

Discontinuous Neural Networks for Finite-Time Solution of Time-Dependent Linear Equations

Mauro Di Marco, Mauro Forti, Paolo Nistri, and Luca Pancioni

Abstract—This paper considers a class of nonsmooth neural networks with discontinuous hard-limiter (signum) neuron activations for solving time-dependent (TD) systems of algebraic linear equations (ALEs). The networks are defined by the subdifferential with respect to the state variables of an energy function given by the L_1 norm of the error between the state and the TD-ALE solution. It is shown that when the penalty parameter exceeds a quantitatively estimated threshold the networks are able to reach in finite time, and exactly track thereafter, the target solution of the TD-ALE. Furthermore, this paper discusses the tightness of the estimated threshold and also points out key differences in the role played by this threshold with respect to networks for solving time-invariant ALEs. It is also shown that these convergence results are robust with respect to small perturbations of the neuron interconnection matrices. The dynamics of the proposed networks are rigorously studied by using tools from nonsmooth analysis, the concept of subdifferential of convex functions, and that of solutions in the sense of Filippov of dynamical systems with discontinuous nonlinearities.

Index Terms—Discontinuous neural networks, finite-time convergence, subdifferential, time-dependent (TD) linear equations.

I. INTRODUCTION

THERE are significant motivations for studying dynamic neural networks to solve linear algebra problems in real-time. Computing the solution of a system of algebraic linear equations (ALEs), or, equivalently, computing the inverse of a matrix, is indeed one of the most basic and frequently encountered problems in several diverse areas of engineering and physical sciences [1]–[3]. Quite often the ALEs involved are high-dimensional and it is important, if not mandatory, to obtain their solution in real-time, as in signal and image processing, electromagnetic problems, statistics and robotics, to name only a few important application fields [4]–[6]. Another motivation is that understanding the neural network behavior in the solution of linear algebra problems can provide useful information also to design neural architectures for solving more general nonlinear optimization problems.

Manuscript received April 23, 2015; revised July 29, 2015; accepted September 14, 2015. This paper was recommended by Associate Editor Q. Liu.

The authors are with the Department of Information Engineering and Mathematical Sciences, University of Siena, Siena 53100, Italy (e-mail: dimarco@dii.unisi.it; forti@dii.unisi.it; pnistri@dii.unisi.it; pancioni@dii.unisi.it).

Digital Object Identifier 10.1109/TCYB.2015.2479118

In 1992, Cichocki and Unbehauen [7], [8] published two relevant papers on the analog neural network approach for solving online ALEs. The approach is based on constructing a computational energy (Lyapunov) function whose minimum is the desired ALE solution, find the (negative) gradient of this function, and interpret the gradient equations by means of a dynamic neural network architecture. Emphasis is given to the hardware implementation of the obtained networks in CMOS technology or techniques based on switched capacitors. It has been shown analytically, or by simulations, that the networks in [7] and [8] work well in the solution of large classes of ALEs. Since then, several contributions have been published on the dynamic neural network approach to solve ALEs and related linear algebra problems as computing the eigenvalues/eigenvectors, finding the lower-upper decomposition or Cholesky factorization, or finding the inverse or the Moore–Penrose pseudo-inverse of a matrix, and this topic is still of current interest (see [9]–[18] and references therein).

The computational energy introduced in [7] and [8] is given by some norm of the error between the neural network state and the ALE solution. Accordingly, we can classify the resulting networks in two main categories: 1) for smooth (differentiable) norms, as the least squares norm, the obtained neural networks are described by smooth gradient systems of differential equations involving C^1 neuron activations and 2) for nonsmooth norms, such as the L_1 (or the Chebishev) norm, the obtained neural networks are instead described by nonsmooth gradient systems of differential equations involving discontinuous hard-comparator (signum) neuron activations. Relative advantages and disadvantages of using the least squares, or the L_1 minimization criteria, are discussed in [8] and [19].

The networks in [7] and [8] are intrinsically conceived and designed by referring to time-invariant (TI) ALEs where all coefficients are held fixed during the networks evolution. However, in practice they are applied in a dynamic context where the ALEs usually have rapidly time-varying coefficients. This is the case for instance of several control applications, such as kinematic control of redundant manipulators, where the nullspace-type solution of the kinematic equation requires the computation of the Moore–Penrose generalized inverse of the time-varying Jacobian matrix of the same equation [14]. Analogous problems are generally encountered also in communications and signal processing applications, such as adaptive equalization for time-varying dispersive channels [20], while it is known that, in general, real-time matrix inversion is a key enabling technology in

multiple-input multiple-output communications systems [21], as well as in real-time reconstruction of a sequence of images of moving objects in electrical capacitance tomography [22]. We also mention that control of nonlinear time-varying systems via instantaneous pole placement requires the solution of a time-varying linear matrix equation (Sylvester equation) in order to stabilize in real time the system linearization [16].

Then, it is important to ask whether the behavior of the networks in [7] and [8], which is satisfactory in the solution of TI-ALEs, continues to be so also in the solution of time-dependent (TD) ALEs. To the authors' knowledge, so far this relevant question has been discussed and addressed only in the smooth case, where it has been shown to have a negative answer. More precisely, Zhang *et al.* [23] has shown that neural networks based on smooth gradient systems in general display nonvanishing errors in finding the inverse of TD matrices. The estimates there obtained show that even in the inversion of simple TD matrices the errors may be quantitatively significant and do not approach 0 as $t \rightarrow +\infty$.

The goal of this paper is to show that the previous question has instead a positive answer in the nonsmooth case. Namely, nonsmooth gradient neural networks with a sufficiently large penalty parameter are well suited also to work in a TD environment and solve TD-ALEs. More specifically, we consider a class of nonsmooth neural networks defined by the subdifferential with respect to the state variables of a nonsmooth energy (barrier) function given by the L_1 norm of the difference between the state and the TD-ALE solution. The networks, which are the natural extension to the TD case of those in [7] and [8], have discontinuous neuron activation functions, and are rigorously analyzed by means of tools from nonsmooth analysis and Lyapunov method. First of all, this paper addresses the definition, existence, and uniqueness of the solutions with respect to initial conditions. Then the main result is that, under reasonable assumptions on the TD coefficients and matrices involved, and when the penalty parameter exceeds a given threshold, each solution converges in finite time to the exact solution of the TD-ALE and exactly tracks the same solution for subsequent times. Then, this paper discusses how tight is the estimated threshold and also points out the key differences in the role played by the threshold with respect to that played in networks for solving TI-ALEs. This paper also provides a robustness analysis of finite-time convergence to the TD-ALE solution with respect to perturbations due to the electronic implementation. Various simulation examples and applications to TD-ALEs are discussed to illustrate the proposed neural network capabilities.

During recent years, techniques from nonsmooth analysis have been widely applied to study neural networks with high-gain nonlinearities (ideal diodes, hard-comparators, or hard-limiters), such as linear and nonlinear programming networks [24]–[28], networks for global optimization, synchronization and identification [29]–[39], and full-range cellular neural networks [40]. Some features and advantages of nonsmooth analysis can be briefly pointed out as follows. Nonsmooth analysis is based on the concept of Filippov solutions, which are good approximations of

solutions of actual dynamic systems with high-gain nonlinearities [41]. The needed mathematical machinery, which relies on subgradient calculus and nonsmooth Lyapunov method, is rigorous [40], [41]. Moreover, nonsmooth analysis is able to highlight salient features of the dynamics as the presence of sliding modes along discontinuity surfaces or the phenomenon of convergence in finite time [29]. Such phenomena would be instead difficult to study via a smooth analysis or by considering the high-gain limit behavior.

The structure of this paper is outlined as follows. In the remaining part of this section, we give the needed preliminary results. Then, Section II discusses the considered neural network model, whereas Section III gives the main results on the global dynamical behavior of the network. Section IV studies the dynamics of a perturbed model, while Section V presents some application examples. Then, Section VI discusses the obtained results in the context of the existing literature and, finally, Section VII collects some concluding remarks.

Notation: If $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ is the scalar product. Moreover, $\|x\|_2 = \sqrt{\langle x, x \rangle} = (\sum_{i=1}^n x_i^2)^{1/2}$ is the Euclidean norm of x . For any $r > 0$, $S(0, r) = \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$ is the closed ball with radius r about the origin. Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. The self-adjoint matrix is given by $A'A$ and is symmetric and positive semidefinite (the prime denotes the transpose). Let $\lambda_i(A'A)$, $i = 1, 2, \dots, n$, be the eigenvalues of $A'A$ and $\lambda_{\max}(A'A) = \max_i \lambda_i(A'A)$ (spectral radius of $A'A$), $\lambda_{\min}(A'A) = \min_i \lambda_i(A'A)$. The induced (spectral) norm of A is $\|A\|_2 = \sqrt{\lambda_{\max}(A'A)}$ and we have $\|Ax\|_2 \leq \|A\|_2 \|x\|_2$ for any $x \in \mathbb{R}^n$. If $\det A \neq 0$, $A'A$ is symmetric and positive definite and so $\lambda_{\min}(A'A) > 0$. In this case, the condition number of A is given by $\|A\|_2 \|A^{-1}\|_2 = (\lambda_{\max}(A'A)/\lambda_{\min}(A'A))^{1/2}$. Hence

$$\|A^{-1}\|_2 = \frac{1}{\sqrt{\lambda_{\min}(A'A)}}. \quad (1)$$

If $x \in \mathbb{R}^n$, we also let $\|x\|_1 = \sum_{i=1}^n |x_i|$. The corresponding induced matrix norm is $\|A\|_1 = \max_{j=1,2,\dots,n} \sum_{i=1}^n |a_{ij}|$ and we have $\|Ax\|_1 \leq \|A\|_1 \|x\|_1$ for any $x \in \mathbb{R}^n$. For any $x \in \mathbb{R}^n$ we have $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2$.

A. Preliminaries

Let $(t, x) \rightarrow F(t, x)$ be a multivalued map from $[0, +\infty) \times \mathbb{R}^m$ into the subsets of \mathbb{R}^m such that $F(t, x) \neq \emptyset$ for any $(t, x) \in [0, +\infty) \times \mathbb{R}^m$. The map F is said to be upper semicontinuous (u.s.c.) at point $(t, x) \in [0, +\infty) \times \mathbb{R}^m$ if for any open set U in \mathbb{R}^m , such that $F(t, x) \subset U$, there exists a neighborhood V in $[0, +\infty) \times \mathbb{R}^m$ of (t, x) such that $F(V) \subset U$ [41]. When F has closed values, and it is bounded in a neighborhood of each point $(t, x) \in [0, +\infty) \times \mathbb{R}^m$, F is u.s.c. on \mathbb{R}^m if and only if its graph $\{(t, x, y) \in ([0, +\infty) \times \mathbb{R}^m) \times \mathbb{R}^m : y \in F(t, x)\}$ is a closed subset.

Let $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be convex. A convex function is continuous and locally Lipschitz in \mathbb{R}^m but it is not in general differentiable. The subdifferential $\partial f(x)$ of f at $x \in \mathbb{R}^m$ is the set of vectors $z \in \mathbb{R}^m$ such that [42, Definition 1.2.1]

$$f(y) \geq f(x) + \langle z, y - x \rangle \quad \forall y \in \mathbb{R}^m$$

where any $z \in \partial f(x)$ is said to be a subgradient of f at x .

