Published in final edited form as:

IEEE Trans Neural Netw. 2008 February ; 19(2): 273–283. doi:10.1109/TNN.2007.905840.

Constrained Least Absolute Deviation Neural Networks

Zhishun Wang [Senior Member, IEEE] and

The MRI Unit & The Division of Child Psychiatry, Columbia College of Physicians & Surgeons & the New York State Psychiatric Institute, New York, NY 10032 USA and also with the Brain Image Laboratory, Columbia University and NYSPI, New York, NY 10032 USA (e-mail: zw2105@columbia.edu)

Bradley S. Peterson

The MRI Unit & The Division of Child Psychiatry, Columbia College of Physicians & Surgeons & the New York State Psychiatric Institute, New York, NY 10032 USA

Abstract

It is well known that least absolute deviation (LAD) criterion or L_1 -norm used for estimation of parameters is characterized by robustness, i.e., the estimated parameters are totally resistant (insensitive) to large changes in the sampled data. This is an extremely useful feature, especially, when the sampled data are known to be contaminated by occasionally occurring outliers or by spiky noise. In our previous works, we have proposed the least absolute deviation neural network (LADNN) to solve unconstrained LAD problems. The theoretical proofs and numerical simulations have shown that the LADNN is Lyapunov-stable and it can globally converge to the exact solution to a given unconstrained LAD problem. We have also demonstrated its excellent application value in time-delay estimation. More generally, a practical LAD application problem may contain some linear constraints, such as a set of equalities and/or inequalities, which is called constrained LAD problem, whereas the unconstrained LAD can be considered as a special form of the constrained LAD. In this paper, we present a new neural network called constrained least absolute deviation neural network (CLADNN) to solve general constrained LAD problems. Theoretical proofs and numerical simulations demonstrate that the proposed CLADNN is Lyapunov stable and globally converges to the exact solution to a given constrained LAD problem, independent of initial values. The numerical simulations have also illustrated that the proposed CLADNN can be used to robustly estimate parameters for nonlinear curve fitting, which is extensively used in signal and image processing.

Index Terms

Equality and inequality constraints; least absolute deviation (LAD); L_1 -norm optimization; neural network (NN); nonlinear curve fitting

I. Introduction

The use of the least absolute deviation (LAD) or L_1 -norm criterion provides in many applications in signal and image processing a useful and reliable alternative to the least squares (LS) criterion (L_2 -norm) or Chebyshev criterion (L_∞ -norm). These applications span the field of deconvolution, state estimation, blind source separation, system identification, parameter estimation, time-frequency analysis, and time-delay estimation [1]–[13]. The L_1 -norm solutions of systems of linear algebraic equations have certain properties not shared by the ordinary LS (L_2 -norm) solutions [1]–[13] as follows. 1) An L_1 -norm solution of an overdetermined system of linear equations always exists although the L_1 -norm solution is not necessarily unique in contrast to the L_2 -norm solution where the solution is always unique when matrix A is of full rank. 2) L_1 -norm solutions are robust which means the solution is

resistant (insensitive) to some large changes in the data. This is an extremely useful property when the data are known to be contaminated by occasional wild points or spiky noise. 3) For fitting a number of data points by a constant, the L_1 -norm estimate can be interpreted as the median while the L_2 -norm estimate is the mean. A general LAD with equality and inequality constraints are formulated as follows:

$$\begin{cases} \text{Minimize} & \|Ax - b\|_1 \\ \text{Subject to} & C^T x = d, E^T x \leq f \end{cases} \quad (1)$$

where $A = [a_{ij}] \in \mathbf{R}^{M \times N}$, $b \in \mathbf{R}^M$, $C = [c_{ij}] \in \mathbf{R}^{N \times P}$, $d \in \mathbf{R}^P$, $E = [e_{ij}] \in \mathbf{R}^{N \times Q}$, $f \in \mathbf{R}^Q$, and $x \in \mathbf{R}^N$ is the solution to be determined.

