison is made to LCA with same initial condition. And it is shown that our proposed system can reach the same sparse solution



as LCA, with 10 nonzero entries, which correspond to the nonzero entries in original sparse signal  $x$ .

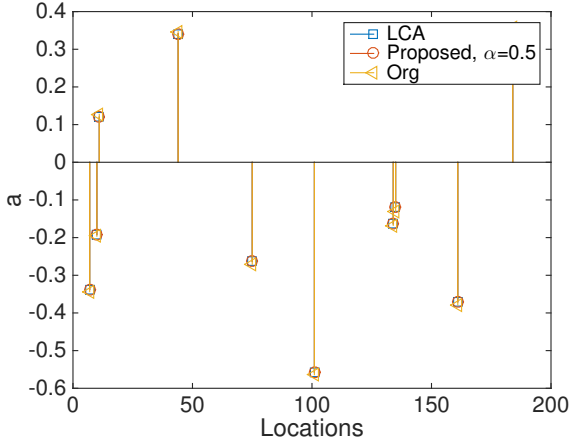


Fig. 2. Output  $a^*$  of LCA and the proposed system after convergence with  $\alpha = 0.5$  and  $\lambda = 0.05$ .

On the other hand, we also plot the evolution of several active nodes and nonactive nodes with respect to time for LCA and our proposed dynamical system in Fig. 3. The initial starting points of states  $u(t)$  for both systems are identical. It is shown that every node of both LCA and our proposed system converge to a fixed point and the convergent points for each node of LCA and our proposed system are identical, while the node of our proposed system converges much faster than LCA.

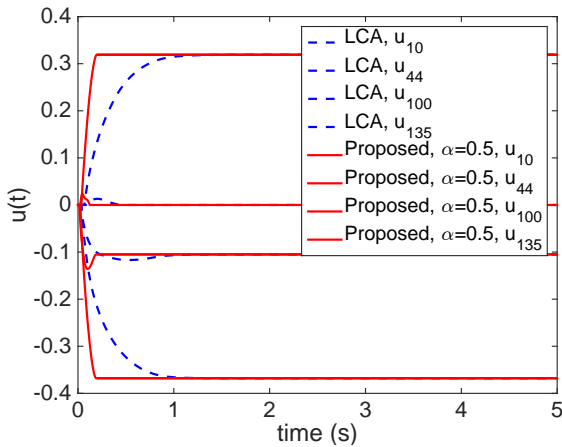


Fig. 3. Evolution of several active nodes (solid lines) and nonactive nodes (dashed lines) with respect to time for LCA and our proposed dynamical system with  $\alpha = 0.5$ .

At last, we evaluate the global convergence property of our proposed system by plotting the trajectories of two randomly selected nodes  $u_{10}$  and  $u_{44}$  starting from 20 randomly generated initial points. And the result is plotted in Fig. 4, from which one can clearly find that the solution is attractive for any of those initial points.

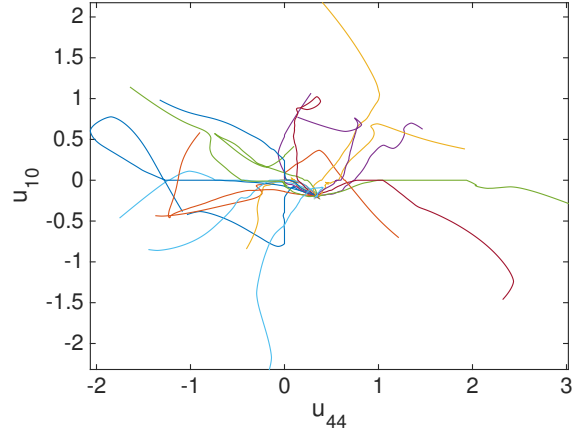


Fig. 4. Trajectories  $u_{44}(t)$  v.s.  $u_{10}(t)$  with 20 different initial conditions via the proposed system with  $\alpha = 0.5$ .