The multivalued map $x \rightarrow \partial f(x)$ has nonempty compact convex values, has a closed graph and is locally bounded [42, Sec. VI-6.2], hence $x \rightarrow \partial f(x)$ sends compact sets of \mathbb{R}^m into compact sets of \mathbb{R}^m . Furthermore, $x \rightarrow \partial f(x)$ is maximal monotone [41, Proposition 1, p. 159], i.e., we have $\langle v_a - v_b, x_a - x_b \rangle \geq 0$ for any $x_a, x_b \in \mathbb{R}^m$ and any $v_a \in \partial f(x_a), v_b \in \partial f(x_b)$. The subdifferential is a generalization of the ordinary concept of differentiation, indeed if f is differentiable at x in the ordinary sense, then $\partial f(x) = \{\nabla f(x)\}$, where $\nabla f(x)$ is the gradient of f at x .

II. NEURAL NETWORK MODEL

Consider the TD-ALE

$$A(t)z(t) - b(t) = 0 \quad (2)$$

where $z(t) \in \mathbb{R}^n$ and functions $t \rightarrow A(t) = (a_{ij}(t)) \in \mathbb{R}^{n \times n}$, $t \rightarrow b(t) = (b_i(t)) \in \mathbb{R}^n$ are globally Lipschitz continuous for $t \geq 0$, i.e., there exist $L_A, L_b \geq 0$ such that $\|A(t) - A(s)\|_2 \leq L_A|t - s|$, $\|b(t) - b(s)\|_2 \leq L_b|t - s|$ for $t, s \geq 0$.

We suppose henceforth that $\det A(t) \neq 0$ for any $t \geq 0$ so that (2) has a unique solution given by $z^*(t) = A^{-1}(t)b(t)$ for $t \geq 0$. Our goal is to design a neural network for which any solution $x(t)$ is such that there exists a finite instant t_h such that we have $x(t_h) = z^*(t_h)$ and $x(t) = z^*(t)$ for any $t \geq t_h$. Said another way, the network should be able to reach the target solution of TD-ALE (2) in finite time and exactly track the same solution for subsequent times. To this end, we need to enforce some additional assumptions. Since $\det A(t) \neq 0$ for $t \geq 0$, the self-adjoint matrix $A'(t)A(t)$ is symmetric and positive definite, i.e., $\lambda_{\min}(A'(t)A(t)) > 0$ for any $t \geq 0$.

Assumption 1: There exists a constant $\mu > 0$ such that $\lambda_{\min}(A'(t)A(t)) \geq \mu > 0$ for any $t \geq 0$.

Assumption 2: There exists $\bar{b} > 0$ such that $\|b(t)\|_2 \leq \bar{b}$ for any $t \geq 0$.

Proposition 1: Suppose that Assumptions 1 and 2 hold. Then

$$\|z^*(t)\|_2 \leq \frac{1}{\sqrt{\mu}}\bar{b} \quad (3)$$

for any $t \geq 0$, i.e., function $t \rightarrow z^*(t)$ is bounded. Moreover, it is Lipschitz continuous, i.e., there exists $L_* \geq 0$ such that $\|z^*(t) - z^*(s)\|_2 \leq L_*|t - s|$ for any $t, s \geq 0$, where

$$L_* = \frac{1}{\sqrt{\mu}}L_b + \frac{1}{\mu}L_A\bar{b}. \quad (4)$$

Proof: We have $\|z^*(t)\|_2 = \|A^{-1}(t)b(t)\|_2 \leq \|A^{-1}(t)\|_2\|b(t)\|_2$. By Assumption 1 and (1), we have

$$\|A^{-1}(t)\|_2 = \frac{1}{\sqrt{\lambda_{\min}(A'(t)A(t))}} \leq \frac{1}{\sqrt{\mu}}$$

whereas, by Assumption 2, $\|b(t)\|_2 \leq \bar{b}$ for any $t \geq 0$. Hence (3) is obtained.

For any $t, s \geq 0$, $\|A^{-1}(t) - A^{-1}(s)\|_2 = \|A^{-1}(t)[A(s) - A(t)]A^{-1}(s)\|_2 \leq \|A^{-1}(t)\|_2\|A^{-1}(s)\|_2\|[A(s) - A(t)]\|_2 \leq L_A|t - s|/\mu$. Then

$$\begin{aligned} \|z^*(t) - z^*(s)\|_2 &\leq \|A^{-1}(t)b(t) - A^{-1}(s)b(s)\|_2 \\ &= \|A^{-1}(t)b(t) - A^{-1}(t)b(s) \\ &\quad + A^{-1}(t)b(s) - A^{-1}(s)b(s)\|_2 \\ &\leq \|A^{-1}(t)\|_2\|b(t) - b(s)\|_2 \\ &\quad + \|A^{-1}(t) - A^{-1}(s)\|_2\|b(s)\|_2 \\ &\leq \frac{1}{\sqrt{\mu}}L_b|t - s| + \|A^{-1}(t) - A^{-1}(s)\|_2\bar{b} \\ &\leq \left(\frac{L_b}{\sqrt{\mu}} + \frac{L_A\bar{b}}{\mu}\right)|t - s| \end{aligned}$$

for any $t, s \geq 0$, and the proof is complete. \blacksquare

Consider the nonsmooth TD barrier function

$$B(t, x) = \|A(t)x - b(t)\|_1 = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}(t)x_j - b_i(t) \right| \quad (5)$$

for $x \in \mathbb{R}^n$ and $t \geq 0$. Since for any fixed $t \geq 0$ function $x \rightarrow B(t, x)$ is convex, the subdifferential function $x \rightarrow \partial_x B(t, x)$ is well defined (see Section I-A). Then, we can introduce the nonsmooth neural network described by the differential system

$$\dot{x} \in -\sigma \partial_x B(t, x)$$

for almost all (a.a.) $t \geq 0$, where $\sigma > 0$ is a threshold which will be chosen at a later point. For any $t \geq 0$ the barrier function $x \rightarrow B(t, x)$ in (5) is the sum of the n convex functions $x \rightarrow |\sum_{j=1}^n a_{ij}(t)x_j - b_i(t)|$. Then, by an argument as in the proof of [24, Proposition 3], we obtain $\partial_x B(t, x) = \sum_{i=1}^n \partial_x |\sum_{j=1}^n a_{ij}(t)x_j - b_i(t)|$, and so $\partial_x B(t, x) = A'(t)\text{SGN}(A(t)x - b(t))$, where $\text{SGN}(y) = (\text{sgn}(y_1), \text{sgn}(y_2), \dots, \text{sgn}(y_n))'$ for and $y \in \mathbb{R}^n$ and

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

Therefore, the neural network equations can be rewritten as

$$\dot{x} \in -\sigma A'(t)\text{SGN}(A(t)x - b(t)) \quad (6)$$

for a.a. $t \geq 0$.

Note that the right-hand side of (6) is a multivalued map with nonempty compact convex values giving the whole set of possible velocities \dot{x} of the neural network at a given instant t and a given point x . In mathematical terms, the neural network is then described by a differential inclusion rather than an ordinary differential equation with a single-valued velocity vector field.

Let $x_0 \in \mathbb{R}^n$. We say that $x(\cdot)$ is a local solution of (6) with initial condition $x(0) = x_0$ if there exists $T > 0$ such that $x(\cdot)$ is an absolutely continuous function on $[0, T]$ and we have $\dot{x}(t) \in -\sigma A'(t)\text{SGN}(A(t)x(t) - b(t))$ for a.a. $t \in [0, T]$. We say that $x(\cdot)$ is a solution of (6) on $[0, +\infty)$ if $x(\cdot)$ is a solution on $[0, T]$ for any $T > 0$.

First, we give an *a priori* estimate for the solutions of (6).

Lemma 1: Suppose that $x_a(\cdot), x_b(\cdot)$ are two solutions of (6) defined in $[0, T]$, for some $T > 0$. Then, we have

$$\|x_a(t) - x_b(t)\|_2 \leq \|x_a(s) - x_b(s)\|_2, \quad 0 \leq s \leq t \leq T$$

i.e., the Euclidean norm of the difference between the solutions is nonincreasing.

Proof: For any $\sigma > 0$, and $t \in [0, T]$, $x \rightarrow \sigma B(t, x)$ is convex, hence $x \rightarrow \sigma \partial_x B(t, x)$ is a maximal monotone operator (Section I-A). Then, for a.a. $t \in [0, T]$ we obtain

$$\begin{aligned} \frac{d}{dt} \frac{1}{2} \|x_a(t) - x_b(t)\|_2^2 &= \langle x_a(t) - x_b(t), \dot{x}_a(t) - \dot{x}_b(t) \rangle \\ &= -\langle x_a(t) - x_b(t), \xi_a(t) - \xi_b(t) \rangle \end{aligned}$$

for some $\xi_a(t) \in -\sigma \partial_x B(t, x_a(t))$, $\xi_b(t) \in -\sigma \partial_x B(t, x_b(t))$. In conclusion, $d\|x_a(t) - x_b(t)\|_2^2/dt \leq 0$ for a.a. $t \in [0, T]$. Hence, $t \rightarrow \|x_a(t) - x_b(t)\|_2^2$ and $t \rightarrow \|x_a(t) - x_b(t)\|_2$ are nondecreasing in $[0, T]$. ■

Proposition 2: For any $x_0 \in \mathbb{R}^n$ there exists a unique solution $x(t, x_0)$ of (6) with initial condition $x(0) = x_0$, which is defined on $[0, +\infty)$.

Proof: Function $y \rightarrow \text{SGN}(y)$ is u.s.c. with nonempty compact convex values in \mathbb{R}^n and functions $(t, x) \rightarrow A(t)x - b(t)$ and $t \rightarrow A'(t)$ are continuous. Hence, for any $\sigma > 0$ the multivalued vector field $(t, x) \rightarrow \sigma \partial_x B(t, x)$ is u.s.c. with nonempty compact convex values. Furthermore, for any $(t, x) \in [0, +\infty) \times \mathbb{R}^n$ the multivalued map $(t, x) \rightarrow \text{SGN}(A(t)x - b(t))$ takes values in the hypercube $[-1, 1]^n$ and for any $\sigma > 0$ function $t \rightarrow \sigma A'(t)$ is bounded in $[0, T]$ for any fixed $T > 0$. Thus, by [41, Th. 3, p. 98] there exists at least a local solution of (6). Moreover, such local solution is easily seen to be bounded on any time interval $[0, T]$, hence it can be prolonged to $[0, +\infty)$. Finally, uniqueness of solution follows immediately from Lemma 1. ■

Proposition 3: The following hold for neural network (6).

- 1) Given $x_a, x_b \in \mathbb{R}^n$, the Euclidean distance between solutions $t \rightarrow \|x(t, x_a) - x(t, x_b)\|_2$ is a nonincreasing function for any $t \geq 0$.
- 2) The solutions of (6) are either all unbounded, or all bounded, on $[0, +\infty)$.

Proof: The proof is divided into two parts.