Because of its excellent properties, the LAD optimization model described by (1) has been extensively studied [1]-[10] and many numerical algorithms for solving this model, such as Simplex and Karmarkar's algorithms, have been proposed [7]-[9]. However, with the increase of the model scale, these numerical algorithms are not adequate for solving real-time problems. One possible and promising approach to real-time optimization is to apply neural networks (NNs). Because of the inherent massive parallelism, the NN-based approach can solve optimization problems within a time constant of the network, for example, several milliseconds. Hopfield and Tank first proposed to use an NN for solving linear programming problems [14], [15]. Their creative work has motivated many researchers to investigate alternative NNs for solving linear and nonlinear programming problems [1], [2], [11]-[13], [16]-[22] and for solving arbitrary roots of polynomials [23], [24]. For instance, Cichocki and Unbehauen proposed to use an NN for solving the unconstrained LAD problem [1]. In the NN proposed by Cichocki and Unbehauen for implementing the LAD optimization model, some complex control circuits, such as "inhibition control circuits" and "switch control circuits," were introduced [1]. This made their NN models complicated and difficult for implementation. For some so-called ill-conditioned problems, their network parameters have to be adjusted, which is unrealistic for practical applications since the exact solutions are unknown. In our previous work, we presented a new NN to solve the unconstrained LAD optimization model [11], [12]. This NN can be implemented by using a number of adders, limiters, and integrators, without any control circuits to be tuned. By mathematical proofs and experimental validations, the proposed unconstrained LAD network is Lyapunov stable and able to converge globally to the exact solution to a given problem. In this paper, we propose another NN to solve the constrained LAD optimization problem, which can be viewed as the generalization of the unconstrained LAD because when the constraints are set to zeros, the constrained LAD model described by (1) is equivalent to an unconstrained LAD model. Likewise, the NN for solving constrained LAD problems, namely, CLADNN, can be implemented without any control circuits and its convergence is proven theoretically and validated experimentally.

The rest of this paper is organized as follows. In Section II, preliminary theories are given, which will be cited in later sessions. In Section III, the LAD NN with equality constraints is presented first and then it is extended to the general case where the equality and inequality constraints exist simultaneously. Section IV presents three groups of numerical experiments for the validation of the convergence performances of the proposed CLADNN. Finally, the conclusions are given in Section V.

II. Preliminary Theories and Definitions

In this section, the necessary mathematical definitions and lemmas are introduced, which will be used to describe the proposed LAD NN in Section II-A.

Definition 1 (Lipschitz Continuous [21], [25])

The mapping $F : \Omega \subset \mathbf{R}^N$ is said to be Lipschitz continuous with a positive constant L on the set Ω , if for each pair of points $\mathbf{x}, \mathbf{y} \in \Omega$, the following inequality holds:

$$\|F(\mathbf{x}) - F(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2$$

Definition 2 (Lyapunov Stable [21], [26]-[29])

Let $\mathbf{x}(t)$ be a solution of the dynamic system $\dot{\mathbf{x}} = f(\mathbf{x})$ ($(d\mathbf{x}/dt) = f(\mathbf{x})$). The equilibrium point $\mathbf{x}^*$, that is $f(\mathbf{x}^*) = 0$, is defined to be Lyapunov stable if, for any $\mathbf{x}_0 = \mathbf{x}(t_0)$ and scalar $\varepsilon > 0$, there exists $\delta > 0$ such that if $\|\mathbf{x}(t_0) - \mathbf{x}^*\|_2 < \delta$, then $\|\mathbf{x}(t) - \mathbf{x}^*\|_2 < \varepsilon$ for all $t > t_0$.

Definition 3 (Global Convergence [21], [26]-[29])

An NN is said to be globally convergent if for any initial point taken in the definition domain, every trajectory of the corresponding dynamic system converges to its unique equilibrium point.

Lemma 1 (Extended Brouwer's Fixed Theorem [30, Th. 2.5])

Let $\Omega \subset \mathbf{R}^N$ be compact and convex and let $F : \Omega \rightarrow \Omega$ be continuous. Then, F admits a fixed point.