### B. Finite Switches

In this subsection, we will empirically verify the result of Lemma 2. The switch occurs as  $|u_i(t)| > \lambda$  decreasing to  $|u_i(t)| \leq \lambda$  or  $|u_i(t)| \leq \lambda$  increasing to  $|u_i(t)| > \lambda$ . In our simulation, the ODE (7) is simulated through a discrete approximation via `ode4` with step size 0.001 and 5 seconds evolution is implemented to guarantee the convergence, thus the solution trajectory is discretized into 5000 points. Then, 1000 trials are carried out with randomly generated initial conditions and noises, then the number of occurrences (for each trial) of switches is counted along the trajectories of all the nodes over these 5000 discrete points. At last, we can plot the histogram of the number of occurrence of switches, as shown in Fig. 5(b). This figure illustrates that the number of switches required for our proposed system before convergence is finite.

Moreover, we also plot the histogram of number of switches for LCA as the comparison, as shown in Fig. 5(a). Similarly, the number of switches required for LCA is also finite. Furthermore, the average number of switches required for LCA is less than that required for our proposed system. Even though, as shown in Fig. 6 where evolutions of the number of active nodes for LCA and our proposed system are plotted, it is clear that the number of active nodes converges faster for our proposed system than LCA. It implies that, although more switches occurred for proposed system, the interval between two contiguous switches is much smaller than that for LCA.

### C. Convergence in Finite Time

According to Theorem 2, after some time  $t_e > 0$ , the proposed system will converge in finite time. As shown in Fig. 7, the evolutions of state error  $\tilde{u}(t)$  and the number of the active nodes with respect to time are put together, where initial state point  $u(0)$  is generated randomly. Instead of exponential convergence rate as LCA (which has been proved in [35]), the proposed system converges largely faster than LCA, and the evolution of state error exhibits a finite-time convergence. On the other hand, the proposed system can find the correct active nodes faster than LCA.

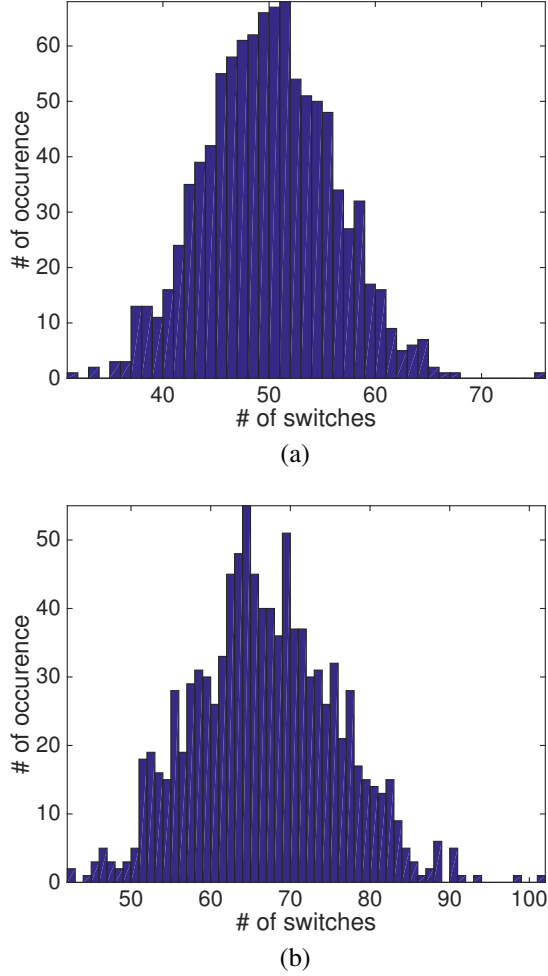


Fig. 5. Histogram of the number of switches required for LCA (a) and the proposed system (b) with  $\alpha = 0.5$  before convergence over 1000 trials.