- 1) Follows from Lemma 1 and Proposition 2.
- 2) Suppose that for some $x_0 \in \mathbb{R}^n$ the solution $x(t, x_0)$ is bounded on $[0, +\infty)$, i.e., there exists $M > 0$ such that $\|x(t, x_0)\|_2 < M$ for any $t \geq 0$. Let $x_a \in \mathbb{R}^n$. We have $\|x(t, x_a)\|_2 \leq \|x(t, x_a) - x(t, x_0)\|_2 + \|x(t, x_0)\|_2$ and, by the result in point 1), $\|x(t, x_a)\|_2 \leq \|x_a - x_0\|_2 + M$ for any $t \geq 0$, i.e., $x(\cdot, x_a)$ is bounded on $[0, +\infty)$.

Conversely, suppose that for some $x_0 \in \mathbb{R}^n$ the solution $x(t, x_0)$ is unbounded on $[0, +\infty)$, i.e., there exists a sequence $t_k \rightarrow +\infty$ such $\|x(t_k, x_0)\|_2 \rightarrow +\infty$. Pick any $x_a \in \mathbb{R}^n$ and note that $\|x(t_k, x_a)\|_2 \geq \|\|x(t_k, x_0)\|_2 - \|x(t_k, x_0) - x(t_k, x_a)\|_2\|$. By the result in point 1), $\|x(t_k, x_a) - x(t_k, x_0)\|_2 \leq \|x_a - x_0\|_2$, and then $\|x(t_k, x_a)\|_2 \rightarrow +\infty$, i.e., $x(\cdot, x_a)$ is unbounded on $[0, +\infty)$. ■

Remarks: We conclude this section with a list of remarks concerning neural network (6).

- 1) We use for (6) the concept of solutions in the sense of Filippov. This kind of solutions are of importance in engineering applications since it is known that they are uniform approximations on compact intervals of solutions of actual nonlinear systems with high-gain nonlinearities [41].
- 2) The proposed network (6) is the natural extension to the case of TD coefficients $a_{ij}(\cdot)$ and $b_i(\cdot)$ of the networks for TI-ALEs, and based on the minimization of the L_1 norm of the error, proposed by Cichocki and Unbehauen [8, p. 620]. Those networks use discontinuous hard-comparator (signum) functions and if we use a typical convexification procedure for defining Filippov solutions it can be seen that the resulting differential inclusion coincides with that describing (6).
- 3) Neural network (6) can also be thought of as being a generalization to a TD setting of the neural network for nonsmooth nonlinear programming problems in [24], or an extension to a nonsmooth and TD setting of the neural network for smooth nonlinear programming problems in [43]. Network (6) has two sets of feedback interconnected amplifiers, namely, linear amplifiers whose outputs are the state variables x_i , $i = 1, 2, \dots, n$, and constraint amplifiers whose input–output activation is a discontinuous (signum) function (see the circuit schematic in [8, Fig. 1]).

The feedback interconnections are given by the two TD matrices $A(t)$ and $A'(t)$ for any $t \geq 0$. Each entry of these matrices, i.e., each TD synaptic weight, can be implemented in VLSI-CMOS technology either via a four-quadrant multiplier or a transconductance amplifier along the lines discussed for example in [8, Sec. V]. The TD inputs $b_i(\cdot)$ can be implemented by means of TD voltage sources as biasing inputs to the network.

III. GLOBAL DYNAMICAL ANALYSIS

After defining the neural network (6), the main question we have to address is whether we can find a threshold $\bar{\sigma}$ such that, when the penalty parameter σ exceeds $\bar{\sigma}$, any solution $x(\cdot, x_0)$ of (6) reaches the solution $z^*(\cdot)$ of the TD-ALE (2) in finite time and exactly tracks $z^*(\cdot)$ thereafter. A positive answer to this question is given in Section III-A. An important related issue is how to choose an effective (candidate) Lyapunov function for (6) in order to study the global dynamical behavior and the reaching and tracking phase. We will see in the same section that, although (6) is defined by the sub-differential with respect to x of $B(t, x) = \|A(t)x - b(t)\|_1$, there are advantages in using as a Lyapunov function for studying the global dynamical behavior the Euclidean distance between $x(t, x_0)$ and $z^*(t)$. In Sections III-B and III-C, we then discuss two additional relevant issues. The first one is how much conservative is the estimated threshold $\bar{\sigma}$. The second one concerns some fundamental differences in the role played by the penalty parameter σ , and the threshold $\bar{\sigma}$, in

the solution of TD-ALEs as those considered in this paper, with respect to TI-ALEs as considered in previous papers in the literature.

A. Reaching and Tracking Phase

Let us discuss how to choose the penalty parameter σ in neural network model (6). Since (6) is aimed at reaching and tracking the solution $z^*(\cdot)$ of the TD-ALE (2), first of all we should ensure that $z^*(\cdot)$ is a solution of (6). The next property says that this is indeed the case if σ exceeds a given threshold.

Proposition 4: Suppose that Assumptions 1 and 2 are satisfied and that

$$\sigma > \bar{\sigma} \doteq \frac{L_*}{\sqrt{\mu}} = \frac{1}{\mu}L_b + \frac{1}{\mu^{3/2}}L_A\bar{b}. \quad (7)$$

Then, $x(t, z^*(0)) = z^*(t)$ for any $t \geq 0$, i.e., $z^*(\cdot)$ is the unique solution of (6) with initial condition $z^*(0)$.

Proof: We need to show that, for a.a. $t \geq 0$, $\dot{z}^*(t) \in -\sigma A'(t)\text{SGN}(A(t)z^*(t) - b(t)) = -\sigma A'(t)\text{SGN}(0) = -\sigma A'(t)[-1, 1]^n$. Since $S(0, 1) \subset [-1, 1]^n$, in order to prove the previous inclusion it is enough to show that, for $\sigma > \bar{\sigma}$, we have $\|\dot{z}^*(t)\|_2 \leq \sigma \min_{\|w\|_2=1} \|A'(t)w\|_2$. Indeed, as it can be easily seen, this in turn implies that $\dot{z}^*(t) \in -\sigma A'(t)S(0, 1)$. Due to Proposition 1, $\|\dot{z}^*(t)\|_2 \leq L_*$. Moreover, $\|A'(t)w\|_2^2 = w'A(t)A'(t)w$ and so, by Assumption 1

$$\min_{\|w\|_2=1} \|A'(t)w\|_2 = \sqrt{\lambda_{\min}(A(t)A'(t))} \geq \sqrt{\mu}.$$

Hence, $\sigma \min_{\|w\|_2=1} \|A'(t)w\|_2 \geq \sigma\sqrt{\mu} > \bar{\sigma}\sqrt{\mu} = L_* \geq \|\dot{z}^*(t)\|_2$ for a.a. $t \geq 0$, thus completing the proof. ■

Taking into account Proposition 3, and that $z^*(\cdot)$ is bounded (Proposition 1), we can prove the following.

Proposition 5: Suppose that the same assumptions as in Proposition 4 hold. Then, for any $x_0 \in \mathbb{R}^n$, we have

$$\|x(t, x_0)\|_2 \leq \|x_0 - z^*(0)\|_2 + \frac{\bar{b}}{\sqrt{\mu}}, \quad t \geq 0$$

that is, any solution of (6) is bounded for $t \geq 0$.

Proof: We have $\|x(t, x_0)\|_2 \leq \|x(t, x_0) - z^*(t)\|_2 + \|z^*(t)\|_2$, $t \geq 0$. By Proposition 3, $t \rightarrow \|x(t, x_0) - z^*(t)\|_2$ is nonincreasing, whereas, by Proposition 1, $\|z^*(t)\|_2 \leq \bar{b}/\sqrt{\mu}$ for any $t \geq 0$. This yields the stated result. ■

Let us now consider the question of which (candidate) Lyapunov function we should use for studying the dynamics of (6). As we have seen in Proposition 3, the Euclidean distance between any pair of solutions of (6) is monotonically nonincreasing with time. We have also proved in Proposition 4 that, when $\sigma > \bar{\sigma}$, $z^*(\cdot)$ is a solution of (6). Then, we are guaranteed that the Euclidean distance between any solution of (6) and $z^*(\cdot)$ is nonincreasing with time and can thus be used as a Lyapunov function for (6). This is stated more formally in the next result.

Proposition 6: Suppose that the same assumptions as in Proposition 4 hold. Then, for any $x_0 \in \mathbb{R}^n$ function

$t \rightarrow \|x(t, x_0) - z^*(t)\|_2$ is nonincreasing for any $t \geq 0$, hence

$$\lim_{t \rightarrow +\infty} \|x(t, x_0) - z^*(t)\|_2 = \Delta(x_0)$$

where $\|x_0 - z^*(0)\|_2 \geq \Delta(x_0) \geq 0$.

Proof: Follows from Proposition 3 and the fact that $x(\cdot, x_0)$ and $z^*(\cdot)$ are solutions of (6). ■

To complete the analysis it remains to show that we have $\Delta(x_0) = 0$ for any $x_0 \in \mathbb{R}^n$. The result in Proposition 6 follows from the monotonicity properties of the subdifferential of a convex function. By a more detailed evaluation of the derivative of function $t \rightarrow \|x(t, x_0) - z^*(t)\|_2$, which exploits the peculiar structure of the neural network equations, we can prove the next basic result on the reaching and tracking phase.

Theorem 1: Suppose that Assumptions 1 and 2 are satisfied, and $\sigma > \bar{\sigma}$, where $\bar{\sigma}$ is as in (7). Then, the following hold.

- 1) For any $x_0 \neq z^*(0) \in \mathbb{R}^n$, and for a.a. $t \geq 0$ such that $x(t, x_0) \neq z^*(t)$, we have

$$\frac{d}{dt} \|x(t, x_0) - z^*(t)\|_2 \leq -\beta^2 < 0$$

where

$$\beta^2 \doteq \sigma\sqrt{\mu} - L_* = \sigma\sqrt{\mu} - \frac{1}{\sqrt{\mu}}L_b - \frac{1}{\mu}L_A\bar{b} > 0. \quad (8)$$

- 2) If $x_0 \neq z^*(0)$, there exists

$$t_h < \frac{\|x_0 - z^*(0)\|_2}{\beta^2} \quad (9)$$

such that $x(t, x_0) \neq z^*(t)$, $t \in [0, t_h)$, and $x(t, x_0) = z^*(t)$ for any $t \geq t_h$. If $x_0 = z^*(0)$, then we have $x(t, x_0) = z^*(t)$ for any $t \geq 0$.

Proof: The proof is divided into two parts.

- 1) Let $V(t, x) = \|x - z^*(t)\|_2 = (\sum_{i=1}^n (x_i - z_i^*(t))^2)^{1/2}$ for any $t \geq 0$ and $x \in \mathbb{R}^n$. For any $t \geq 0$, $x \rightarrow V(t, x)$ is convex. Moreover, it is differentiable with respect to x for $x \neq z^*(t)$ and we have $\partial_x V(t, x) = (x - z^*(t))/\|x - z^*(t)\|_2$. Note that $\|\partial_x V(t, x)\|_2 = 1$ for $x \neq z^*(t)$, hence $\lim_{x \rightarrow z^*(t)} \|\partial_x V(t, x)\|_2 = 1$. Function $t \rightarrow V(t, x(t, x_0))$ is differentiable for a.a. $t \geq 0$ such that $x(t, t_0) \neq z^*(t)$ and

$$\begin{aligned} \frac{d}{dt} V(t, x(t, x_0)) &= \left\langle \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2}, \dot{x}(t, x_0) - \dot{z}^*(t) \right\rangle \\ &= \left\langle \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2}, -\sigma A'(t)\gamma(t) \right\rangle \\ &\quad - \left\langle \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2}, \dot{z}^*(t) \right\rangle \\ &= -\sigma \left\langle A(t) \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2}, \gamma(t) \right\rangle \\ &\quad - \left\langle \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2}, \dot{z}^*(t) \right\rangle \end{aligned}$$

where $\gamma(t) \in \text{SGN}(A(t)x(t, x_0) - b(t)) = \text{SGN}(A(t)(x(t, x_0) - z^*(t)))$. Let $e(t) = x(t, x_0) - z^*(t)$ and note that $\langle A(t)(x(t, x_0) - z^*(t)), \gamma(t) \rangle = \langle A(t)e(t), \gamma(t) \rangle = \sum_{i=1}^n (A(t)e(t))_i \gamma_i(t)$, where

$\gamma_i(t) \in \text{sgn}(A(t)e(t))_i$. Then, we have $\langle A(t)(x(t, x_0) - z^*(t)), \gamma(t) \rangle = \sum_{i=1}^n |(A(t)e(t))_i| = \|A(t)e(t)\|_1$, and so

$$\begin{aligned} \frac{d}{dt}V(t, x(t, x_0)) &\leq -\sigma \left\| A(t) \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2} \right\|_1 \\ &\quad - \left\langle \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2}, \dot{z}^*(t) \right\rangle \\ &\leq -\sigma \left\| A(t) \frac{x(t, x_0) - z^*(t)}{\|x(t, x_0) - z^*(t)\|_2} \right\|_2 \\ &\quad + \|\dot{z}^*(t)\|_2 \\ &\leq -\sigma \min_{\|w\|_2=1} \|A(t)w\|_2 + \|\dot{z}^*(t)\|_2 \end{aligned}$$

where we considered that the Euclidean norm of vector $(x(t, x_0) - z^*(t))/\|x(t, x_0) - z^*(t)\|_2$ is equal to 1. Note that $\|A(t)w\|_2^2 = (A(t)w)'A(t)w = w'A'(t)A(t)w \geq \lambda_{\min}(A'(t)A(t))\|w\|_2^2$. Due to Assumption 1, for $\|w\|_2 = 1$ we obtain $\|A(t)w\|_2^2 \geq \lambda_{\min}(A'(t)A(t)) \geq \mu$ and so

$$\frac{d}{dt}V(t, x(t, x_0)) \leq -\sigma\sqrt{\mu} + L_* = -\beta^2$$

for a.a. $t \geq 0$ such that $x(t, x_0) \neq z^*(t)$, where we have taken into account that $\|\dot{z}^*(t)\|_2 \leq L_*$ for a.a. $t \geq 0$.

- 2) Let $x_0 \neq z^*(0)$ and let $\tau > 0$ be such that $x(t, x_0) \neq z^*(t)$, $t \in [0, \tau]$. By point 1) we have $dV(t, x(t, x_0))/dt = d\|x(t, x_0) - z^*(t)\|_2/dt < -\beta^2 < 0$ and hence

$$\|x(\tau, x_0) - z^*(\tau)\|_2 - \|x_0 - z^*(0)\|_2 < -\beta^2\tau.$$

This implies there exists $t_h < \|x_0 - z^*(0)\|_2/\beta^2$ such that $x(t_h, x_0) = z^*(t_h)$ and $x(t, x_0) \neq z^*(t)$, $t \in [0, t_h)$. Moreover, since $x(\cdot, x_0)$ and $z^*(\cdot)$ are solutions of (6), due to Proposition 6 the Euclidean distance $t \rightarrow \|x(t, x_0) - z^*(t)\|_2$ is nonincreasing for $t \geq 0$. Then, $x(t, x_0) = z^*(t)$, $t \geq t_h$. Suppose now that $x_0 = z^*(0)$. Arguing as before we conclude that $x(t, x_0) = z^*(t)$ for any $t \geq 0$. ■

Corollary 1: Suppose that the same assumptions as in Theorem 1 are satisfied and let $r > 0$. Then, the time needed for any solution $x(t, x_0)$ of (6), with $x_0 \in S(0, r)$, to reach $z^*(\cdot)$, is uniformly bounded above by

$$t_h(r) = \frac{\bar{b}}{\beta^2} + r.$$

Moreover, the solutions of (6) starting in $S(0, r)$ display a common asymptotic behavior for $t \geq t_h(r)$, namely, we have $x(t, x_0) = z^*(t)$ for any $x_0 \in S(0, r)$ and $t \geq t_h(r)$.

Proof: The result is an immediate consequence of Theorem 1 and Proposition 1. ■

B. Tightness of Estimated Threshold

Let us consider the class $\mathcal{L}(L_A, L_b, \bar{b}, \mu)$ of TD-ALEs (2) such that $t \rightarrow A(t)$ is Lipschitz with Lipschitz constant $L_A \geq 0$ for $t \geq 0$, $t \rightarrow b(t)$ is Lipschitz with Lipschitz constant $L_b \geq 0$ for $t \geq 0$, and such that Assumptions 1 and 2 are satisfied with given constants $\mu > 0$ and $\bar{b} > 0$.

Given a specific TD-ALE in the class $\mathcal{L}(L_A, L_b, \bar{b}, \mu)$, it may happen that the threshold $\bar{\sigma} = \bar{\sigma}(L_A, L_b, \bar{b}, \mu)$ in (7) of

Theorem 1 is quite a conservative estimate for ensuring the reaching and tracking phase of $z^*(\cdot)$ (see Section V for an example of this kind). However, $\bar{\sigma}$ is a tight estimate with respect to the whole class $\mathcal{L}(L_A, L_b, \bar{b}, \mu)$ in the following sense. Suppose to choose any $0 < \sigma < \bar{\sigma}$. Then, we can find TD-ALEs in the class $\mathcal{L}(L_A, L_b, \bar{b}, \mu)$ such that Theorem 1 fails. In particular, $z^*(\cdot)$ is no longer a solution of (6) and for any solution $x(\cdot, x_0)$ of (6) such that $x_0 \neq z^*(0)$ the tracking phase of $z^*(\cdot)$ fails, namely, there exists a sequence $t_k \rightarrow +\infty$ such that $x(t_k, x_0) \neq z^*(t_k)$. Next, we discuss this point with a simple 1-D example.

Example: Consider the 1-D TD-ALE

$$\dot{z}(t) - b(t) = 0$$

where $z(t) \in \mathbb{R}$

$$b(t) = \begin{cases} 1-t & t \in [0, 2) \\ t-3 & t \in [2, 4) \end{cases}$$

and $b(t+4) = b(t)$ for any $t \geq 0$, is a triangular wave with period 4 as depicted in Fig. 1. The solution is $z^*(t) = b(t)$, $t \geq 0$, and the corresponding neural network is

$$\dot{x} \in -\sigma \text{sgn}(x - b(t)) \quad (10)$$

for a.a. $t \geq 0$. Clearly, $L_A = 0$, $L_b = 1$ and $L_* = 1$, $\bar{\sigma} = 1$.

Suppose to pick any $0 < \sigma < \bar{\sigma} = 1$. Since $|z^*(t)| = 1$ for a.a. $t \geq 0$, and $[-\sigma \text{sgn}(0)] = [-\sigma, \sigma]$, $z^*(\cdot)$ is not a solution of (10) and it is expected that (10) would not be able to track $z^*(\cdot)$. The global behavior of (10) can be indeed described as follows. First of all, a straightforward computation shows that, for any $0 < \sigma < 1$, (10) has a periodic solution given by

$$\tilde{z}(t) = \begin{cases} \sigma^2 + \sigma t & t \in [0, 1 - \sigma) \\ 2\sigma - \sigma^2 - \sigma t & t \in [1 - \sigma, 3 - \sigma) \\ -4\sigma + \sigma^2 + \sigma t & t \in [3 - \sigma, 4) \end{cases}$$

and $\tilde{z}(t+4) = \tilde{z}(t)$ for any $t \geq 0$, i.e., $\tilde{z}(\cdot)$ is a triangular wave with period 4, slope σ for a.a. $t \geq 0$, which is shifted by $1 - \sigma$ with respect to $z^*(\cdot)$ (Fig. 1). Note that $|z^*(t) - \tilde{z}(t)|$ has a maximum at the instants $t_k = 2k = 2, 4, 6, \dots$, given by $|z^*(t_k) - \tilde{z}(t_k)| = 1 - \sigma^2 > 0$. We can now show that, for any $x_0 \in \mathbb{R}$, the unique solution $x(t, x_0)$ of (10) with initial condition x_0 converges to $\tilde{z}(\cdot)$ as $t \rightarrow +\infty$. Then, $\liminf_{t \rightarrow +\infty} |x(t, x_0) - z^*(t)| = \limsup_{t \rightarrow +\infty} |x(t, x_0) - z^*(t)| = 1 - \sigma^2 > 0$, i.e., the tracking error $|x(t, x_0) - z^*(t)|$ does not vanish as $t \rightarrow +\infty$. As an example, Fig. 1 depicts the time-domain behavior of the solution of (10) starting at $x_0 = 2$ when $\sigma = 0.85 < \bar{\sigma} = 1$, which is seen to converge to $\tilde{z}(\cdot)$ and display a residual error with peaks $1 - 0.85^2 = 0.228$ for large times.

Suppose without loss of generality that $x_0 > \tilde{z}(0)$ and let us show global convergence of $x(\cdot, x_0)$ to $\tilde{z}(\cdot)$. Since $\tilde{z}(\cdot)$ is a bounded solution of (10), due to Proposition 6, $x(\cdot, x_0)$ is bounded on $[0, +\infty)$ and the distance $t \rightarrow \delta(t) \doteq |x(t, x_0) - \tilde{z}(t)|$ is a monotonically decreasing function on $[0, +\infty)$. Hence, $x(t, x_0) \geq \tilde{z}(t)$ for any $t \geq 0$ and $\lim_{t \rightarrow +\infty} \delta(t) = \Delta(x_0) \geq 0$. Suppose for contradiction that $\Delta(x_0) > 0$. We have $\dot{\delta}(t) \in -\sigma[\text{sgn}(x(t, x_0) - b(t)) - \text{sgn}(\tilde{z}(t) - b(t))]$ for a.a. $t \geq 0$. Let $\bar{a} = \min\{2, \Delta(x_0)/(1 + \sigma)\} > 0$ and consider for any $k = 1, 2, \dots$, the intervals $J_k = (4k + 3 - \sigma - \bar{a}, 4k + 3 - \sigma)$.