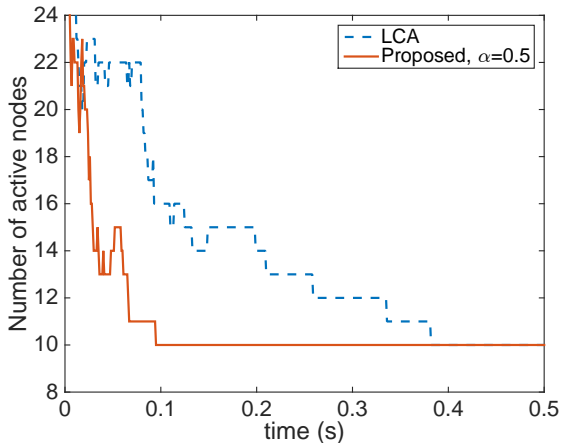


Fig. 6. Number of active nodes for the proposed system with  $\alpha = 0.5$ .

In order to verify Theorem 2, simulations with different settings are carried out, as shown in Fig. 8. We firstly fix the sparsity level  $s = 10$ , measurement number  $M = 100$  and the threshold  $\lambda = 0.05$ , then implement the simulation

with various signal length  $N \in [200, 400, 600, 800]$ . The evolutions of state error  $\tilde{u}(t) = u(t) - u^*$  for both LCA and the proposed system are plotted in Fig. 8(a). Similarly, the convergence performances comparing to LCA with respect to sparsity level  $s$ , measurement number  $M$  and the threshold  $\lambda$  are respectively considered, as shown in Fig. 8 from (b) to (d). It is obvious that the proposed system converges much faster than LCA with different signal length, measurement number, sparsity level and threshold, and performs the finite-time convergent property.

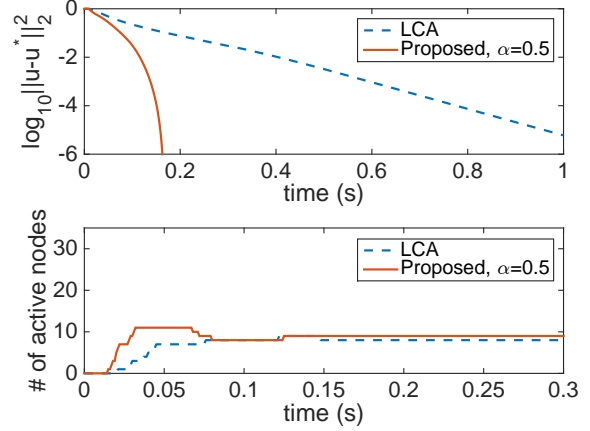


Fig. 7. Evolutions of state error  $\tilde{u}(t)$  and the number of active nodes with respect to time.

#### D. Influence of $\alpha$

In this subsection, the performance with respect to  $\alpha$  is analyzed, where simulations are carried out by ranging  $\alpha$  from 0.2 to 1 (when  $\alpha = 1$  it is equivalent to LCA) and let other parameters be fixed. The results are shown in Fig. 9, and one can find that the convergence rate is decreasing as  $\alpha$  increasing, which verifies the result in the proof of Theorem 2.

On the other hand, it is worth mentioning that simulations of dynamical system might induce oscillations when parameter  $\alpha$  is getting smaller. For instance, in Fig. 9 (the left and middle subfigure), oscillation happens when ODE is realized by approximating with low order ODE solvers, such as `ode1` with fixed time step  $10^{-3}$ . This phenomenon is due to the fact that the function  $[\cdot]^\alpha$  with  $\alpha < 1$  will result in some numerical computational problems when the variables are getting close to zero. In numerical simulations, it can be alleviated by either reducing the time step for ODE solvers or alternating to use higher order ODE solvers. As shown in Fig. 9, one can find that the oscillations disappeared when reducing the time step from  $10^{-3}$  to  $10^{-4}$  or replacing `ode1` solver by `ode4` solver, i.e., Runge-Kutta method.