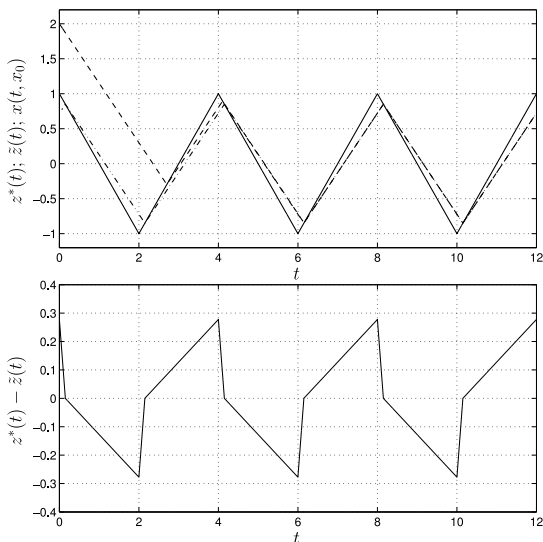


Fig. 1. Upper plot: triangular waves $b(t) = z^*(t)$ (solid) and $\tilde{z}(\cdot)$ (dashed-dotted) and solution $x(t, x_0)$ of neural network (10) when $x_0 = 2$ and $\sigma = 0.85 < \bar{\sigma} = 1$. Lower plot: error $z^*(t) - \tilde{z}(t)$ when $\sigma = 0.85$.

We can check that in J_k we have $x(t, x_0) - b(t) = \tilde{z}(t) + \delta(t) - b(t) > 0$ and $\tilde{z}(t) - b(t) < 0$. Hence, $\dot{\delta}(t) = -\sigma[\text{sgn}(\tilde{z}(t) + \delta(t) - b(t)) - \text{sgn}(\tilde{z}(t) - b(t))] = -2\sigma < 0$ for a.a. $t \in J_k$, i.e., for any k the distance $\delta(\cdot)$ decreases of at least $-2\tilde{a}\sigma$ in J_k , which yields a contradiction. Then, $\Delta(x_0) = 0$ and so $x(t, x_0) \rightarrow \tilde{z}(t)$ as $t \rightarrow +\infty$.

C. Application to TI-ALEs

Theorem 1 can be applied also in the limiting case where A and b do not depend on time, i.e., to TI-ALEs

$$Az - b = 0. \quad (11)$$

In this case, we have $L_A = L_b = L_* = 0$ and so $\bar{\sigma} = 0$. Suppose A is nonsingular. Then, Assumption 1 is satisfied with $\mu = \min_i \lambda_i(A'A) > 0$ and also Assumption 2 is satisfied with $\bar{b} = \|b\|_2$. The unique solution of (11) is $z^* = A^{-1}b$.

The neural network equations for TI-ALEs are

$$\dot{x} \in -\sigma \partial_x \|Ax - b\|_1 = -\sigma A' \text{SGN}(Ax - b) \quad (12)$$

where $\sigma > 0$. Since $0 \in -\sigma A' \text{SGN}(Az^* - b) = -\sigma A' \text{SGN}(0) = -\sigma A'[-1, 1]^n$, it follows that z^* is an equilibrium point of (12), i.e., it is a constant solution of (12). Theorem 1 yields the following.

Theorem 2: Let $\sigma > 0$. Then, for any $x_0 \neq z^* \in \mathbb{R}^n$ and a.a. $t \geq 0$ such that $x(t, x_0) \neq z^*$

$$\frac{d}{dt} \|x(t, x_0) - z^*\|_2 \leq -\beta^2 < 0$$

where $\beta^2 = \sigma \sqrt{\mu} > 0$ and $x(\cdot, x_0)$ is the unique solution of (12) with $x(0) = x_0$. If $x_0 \neq z^*$, there exists

$$t_h < \frac{\|x_0 - z^*\|_2}{\beta^2} \quad (13)$$

such that $x(t, x_0) \neq z^*$, $t \in [0, t_h)$, and $x(t, x_0) = z^*$ for any $t \geq t_h$. If $x_0 = z^*$, then $x(t, x_0) = z^*$, $t \geq 0$.

An analogous result has been obtained in [11, Th. 1], although that result does not ensure the uniqueness of solutions with respect to initial conditions.

We stress that Theorem 2 holds for any $\sigma > 0$. Namely, for any $\sigma > 0$, z^* is the unique equilibrium point of (12) and there is global uniform convergence in finite time to z^* . The value of σ only influences the finite convergence time t_h to z^* [see (13)]. Quite on the contrary, Theorem 1, and the results in Section III-A, show that for TD-ALEs the penalty parameter σ needs to be sufficiently large, i.e., it needs to exceed threshold $\bar{\sigma}$, in order that $z^*(\cdot)$ is a solution of (6) and that the reaching and tracking phase can be performed by (6). Once condition $\sigma > \bar{\sigma}$ is satisfied, σ also influences the reaching time t_h , i.e., the larger σ , the smaller t_h is [see (9)].

IV. PERTURBATION RESULT

The interconnection matrices $A'(t)$, $A(t)$ in neural network model (6) are electronically implemented by two distinct sets of conductances (see the circuit schematic in [8, Fig. 1(a)]). Therefore, a more realistic model accounting for tolerances in the implementation of the interconnection matrices may be written as follows:

$$\dot{y} \in -\sigma(A(t) + \Delta A_1(t))' \text{SGN}((A(t) + \Delta A_2(t))'y - b(t)) \quad (14)$$

for a.a. $t \geq 0$, where $\Delta A_1(t)$, $\Delta A_2(t) \in \mathbb{R}^{n \times n}$. Actually, perturbations $\Delta A_1(t)$, $\Delta A_2(t)$ are independent of each other, hence, $\Delta A_1(t) \neq \Delta A_2(t)$. Our goal is to show that the proposed neural network enjoys some robustness properties with respect to these implementation errors.

For the sake of brevity, we consider in what follows the simplified perturbed model:

$$\dot{y} \in -\sigma(A(t) + \Delta A(t))' \text{SGN}(A(t)y - b(t)) \quad (15)$$

for a.a. $t \geq 0$, where $t \rightarrow \Delta A(t) \in \mathbb{R}^{n \times n}$ is Lipschitz continuous on $[0, +\infty)$. This is reasonable due to the following considerations. The role of the constraint neurons with a hard-comparator input-output activation function in model (14) is to force their input $(A(t) + \Delta A_2(t))'y - b(t)$ to go to 0 with time (see Section II). Then, the effect of perturbation $\Delta A_2(t)$ in (14) is to change the target solution $\tilde{z}^*(t) = (A(t) + \Delta A_2(t))^{-1}b(t)$ the neural network tries to reach and track. The problem of how $\tilde{z}^*(t)$ approximates the desired solution $z^*(t)$ can be treated by standard algebraic techniques and may be omitted here. Instead, $\Delta A_1(t)$ affects more substantially the dynamics of the neural network. Indeed, even if the constraint neurons force (14) to reach and track the ideal solution $z^*(t)$, yet when $\Delta A_1(t) \neq 0$ the dynamics of (14) is no longer described by a gradient-type system with respect to x , and this may disrupt the reaching and tracking phase analyzed in Theorem 1.

We can prove for the perturbed system (15) the same results on existence of global solutions as in Proposition 2. Instead, since we are dealing with multivalued perturbations of a maximal monotone operator, uniqueness of solutions with respect to initial conditions might fail. Then, in what follows, we give a result on finite convergence holding for any possible solution of (15) starting at a given initial condition.

Suppose we are given $\underline{\Delta A}, \overline{\Delta A} \in \mathbb{R}^{n \times n}$ satisfying $\underline{\Delta A} \leq \overline{\Delta A}$, where the inequality holds with respect to each entry. We assume henceforth that $\underline{\Delta A} \leq \Delta A(t) \leq \overline{\Delta A}$ for any $t \geq 0$, i.e., $\Delta A(\cdot)$ belongs to a class of interval matrices. Interval matrices are widely used in the neural network literature for modeling perturbations of interconnections and studying robustness of convergence and global stability (see [44]–[46] and references therein).

By means of [44, Lemma 1], we can obtain an estimate of the maximal norm of interval matrices. Let $A^* = (1/2)(\overline{A} + \underline{A})$, $A_* = (1/2)(\overline{A} - \underline{A})$, and

$$N_{\max}(\underline{A}, \overline{A}) \doteq \sqrt{\| (A^*)' A^* \| + 2 \| (A^*)' A_* + A_*' A^* \|} \quad (16)$$

where for any $M \in \mathbb{R}^{n \times n}$ we have let $|M| = (|m_{ij}|)$. Then, we have $\|A(t)\|_2 \leq N_{\max}(\underline{A}, \overline{A})$ for any $\underline{A} \leq A(t) \leq \overline{A}$.

Theorem 3: Suppose that Assumptions 1 and 2 hold, and that

$$N_{\max}(\underline{\Delta A}, \overline{\Delta A}) < \sqrt{\frac{\mu}{n}} \quad (17)$$

$$\sigma > \hat{\sigma} \doteq \frac{L_*}{\sqrt{\mu} - \sqrt{n} N_{\max}(\underline{\Delta A}, \overline{\Delta A})} \quad (18)$$

where L_* is as in (4). Let $y_0 \in \mathbb{R}^n$ and let $y(t)$, $t \geq 0$, be a solution of (15) with $y(0) = y_0$. Then, the following hold.

- 1) For any $y_0 \neq z^*(0) \in \mathbb{R}^n$, and for a.a. $t \geq 0$ such that $y(t) \neq z^*(t)$, we have

$$\frac{d}{dt} \|y(t) - z^*(t)\|_2 \leq -\hat{\beta}^2 < 0$$

where

$$\hat{\beta}^2 \doteq \sigma (\sqrt{\mu} - \sqrt{n} N_{\max}(\underline{\Delta A}, \overline{\Delta A})) - L_* > 0.$$

- 2) If $y_0 \neq z^*(0)$, there exists

$$t_h < \frac{\|y_0 - z^*(0)\|_2}{\hat{\beta}^2}$$

such that $y(t) \neq z^*(t)$, $t \in [0, t_h)$ and $y(t) = z^*(t)$, $t \geq t_h$.

If $y_0 = z^*(0)$, then $y(t) = z^*(t)$ for any $t \geq 0$.

Proof: See the Appendix. ■

Theorem 3 states that the reaching and tracking behavior in Theorem 1 is robust with respect to some implementation errors. Note that (17) is the maximum allowable perturbation still ensuring convergence to $z^*(\cdot)$ for the considered class of perturbation matrices. Theorem 3 is one of the few existing results on robustness of convergence with respect to perturbations of neural gradient-type systems. Other robustness results of this kind for TI gradient systems have been obtained in [47] and [48]. We note that in the general case even arbitrarily small perturbations of convergent gradient systems may instead lead to nonconvergent dynamics [49].

V. EXAMPLES

Example 1: Consider the 2-D TD-ALE (2) defined by the rotation matrix with speed $\omega = 3$ rad/s

$$A(t) = \begin{pmatrix} \cos(3t) & \sin(3t) \\ -\sin(3t) & \cos(3t) \end{pmatrix}$$

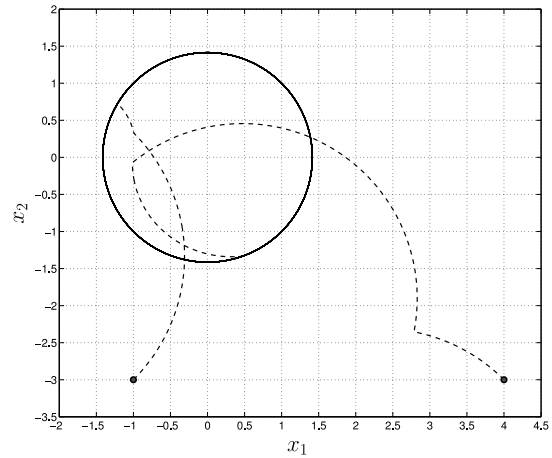


Fig. 2. Trajectories of neural network (6) with $\sigma = 5$ starting at the initial conditions $x_0 = (4, -3)'$ and $x_a = (-1, -3)'$ and converging in finite time to the target solution $z^*(\cdot)$ of the TD-ALE problem in Example 1. The trajectory corresponding to the solution $z^*(\cdot)$ is depicted with a solid line.

and the constant vector $b(t) = (1, 1)'$, $t \geq 0$. We have $A^{-1}(t) = A'(t)$, hence for any $t \geq 0$ the theoretical solution is $z^*(t) = (\cos(3t) - \sin(3t), \sin(3t) + \cos(3t))'$.

Note that $A(\cdot)$ is a smooth function of t and we have $L_A = \omega = 3$. Since, for any $t \geq 0$, $A'(t)A(t)$ is the identity matrix, Assumption 1 is satisfied with $\mu = 1$. Clearly, $L_b = 0$ and also Assumption 2 holds with $\bar{b} = \sqrt{2}$. Then, $L_* = \omega \bar{b} = 4.24$ and, by (7), $\bar{\sigma} = 4.24$. We have numerically simulated the unperturbed neural network (6) by means of MATLAB when $\sigma = 5 > \bar{\sigma}$. Fig. 2 depicts the trajectory $x(\cdot, x_0)$ of (6) starting at $x_0 = (4, -3)' \neq z^*(0) = (1, 1)'$. Fig. 3 depicts the corresponding time evolution of $x(\cdot, x_0)$ and $z^*(\cdot)$, whereas Fig. 4 depicts the time evolution of the error $\|x(\cdot, x_0) - z^*(\cdot)\|_2$. It is seen that $x(\cdot, x_0)$ reaches in finite time $t_h = 1.43 < \|x_0 - z^*(0)\|_2 / \beta^2 = 6.58$ the solution $z^*(\cdot)$ [see (9)], i.e., we have $x(1.43, x_0) = z^*(1.43)$. Moreover, $x(t, x_0) = z^*(t)$ for any $t \geq 1.43$, i.e., $x(\cdot, x_0)$ exactly tracks $z^*(\cdot)$ for $t \geq 1.43$. Fig. 2 also shows a second trajectory of (6) starting at a different initial condition $x_a = (-1, -3)' \neq z^*(0)$, which reaches $z^*(\cdot)$ at $t_h = 0.61$ and tracks $z^*(\cdot)$ thereafter. These results are in accordance with those predicted by Theorem 1.

We remark that, as it happens in practice, the threshold estimates may be quite conservative in specific cases (see Section III-B). For instance, we have verified numerically that the tracking and reaching phase of $z^*(\cdot)$ still works when $\sigma = 3.5$. Instead, if we further reduce σ , (6) may not be able to exactly perform the tracking phase. For example, when $\sigma = 2.5$ there is a significant residual tracking error that does not tend to 0 with time.

Consider now the perturbed neural network (15) for solving the same TD-ALE, and suppose $\Delta A(t) = (\delta a_{ij}(t))$ satisfies $\underline{\Delta A} \leq \Delta A(t) \leq \overline{\Delta A}$ for any $t \geq 0$, where

$$\overline{\Delta A} = -\underline{\Delta A} = \begin{pmatrix} k & k \\ k & k \end{pmatrix}$$

and $k > 0$ is a parameter. This implies that $\delta a_{ij}(t)$, $i, j \in 1, 2$, can be any bounded Lipschitz-continuous functions taking values in $[-k, k]$ for $t \geq 0$. It can be checked from (16) that

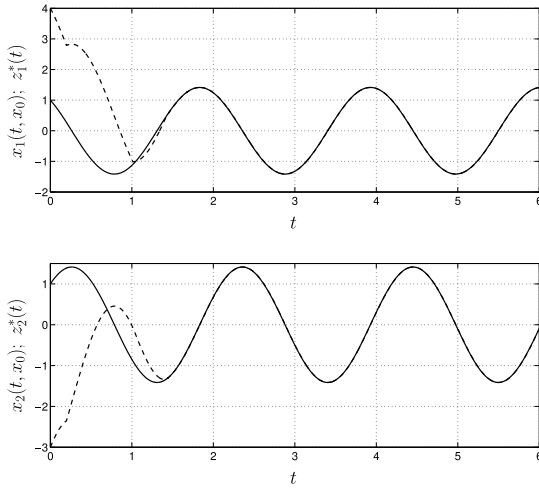


Fig. 3. Time-domain behavior of $x_1(t, x_0)$, $z_1^*(t)$ (upper plot) and $x_2(t, x_0)$, $z_2^*(t)$ (lower plot) for $x_0 = (4, -3)'$ and $\sigma = 5$.

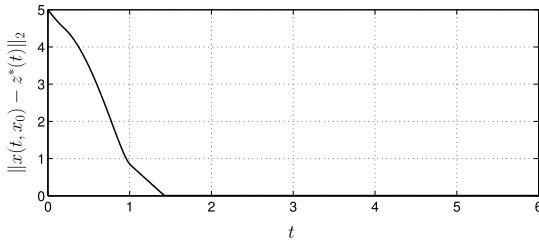


Fig. 4. Time-domain behavior of $\|x(t, x_0) - z^*(t)\|_2$ when $x_0 = (4, -3)'$ and $\sigma = 5$.

we have $N_{\max}(\underline{\Delta A}, \overline{\Delta A}) = \|\overline{\Delta A}\|_2 = 2k$, hence the maximum allowable perturbation corresponds to $k_{\max} = (1/2)\sqrt{\mu/n} = 1/(2\sqrt{2}) = 0.35$ [see (17)].

Let $k = 0.2$, in which case, by (18), $\hat{\sigma} = 9.8$. For numerical verification let us choose

$$\Delta A(t) = 0.2 \begin{pmatrix} \cos(0.3t) & -\sin(0.3t) \\ \sin(0.3t) & \cos(0.3t) \end{pmatrix}$$

as a lower frequency additive perturbation to $A(t)$ belonging to the considered class of interval matrices. We simulated (15) when $\sigma = 10 > \hat{\sigma}$. Fig. 5 depicts the time evolution for a solution $y(\cdot)$ of (15) starting at $y_0 = (4, -3)' \neq z^*(0) = (1, 1)'$. It can be verified that $y(\cdot)$ reaches $z^*(\cdot)$ in finite time $t_h = 0.68$, and exactly tracks $z^*(\cdot)$ thereafter, in agreement with Theorem 3. Several other numerical experiments have been carried out picking different waveforms for the entries of $\Delta A(t)$, such as low- or high-frequency sinusoidal waveforms, or combinations of sinusoidal, absolute value and other bounded Lipschitz-continuous functions. All of them produced results in agreement with Theorem 3 (we do not report them here for space limitations).

Example 2: Consider the 2-D TD-ALE (2), where

$$A(t) = \begin{pmatrix} w(t) & w(t-1) \\ -w(t-1) & w(t) \end{pmatrix}$$

and $w(\cdot)$ is a periodic triangular wave defined as $w(\rho) = 1 - \rho$ for $\rho \in [0, 2)$, $w(\rho) = \rho - 3$ for $\rho \in (2, 4]$, and $w(\rho + 4) = w(\rho)$ for any $\rho \geq 0$. Moreover, $b(t) = (1, 1)'$ for any $t \geq 0$.

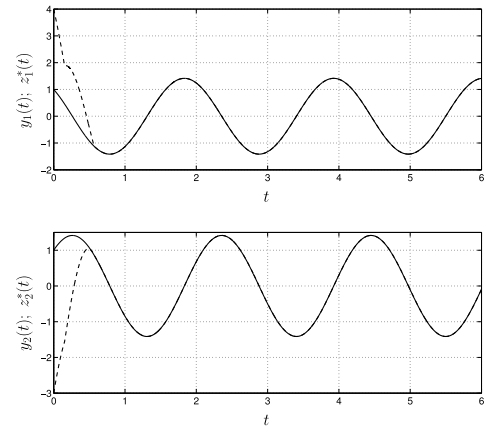


Fig. 5. Time-domain behavior of $y_1(t)$, $z_1^*(t)$ (upper plot) and $y_2(t)$, $z_2^*(t)$ (lower plot), where $y(\cdot)$ is a solution of the perturbed neural network (15) with $\sigma = 10$ with $y(0) = (4, -3)'$.

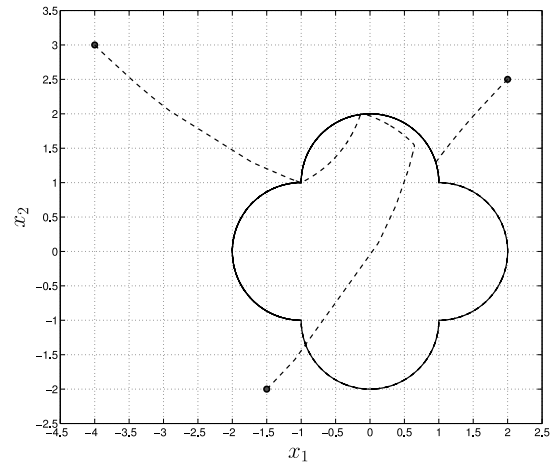


Fig. 6. Three trajectories of neural network (6) with $\sigma = 9$ starting at different initial conditions and converging in finite time to the solution $z^*(\cdot)$ of the TD-ALE of Example 2. The trajectory corresponding to the target solution $z^*(\cdot)$ is displayed with a solid line.

Note that $t \rightarrow A(t)$ is Lipschitz with $L_A = \sqrt{2}$ but it is not differentiable at the instants $t = 1, 2, 3, \dots$. It can be checked that Assumption 1 is satisfied with $\mu = 0.5$. Moreover, $L_b = 0$ and also Assumption 2 holds with $\bar{b} = \sqrt{2}$. Then, $L_* = 4$ and $\bar{\sigma} = 4\sqrt{2} \simeq 5.66$. We simulated neural network (6) for solving this problem when $\sigma = 9$. Fig. 6 depicts the trajectories of (6) starting at $x_0 = (-4, 3)'$, $x_a = (2, 2.5)'$, and $x_b = (-1.5, -2)'$. It can be checked that each trajectory reaches $z^*(\cdot)$ in finite time, and tracks $z^*(\cdot)$ thereafter, and that the reaching time agrees with that predicted by Theorem 1.

Consider the perturbed neural network (15) for solving the same TD-ALE, and suppose $\overline{\Delta A}$ is a biased perturbation such that $\underline{\Delta A} \leq \Delta A(t) \leq \overline{\Delta A}$ for any $t \geq 0$, where

$$\underline{\Delta A} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}; \quad \overline{\Delta A} = \begin{pmatrix} k & k \\ k & k \end{pmatrix}$$

and $k > 0$ is a parameter. From (16) it turns out that $N_{\max}(\underline{\Delta A}, \overline{\Delta A}) = 2\sqrt{3}k$, and the maximum allowable perturbation corresponds to $k_{\max} = 1/(4\sqrt{3}) \simeq 0.14$. Letting $k = 0.1$ gives $\hat{\sigma} = 18.2$ [see (18)]. Several numerical

simulations were carried out by choosing $\sigma > \hat{\sigma}$ and using different bounded Lipschitz-continuous waveforms for the entries of $\Delta A(t)$. Once more the reaching and tracking results of the perturbed network were in agreement with those predicted by Theorem 3.