## VI. EXTENSION TO TIME-VARYING PROBLEMS

In previous sections, it has been proved that the proposed dynamical system (7) has the finite-time convergent property, and empirically, it converge much faster than LCA. This

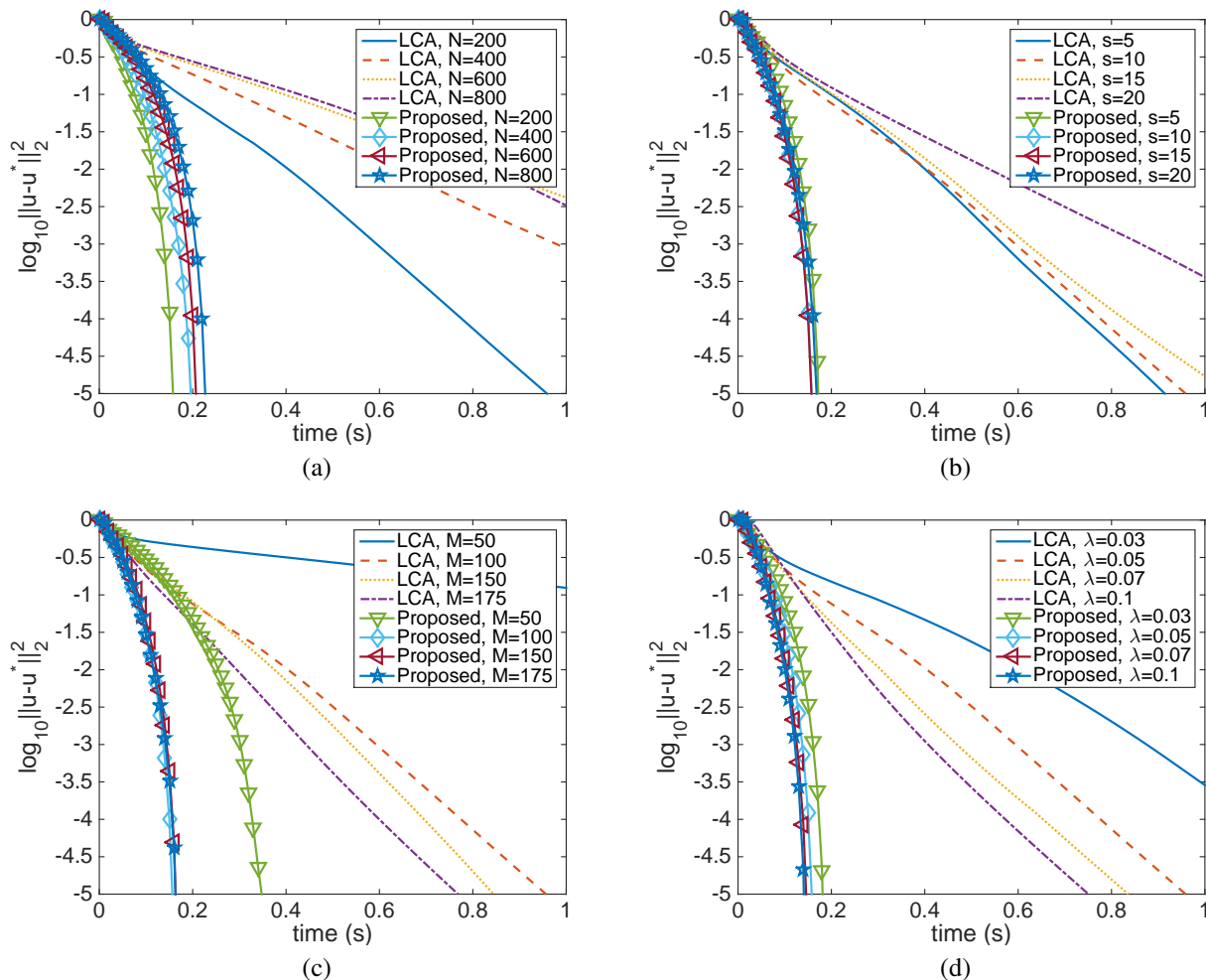


Fig. 8. Evolutions of  $\log_{10} \|\tilde{u}(t)\|_2^2$  for LCA (dotted lines) and the proposed dynamical system (7) (solid lines) with  $\alpha = 0.5$  as problem settings are varied with respect to (a) the signal length  $N$ , (b) the sparsity level  $s$ , (c) the measurement number  $M$  and (d) the threshold  $\lambda$ .