VI. DISCUSSION

A. Nonsmooth Gradient-Type Neural Networks for TI-ALEs

Cichocki and Unbehauen [7], [8] introduced some classes of nonsmooth neural networks for the solution of TI-ALEs. The networks are the gradient of a nonsmooth energy function given by the L_1 norm of the error between the state and the ALE solution. Due to the difficulties to work with the gradient of nonsmooth energy functions, [7], [8] do not provide a theoretic analysis of the dynamical behavior, yet they show by means of simulations that those nonsmooth networks work well in the solution of classes of TI-ALEs and matrix inversion problems. In this paper, we have considered the same nonsmooth networks for TI-ALEs as in [7] and [8] and we have shown that it is possible to rigorously analyze them by means of tools from nonsmooth analysis, the concept of subdifferential (generalized gradient) and that of solutions in the sense of Filippov (Theorem 2). The results in this paper show that the nonsmooth networks in [7] and [8] are well suited also to work in a dynamic environment where they are able to reach in finite time and exactly track the solution of TD-ALEs when the threshold values are suitably chosen (Theorem 1).

Other relevant works in the literature have dealt with nonsmooth gradient neural networks for the solution of TI-ALEs by minimizing the L_1 norm of the error [11], [12]. In particular, Ferreira *et al.* [11] provided a dynamic analysis of a nonsmooth neural network using hard-comparator nonlinearities by writing the network as a Persidskii-type system and using nonsmooth diagonal Lyapunov functions and techniques proposed by Utkin for realizability of sliding manifolds. The idea of using a Lyapunov function to estimate penalty parameters that cause finite-time convergence of nonsmooth neural networks for solving TI-ALEs is also described in detail in [50, Sec. 4.2]. This paper actually shows that the networks in [11] can work in a dynamic environment and they are well suited also in the solution of TD-ALEs. Moreover, the present analysis permits to show, in addition to the results in [11], some more precise dynamical results, such as the uniqueness of solutions of the nonsmooth networks with respect to initial conditions. We also remark that the techniques from nonsmooth analysis used in this paper do not suffer from the drawbacks to handle sliding modes via Utkin approach as those pointed out in [12, p. 149].

B. Neural Networks for TD-ALEs

To the authors' knowledge the only systematic approach proposed so far for the solution of TD-ALEs is that based on the so-called Zhang neural network (ZNN) [14], [16], [18]. The main motivation for introducing ZNN is that smooth gradient-type neural networks may give rise to significant

residual errors in the solution of TD-ALEs [23]. ZNN is defined through an implicit dynamics (not a gradient dynamics) and exploits a vector-valued error function and the information on the time-derivative of the coefficients and parameters involved. Under suitable assumptions it is shown that ZNN is globally exponentially convergent to the solution of TD-ALEs or the inverse of TD matrices [14].

The most important aspect of the networks studied in this paper is that they are defined by the gradient of a nonsmooth error function. The main result here obtained is that nonsmooth gradient-type neural networks are well suited for finding in finite time the exact solution of TD-ALEs. We also remark that, unlike for ZNN, we do not need to use the information on the derivative of the TD coefficients involved, which is expected not to be simply/frequently available in practical applications. Moreover, we do not require that the coefficients are differentiable with respect to time, indeed we only assume that the coefficients satisfy the much weaker condition that they are Lipschitz functions of time.

C. Neural Networks for Convex TD Constraint Satisfaction Problems

In a recent paper, a nonsmooth neural network for solving convex TD constraint satisfaction problems has been introduced [51]. Under suitable assumptions on the constraints involved the network is able to reach in finite time the TD convex feasibility set and stay within the set thereafter, i.e., it is able to find a feasible solution in finite time and also yield an exact feasible solution for subsequent times. The solution of a TD-ALE is actually equivalent to finding a feasible solution of a convex problem defined by linear equality constraints. This notwithstanding, the analysis and results in [51] cannot be applied to the solution of TD-ALEs. The main reason is that in [51] it is required that the TD convex set has nonempty interior for any t , whereas in the solution of the TD-ALEs here considered the feasible set is a unique point $z^*(t)$ for any $t \geq 0$. The techniques here used for the dynamic analysis of (6) have been *ad hoc* developed to guarantee that the network is able to reach and track feasible sets with empty interior. In this sense they may be thought of as being a generalization to the TD case of techniques previously devised in [25] for finding in finite time feasible solutions of TI optimization problems with linear constraints.

D. Under- and Over-Determined ALEs

The TD-ALEs (2) considered in this paper have n equations in the same number of unknowns and they admit for any $t \geq 0$ a unique solution $z^*(t)$ which is Lipschitz continuous in t . The neural networks studied in [7] and [8] can be applied also to over-determined TI-ALEs where the number of equations m is larger than the number n of unknowns. Such ALEs are of importance, for example, in identification problems. The networks studied in [11] can be instead used also to solve under-determined TI-ALEs where $m < n$, which are of interest for example in support vector machines. It is then natural to ask whether the network studied in this paper can be extended and

applied to over- or under-determined TD-ALEs. Here, we only point out one main difficulty preventing this direct application. Namely, in the over- or under-determined case the set of minimizers $M(t)$ of $\|A(t)x - b(t)\|_1$ is in general a discontinuous function of t . As an example, consider the over-determined 2-D TD-ALE

$$A(t)z(t) - b(t) = 0$$

where $z(t) \in \mathbb{R}^2$, $m = 4$, $n = 2$

$$A(t) = \begin{pmatrix} 1 & -\sin(\pi t) \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad b(t) = \begin{pmatrix} 0 \\ -1.5 \\ -1 \\ 0 \end{pmatrix}$$

and suppose we wish to minimize $\|A(t)z(t) - b(t)\|_1$, i.e., to find the point (or the points) $z(t) \in \mathbb{R}^2$ minimizing the least absolute deviation from the trapezoidal set defined by the four straight lines $z_1(t) - \sin(\pi t)z_2(t) = 0$, $z_1(t) = 1.5$, $z_2(t) = 1$, and $z_2(t) = 0$. It can be checked that for any $t \in (2k, 2k+1)$, and $k = 0, 1, 2, \dots$, this minimization problem has a unique solution $M(t) = (1.5, 1)'$, whereas for $t \in (2k+1, 2(k+1))$ the unique minimizer is given by $M(t) = (1.5, 0)'$. At the instants $t_k = k$ we have $M(k) = [0, 1.5] \times [0, 1]$, i.e., $M(k)$ is a whole rectangle in \mathbb{R}^2 . Note that, although matrix $A'(t)A(t)$ has full row-rank for any $t \geq 0$, we have $M(t_k^-) = (1.5, 1)' \neq M(t_k^+) = (0, 1.5)'$, i.e., $M(\cdot)$ is discontinuous at any instant t_k . Clearly, an absolutely continuous solution of neural network (6) would not be able to exactly track $M(\cdot)$. We believe that an interesting topic for future research would be to investigate whether suitable modifications of the proposed neural network dynamics can permit to handle also these more difficult discontinuous cases.

VII. CONCLUSION

This paper has studied a neural network for solving TD-ALEs that is defined by the nonsmooth gradient with respect to the state variables of the L_1 norm of the error. It has been shown that, under suitable invertibility assumptions, there is a threshold for the penalty parameter beyond which the network is able to reach and exactly track the TD-ALE solution. Moreover, it has been shown that the estimated threshold is tight with respect to the whole set of considered TD-ALEs and that the finite time convergence results are robust with respect to perturbations of the nominal interconnection matrices.

One message conveyed by this paper is that neural networks defined by nonsmooth gradient systems can be effectively and rigorously analyzed via nonsmooth techniques. Moreover, by exploiting the sliding modes of the nonsmooth dynamics they are well suited to find in finite time the exact solutions of linear algebra problems also in a TD context. This differs from neural networks defined by smooth gradient systems, which give rise to significant residual errors in the solution of TD-ALEs.

APPENDIX PROOF OF THEOREM 3

Let $y(t)$, $t \geq 0$, be a solution of (15) with $y(0) = y_0 \in \mathbb{R}^n$. Consider the TD function $t \rightarrow V(t, y(t)) = (\sum_{i=1}^n (y_i(t) - z_i^*(t))^2)^{1/2}$. For a.a. $t \geq 0$ such that $y(t) \neq z^*(t)$ we have

$$\begin{aligned} \frac{d}{dt}V(t, y(t)) &= \left\langle \frac{y(t) - z^*(t)}{\|y(t) - z^*(t)\|_2}, \dot{y}(t) - \dot{z}^*(t) \right\rangle \\ &= -\sigma \left\langle A(t) \frac{y(t) - z^*(t)}{\|y(t) - z^*(t)\|_2}, \gamma(t) \right\rangle \\ &\quad - \sigma \left\langle \frac{y(t) - z^*(t)}{\|y(t) - z^*(t)\|_2}, \Delta A'(t)\gamma(t) \right\rangle \\ &\quad - \left\langle \frac{y(t) - z^*(t)}{\|y(t) - z^*(t)\|_2}, \dot{z}^*(t) \right\rangle \end{aligned}$$

where $\gamma(t) \in \text{SGN}(A(t)y(t) - b(t))$. Then, by a calculation as in the proof of Theorem 1, and considering that $\|\gamma(t)\|_2 \leq \sqrt{n}$ for a.a. $t \geq 0$, we obtain

$$\begin{aligned} \frac{d}{dt}V(t, y(t)) &\leq -\sigma \left\| A(t) \frac{y(t) - z^*(t)}{\|y(t) - z^*(t)\|_2} \right\|_2 \\ &\quad + \sigma \left\| \Delta A'(t)\gamma(t) \right\|_2 + \|\dot{z}^*(t)\|_2 \\ &\leq -\sigma \min_{\|w\|_2=1} \|A(t)w\|_2 \\ &\quad + \sigma \sqrt{n} \|\Delta A(t)\|_2 + \|\dot{z}^*(t)\|_2 \\ &\leq -\sigma \sqrt{\mu} + \sigma \sqrt{n} \|\Delta A(t)\|_2 + L_* \end{aligned}$$

By (16), $dV(t, y(t))/dt \leq -\sigma \sqrt{\mu} + \sigma \sqrt{n} N_{\max}(\overline{\Delta A}, \underline{\Delta A}) + L_*$ for a.a. $t \geq 0$ such that $y(t) \neq z^*(t)$. Since $N_{\max}(\overline{\Delta A}, \underline{\Delta A}) < \sqrt{\mu/n}$ and $\sigma > \hat{\sigma} > 0$ we obtain $dV(t, y(t))/dt \leq -\hat{\beta}^2 < 0$ for a.a. $t \geq 0$ such that $y(t) \neq z^*(t)$. We can now proceed as in the proof of Theorem 1 to obtain the stated result on convergence in finite time.