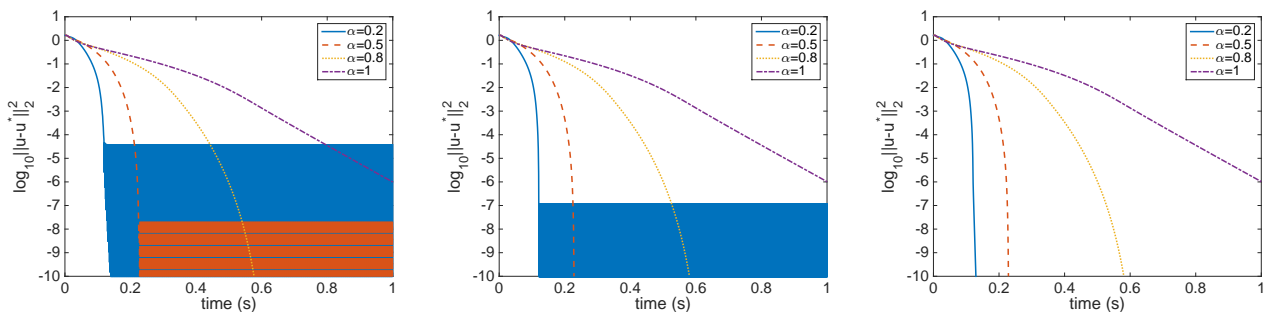


Fig. 9. Convergence of  $\log_{10} \|u - u^*\|_2^2$  for the proposed dynamical system with different value of  $\alpha \in [0.2, 0.5, 0.8, 1]$ . Different ODE solvers are used (left) ode1 with time step  $1e-3$ , (middle) ode1 with time step  $1e-4$  and (right) ode4 with fixed time step  $1e-3$ .

property is more appalusive than LCA especially in real applications, where sparse signals encountered are time-varying, i.e.,

$$y(t) = \Phi x(t) + \epsilon(t) \quad (20)$$

with  $y$  and  $x$  being both varying with respect to time.

In order to approximate the time-varying sparse signals  $x(t)$ , in [32], a maximum sampling rate and a large gradient step

size are required for convergence. However, it is straightforward in our proposed system, where the only requirement is to plug the time-varying measurements  $y(t)$  into the system (7) without changing any parameters.

To demonstrate the superiority of our proposed system, a toy example is given here.

**Example 1.** A length  $N$  time-varying sparse signal  $x(t)$  is

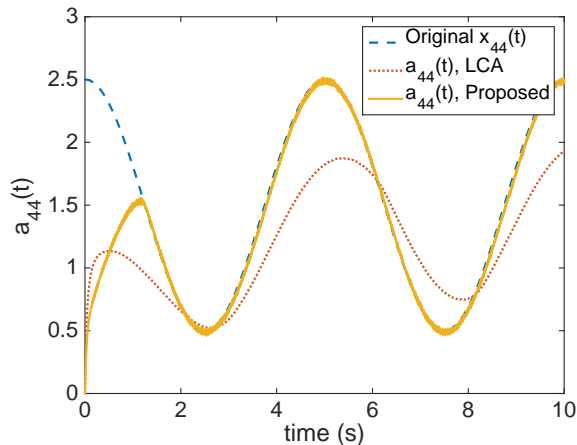


Fig. 10. Estimation of time-varying sparse signals via LCA and the proposed system.

generated with sparsity  $s = 5$ , where 4 of nonzero entries are drawn randomly and stay constant with respect to time. And the last nonzero entry is varying according to the following function

$$x_{44}(t) = \cos(0.4\pi t) + 1.5$$

Then measurements are gathered according to (20) with normal Gaussian noise with derivation  $\sigma = 0.016$ .

The estimations are obtained by evolving both LCA and our proposed system with  $\alpha = 0.5$  and threshold  $\lambda = 0.05$ , as shown in Fig. 10. Obviously, LCA cannot tracking the signal, while our proposed system can successfully tracking the changing of signal.