Next we show that $y(t) = z^*(t)$ for any $t \geq t_h$. For contradiction, assume there exists $t_e > t_h$ such that $y(t_e) \neq z^*(t_e)$, i.e., $\|y(t_e) - z^*(t_e)\|_2 > 0$. Let $t_a = \inf\{t > t_h : y(t) \neq z^*(t)\}$. Since $t \rightarrow \|y(t) - z^*(t)\|_2$ is continuous, there exists $h > 0$ such that $\|y(t) - z^*(t)\|_2 > 0$, $t \in (t_a, t_a + h]$. We have $d\|y(t) - z^*(t)\|_2/dt < -\beta^2 < 0$ for a.a. $t \in (t_a, t_a + h)$, hence $\|y(t_a + h) - z^*(t_a + h)\|_2 - \|y(t_a) - z^*(t_a)\|_2 = \|y(t_a + h) - z^*(t_a + h)\|_2 < -\beta^2 h < 0$, a contradiction.

Finally, let $y_0 = z^*(0)$. Arguing as before we conclude that $y(t) = z^*(t)$ for any $t \geq 0$. ■

REFERENCES

- [1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Solutions Manual to Accompany Nonlinear Programming: Theory and Algorithms*. Hoboken, NJ, USA: Wiley, 2013.
- [2] S. Huffel and J. Vandewalle, "Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error," *SIAM J. Matrix Anal. Appl.*, vol. 10, no. 3, pp. 294–315, 1989.
- [3] S.-Y. Kung, "VLSI array processors," *IEEE ASSP Mag.*, vol. 2, no. 3, pp. 4–22, Jul. 1985.
- [4] R. J. Steriti and M. A. Fiddy, "Regularized image reconstruction using SVD and a neural network method for matrix inversion," *IEEE Trans. Signal Process.*, vol. 41, no. 10, pp. 3074–3077, Oct. 1993.
- [5] T. Sarkar, K. Siarkiewicz, and R. Stratton, "Survey of numerical methods for solution of large systems of linear equations for electromagnetic field problems," *IEEE Trans. Antennas Propag.*, vol. 29, no. 6, pp. 847–856, Nov. 1981.

- [6] R. H. Sturges, "Analog matrix inversion," *IEEE J. Robot. Autom.*, vol. 4, no. 2, pp. 157–162, Apr. 1988.
- [7] A. Cichocki and R. Unbehauen, "Neural networks for solving systems of linear equations and related problems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 39, no. 2, pp. 124–138, Feb. 1992.
- [8] A. Cichocki and R. Unbehauen, "Neural networks for solving systems of linear equations. II. Minimax and least absolute value problems," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 39, no. 9, pp. 619–633, Sep. 1992.
- [9] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. Chichester, U.K.: Wiley, 1993.
- [10] J. Wang, "Recurrent neural networks for solving linear matrix equations," *Comput. Math. Appl.*, vol. 26, no. 9, pp. 23–34, Nov. 1993.
- [11] L. V. Ferreira, E. Kaszkurewicz, and A. Bhaya, "Solving systems of linear equations via gradient systems with discontinuous righthand sides: Application to LS-SVM," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 501–505, Mar. 2005.
- [12] L. Hsu, E. Kaszkurewicz, and A. Bhaya, "Matrix-theoretic conditions for the realizability of sliding manifolds," *Syst. Control Lett.*, vol. 40, no. 3, pp. 145–152, Jul. 2000.
- [13] Z. Zeng, X. Zhou, and B. Zhang, "A neural-network approach solving symmetric linear system," *J. Inf. Comput. Sci.*, vol. 9, no. 8, pp. 2305–2311, 2012.
- [14] Y. Zhang and S. S. Ge, "Design and analysis of a general recurrent neural network model for time-varying matrix inversion," *IEEE Trans. Neural Netw.*, vol. 16, no. 6, pp. 1477–1490, Nov. 2005.
- [15] J. Wang, "Recurrent neural networks for computing pseudoinverses of rank-deficient matrices," *SIAM J. Sci. Comput.*, vol. 18, no. 5, pp. 1479–1493, 1997.
- [16] Y. Zhang, D. Jiang, and J. Wang, "A recurrent neural network for solving Sylvester equation with time-varying coefficients," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1053–1063, Sep. 2002.
- [17] K. Chen, "Recurrent implicit dynamics for online matrix inversion," *Appl. Math. Comput.*, vol. 219, no. 20, pp. 10218–10224, Jun. 2013.
- [18] S. Li and Y. Li, "Nonlinearly activated neural network for solving time-varying complex Sylvester equation," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1397–1407, Aug. 2014.
- [19] Z. Wang and B. S. Peterson, "Constrained least absolute deviation neural networks," *IEEE Trans. Neural Netw.*, vol. 19, no. 2, pp. 273–283, Feb. 2008.
- [20] H.-N. Lee and G. J. Pottie, "Fast adaptive equalization/diversity combining for time-varying dispersive channels," *IEEE Trans. Commun.*, vol. 46, no. 9, pp. 1146–1162, Sep. 1998.
- [21] L. Ma, K. Dickson, J. McAllister, and J. McCanny, "QR decomposition-based matrix inversion for high performance embedded MIMO receivers," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1858–1867, Apr. 2011.
- [22] S. Teniou, M. Meribout, and K. Belarbi, "Real-time reconstruction of moving objects in an electrical capacitance tomography system using inter-frame correlation," *IEEE Sensors J.*, vol. 12, no. 7, pp. 2517–2525, Jul. 2012.
- [23] Y. Zhang, K. Chen, and H.-Z. Tan, "Performance analysis of gradient neural network exploited for online time-varying matrix inversion," *IEEE Trans. Autom. Control*, vol. 54, no. 8, pp. 1940–1945, Aug. 2009.
- [24] M. Forti, P. Nistri, and M. Quincampoix, "Generalized neural network for nonsmooth nonlinear programming problems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 9, pp. 1741–1754, Sep. 2004.
- [25] Q. Liu and J. Wang, "Finite-time convergent recurrent neural network with a hard-limiting activation function for constrained optimization with piecewise-linear objective functions," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 601–613, Apr. 2011.
- [26] W. Bian and X. Chen, "Smoothing neural network for constrained non-Lipschitz optimization with applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 399–411, Mar. 2012.
- [27] Q. Liu and J. Wang, "A one-layer recurrent neural network for constrained nonsmooth optimization," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 41, no. 5, pp. 1323–1333, Oct. 2011.
- [28] W. Bian and X. Xue, "Subgradient-based neural networks for nonsmooth nonconvex optimization problems," *IEEE Trans. Neural Netw.*, vol. 20, no. 6, pp. 1024–1038, Jun. 2009.
- [29] M. Forti and P. Nistri, "Global convergence of neural networks with discontinuous neuron activations," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 11, pp. 1421–1435, Nov. 2003.
- [30] W. Lu and T. Chen, "New approach to synchronization analysis of linearly coupled ordinary differential systems," *Phys. D, Nonlin. Phenom.*, vol. 213, no. 2, pp. 214–230, Jan. 2006.
- [31] D. Wang and L. Huang, "Periodicity and global exponential stability of generalized Cohen–Grossberg neural networks with discontinuous activations and mixed delays," *Neural Netw.*, vol. 51, pp. 80–95, Mar. 2014.
- [32] G. Bao and Z. Zeng, "Analysis and design of associative memories based on recurrent neural network with discontinuous activation functions," *Neurocomputing*, vol. 77, no. 1, pp. 101–107, 2012.
- [33] Y. Huang, H. Zhang, and Z. Wang, "Dynamical stability analysis of multiple equilibrium points in time-varying delayed recurrent neural networks with discontinuous activation functions," *Neurocomputing*, vol. 91, pp. 21–28, Aug. 2012.
- [34] X. Hu and J. Wang, "Design of general projection neural networks for solving monotone linear variational inequalities and linear and quadratic optimization problems," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1414–1421, Oct. 2007.
- [35] X. Liu and J. Cao, "Robust state estimation for neural networks with discontinuous activations," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 40, no. 6, pp. 1425–1437, Dec. 2010.
- [36] Y. Wang and L. Huang, "Global stability analysis of competitive neural networks with mixed time-varying delays and discontinuous neuron activations," *Neurocomputing*, vol. 152, pp. 85–96, Mar. 2015.
- [37] Z. Cai, L. Huang, Z. Guo, L. Zhang, and X. Wan, "Periodic synchronization control of discontinuous delayed networks by using extended Filippov-framework," *Neural Netw.*, vol. 68, pp. 96–110, Aug. 2015.
- [38] X. Nie and W. X. Zheng, "Dynamical behaviors of multiple equilibria in competitive neural networks with discontinuous nonmonotonic piecewise linear activation functions," *IEEE Trans. Cybern.*, Doi: 10.1109/TCYB.2015.2413212.
- [39] A.-M. Zou, K. D. Kumar, Z.-G. Hou, and X. Liu, "Finite-time attitude tracking control for spacecraft using terminal sliding mode and Chebyshev neural network," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 950–963, Aug. 2011.
- [40] M. Di Marco, M. Forti, M. Grazzini, P. Nistri, and L. Pancioni, "Lyapunov method and convergence of the full-range model of CNNs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 11, pp. 3528–3541, Dec. 2008.
- [41] J. P. Aubin and A. Cellina, *Differential Inclusions. Set-Valued Maps and Viability Theory*. Berlin, Germany: Springer, 1984.
- [42] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Part 1: Fundamentals*. Berlin, Germany: Springer, 1996.
- [43] M. P. Kennedy and L. O. Chua, "Neural networks for nonlinear programming," *IEEE Trans. Circuits Syst.*, vol. 35, no. 5, pp. 554–562, May 1988.
- [44] O. Faydasicok and S. Arik, "A new upper bound for the norm of interval matrices with application to robust stability analysis of delayed neural networks," *Neural Netw.*, vol. 44, pp. 64–71, Aug. 2013.
- [45] X. Liao and K.-W. Wong, "Robust stability of interval bidirectional associative memory neural network with time delays," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 34, no. 2, pp. 1142–1154, Apr. 2004.
- [46] M. Di Marco, M. Grazzini, and L. Pancioni, "Global robust stability criteria for interval delayed full-range cellular neural networks," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 666–671, Apr. 2011.
- [47] K. Wang and A. N. Michel, "Robustness and perturbation analysis of a class of nonlinear systems with applications to neural networks," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 41, no. 1, pp. 24–32, Jan. 1994.
- [48] M. Di Marco, M. Forti, and M. Grazzini, "Robustness of convergence in finite time for linear programming neural networks," *Int. J. Circuit Theory Appl.*, vol. 34, no. 3, pp. 307–316, May/June 2006.
- [49] M. Di Marco, M. Forti, and A. Tesi, "Bifurcations and oscillatory behavior in a class of competitive cellular neural networks," *Int. J. Bifurcat. Chaos*, vol. 10, no. 6, pp. 1267–1293, Jun. 2000.
- [50] A. Bhaya and E. Kaszkurewicz, *Control Perspectives on Numerical Algorithms and Matrix Problems*, vol. 10. Philadelphia, PA, USA: SIAM, 2006.
- [51] M. Di Marco, M. Forti, P. Nistri, and L. Pancioni, "Nonsmooth neural network for convex time-dependent constraint satisfaction problems," *IEEE Trans. Neural Netw. Learn. Syst.*, Doi: 10.1109/TNNLS.2015.2404773.