## VII. CONCLUSION

In this paper, we proposed a new dynamical system that can solve the sparse representations. It is with the finite-time convergence property. Comparing to LCA, the proposed system can converge to the same equilibrium point but with much faster convergence, which is very applaudable in real-time sparse representation applications.

Moreover, connections between continuous dynamical systems and discrete optimization algorithms for sparse regularized inversion problems have been investigated [45]. Meanwhile, it is also claimed in [32] that the iterative soft-thresholding algorithm can be considered as the discretized version to LCA. Thus, the future works would be focused on investigating the discretized version of our proposed dynamical system, which might result in a new sparse representation algorithm with faster convergent property.

## APPENDIX

**Lemma 6.** *If the matrix  $\Phi \in \mathbb{R}^{M \times N}$  satisfies the  $s$ -order RIP with constant  $\delta_s$ , then the eigenvalue of  $\Phi^T \Phi$  is upper bounded by  $N(1 + \delta_s)/s$ .*

*Proof.* Denote by  $S$  the all possible subset with size  $s$  of  $\{1, \dots, N\}$ , thus  $|S| = \binom{N}{s}$ . Then let  $x \in \mathbb{R}^N$  be an arbitrary

vector with norm 1 and denote by  $X_S$  the set of  $s$ -sparse vectors

$$X_S = \{y \in \mathbb{R}^N | y_i = 0, \forall i \notin \Gamma \text{ and } y_i = x_i, \forall i \in \Gamma, \Gamma \subset S\}$$

For each element of  $x$ , it will appear  $k = \binom{N-1}{s-1}$  times in all possible  $s$ -sparse vectors in  $X_S$ . Thus we have the following equation

$$kx = \sum_{y \in X_S} y$$

Then,

$$\begin{aligned} k\|\Phi x\|_2 &= \left\| \sum_{y \in X_S} \Phi y \right\|_2 \\ &\leq \sum_{y \in X_S} \|\Phi y\|_2 \leq \sqrt{|S|} \sqrt{\sum_{y \in X_S} \|\Phi y\|_2^2} \\ &\leq \sqrt{|S|} \sqrt{\sum_{y \in X_S} (1 + \delta_s) \|y\|_2^2} = \sqrt{k|S|(1 + \delta_s)} \end{aligned}$$

Thus,  $\|\Phi x\|_2^2 \leq N(1 + \delta_s)/s$ .  $\square$

## REFERENCES

- [1] E. Candès and M. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, mar 2008.
- [2] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, nov 2006.
- [4] J. Mairal, M. Elad, and G. Sapiro, "Sparse Representation for Color Image Restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.
- [5] X. Xu, X. Wei, and Z. Ye, "DOA Estimation based on Sparse Signal Recovery Utilizing Weighted-norm Penalty," *IEEE Signal Processing Letters*, vol. 19, no. 3, pp. 155–158, 2012.
- [6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [7] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Coatrieux, and W. Chen, "Artifact Suppressed Dictionary Learning for Low-dose CT Image Processing," *IEEE Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2271–2292, 2014.
- [8] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [9] M. Tan, I. W. Tsang, and L. Wang, "Matching pursuit lasso part ii: Applications and sparse recovery over batch signals," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 742–753, Feb 2015.
- [10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [11] E. Candès and T. Tao, "The dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ ," *The Annals of Statistics*, pp. 2313–2351, 2007.
- [12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [13] I. Daubechies, M. Defrise, and C. De Mol, "An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [14] E. J. Candès and T. Tao, "Decoding by Linear Programming," *IEEE Transaction on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [15] D. P. Wipf, B. D. Rao, D. P. Wipf, and D. D. Rao, "Sparse Bayesian Learning for Basis Selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

- [16] D. Wipf and S. Nagarajan, "Iterative Reweighted  $\ell_1$  and  $\ell_2$  Methods for Finding Sparse Solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, 2010.
- [17] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian Compressive Sensing for Cluster Structured Sparse Signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.
- [18] L. Yu, H. Sun, G. Zheng, and J. Pierre Barbot, "Model based Bayesian compressive sensing via Local Beta Process," *Signal Processing*, vol. 108, no. 3, pp. 259–271, 2015.
- [19] J. Tropp, "Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [20] J. A. Tropp and A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [21] D. Needell and J. a. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301–321, 2009.
- [22] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1370–1382, March 2017.
- [23] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6215–6221, 2011.
- [24] R. Baraniuk and P. Steeghs, "Compressive Radar Imaging," in *2007 IEEE Radar Conference*. IEEE, apr 2007, pp. 128–133.
- [25] T. Hu and D. B. Chklovskii, "Sparse LMS via online linearized bregman iteration," in *ICASSP 2014*. IEEE, 2014, pp. 7213–7217.
- [26] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *ICASSP 2009*. IEEE, 2009, pp. 3125–3128.
- [27] D. Angelosante, J. Bazerque, and G. Giannakis, "Online adaptive estimation of sparse signals: Where rls meets the  $\ell_1$  -norm," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3436–3447, July 2010.
- [28] K. Themelis, A. Rontogiannis, and K. Koutroumbas, "A variational Bayes framework for sparse adaptive estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4723–4736, Sept 2014.
- [29] L. Yu, C. Wei, and G. Zheng, "Adaptive Bayesian Estimation with Cluster Structured Sparsity," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2309–2313, 2015.
- [30] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. Olshausen, "Sparse coding via thresholding and local competition in neural circuits," *Neural computation*, vol. 20, pp. 2526–2563, 2008.
- [31] A. Balavoine, J. Romberg, and C. J. Rozell, "Convergence and Rate Analysis of Neural Networks for Sparse Approximation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 9, pp. 1377–1389, sep 2012.
- [32] A. Balavoine, C. J. Rozell, and J. Romberg, "Discrete and continuous-time soft-thresholding for dynamic signal recovery," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3165–3176, 2015.
- [33] L. S. Pontryagin, "The mathematical theory of optimal processes," *Classics of Soviet Mathematics, ISBN-13: 978-2881240775*, 1962.
- [34] R. Bellman and R. Kalaba, *Dynamic programming and modern control theory*, ser. Academic paperbacks. Academic Press, 1965.
- [35] A. Balavoine, C. J. Rozell, and J. Romberg, "Convergence speed of a dynamical system for sparse recovery," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4259–4269, 2013.
- [36] A. F. Fillipov, *Differential equations with discontinuous right-hand side*. Kluwer Academic Publishers, collection : Mathematics and its Applications, 1988.
- [37] J. LaSalle, "An invariance principle in the theory of stability," in *Differential Equations and Dynamical Systems Stability and Control*, Academic Press, pp. 277–286, 1967.
- [38] H. Khalil, "Nonlinear systems," *Prentice Hall, Upper Saddle River, NJ 07458*, 1996.
- [39] F. H. Clarke, Y. S. Ledyev, R. J. Stern, and P. R. Wolenski, "Nonsmooth analysis and control theory," *Graduate Texts in Mathematics, Springer, ISBN-13: 978-0387983363*, no. 7, 1998.
- [40] A. Levant, "Sliding order and sliding accuracy in sliding mode control," *Int. J. of control*, vol. 58, no. 6, pp. 1247–1253, 1993.
- [41] L. Yu, J.-P. Barbot, D. Boutat, and D. Benmerzouk, "Observability Forms for Switched Systems With Zeno Phenomenon or High Switching Frequency," *IEEE T. Automat. Contr.*, vol. 56, no. 2, pp. 436–441, 2011.
- [42] S. P. Bhat and D. S. Bernstein, "Finite-time stability of continuous autonomous systems," *SIAM Journal on Control and Optimization*, vol. 38, no. 3, pp. 751–766, 2000.
- [43] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the national academy of sciences*, vol. 81, no. 10, pp. 3088–3092, 1984.
- [44] G. W. Roberts and V. W. Leung, *Design and analysis of integrator-based log-domain filter circuits*. Springer Science & Business Media, 2000.
- [45] 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*, pp. 2510–2518, 2014.