Lemma 2

If $f(\mathbf{x})$ and $g(\mathbf{x})$ are Lipschitz continuous on $\Omega \subset \mathbf{R}^N$, then, for any constant k_1 and k_2 , $k_1 f(\mathbf{x}) + k_2 g(\mathbf{x})$ is also Lipschitz continuous on $\Omega \subset \mathbf{R}^N$.

Proof—See [11].

Lemma 3 (LaSalle's Theorem [28, Th. VI])

Let $V(\mathbf{x})$ be a scalar function with continuous first partial derivations. Let Ω_l designate the region where $V(\mathbf{x}) < l$. Assume that Ω_l is bounded and that within Ω_l

$$V(\mathbf{x}) < 0 \text{ for } \mathbf{x} \neq 0 \quad (2.a)$$

$$\dot{V}(\mathbf{x}) = \frac{dV(\mathbf{x})}{dt} \leq 0. \quad (2.b)$$

Let R be the set of all points within Ω_l where $\dot{V}(\mathbf{x}) = 0$ and let M be the largest invariant subset in R . Then, every solution $\mathbf{x}(t)$ in Ω_l tends to M as $t \rightarrow +\infty$.

III. Theories of CLADNNs**A. Theories of an NN of the LAD With Equality Constraints**

We first construct an NN to solve the LAD problem only with equality constraints and prove its global convergence. This LAD problem is formulated as

$$\begin{cases} \text{Minimize} & \|\mathbf{Ax}-\mathbf{b}\|_1 \\ \text{Subject to} & \mathbf{C}^T \mathbf{x}=\mathbf{d} \end{cases} \quad (3)$$

Using Proposition 1, we turn the problem described in (3) into another form, which is easier to solve.

Proposition 1—The optimization model described in (3) is equivalent to the following optimization model:

$$\begin{cases} \text{Minimize} & \{\max_{\mathbf{y} \in \Omega} \{\mathbf{y}^T (\mathbf{Ax}-\mathbf{b})\}\} \\ \text{Subject to} & \mathbf{C}^T \mathbf{x}=\mathbf{d} \end{cases} \quad (4)$$

where $\Omega = \{\omega \in \mathbf{R}^M \mid \|\omega_i\| \leq 1, i = 1, 2, \dots, M\} \subset \mathbf{R}^M$.

Proof: For (4), we only need to prove $\|\mathbf{Ax}-\mathbf{b}\|_1 = \max_{\mathbf{y} \in \Omega} \{\mathbf{y}^T (\mathbf{Ax}-\mathbf{b})\}$. Let $\mathbf{u} = (\mathbf{Ax}-\mathbf{b}) \in \mathbf{R}^M$; then, for any $\mathbf{y} \in \Omega$, and $i = 1, 2, \dots, M$, we have $|y_i| \leq 1$, thus

$$\mathbf{y}^T (\mathbf{Ax}-\mathbf{b}) = \mathbf{y}^T \mathbf{u} = \sum_{i=1}^M y_i u_i \leq \sum_{i=1}^M |y_i| \cdot |u_i| \leq \sum_{i=1}^M |u_i| = \|\mathbf{Ax}-\mathbf{b}\|_1.$$

Specifically, let $\mathbf{y}^0 = [\text{sign}(u_1), \dots, \text{sign}(u_M)]^T$; then, $\mathbf{y}^0 \in \Omega$ and

$$(\mathbf{y}^0)^T (\mathbf{Ax}-\mathbf{b}) = \sum_{i=1}^M \text{sign}(u_i) u_i = \sum_{i=1}^M |u_i| = \|\mathbf{Ax}-\mathbf{b}\|_1.$$

Thus

$$\max_{\mathbf{y} \in \Omega} \mathbf{y}^T (\mathbf{Ax}-\mathbf{b}) = \|\mathbf{Ax}-\mathbf{b}\|_1.$$

This completes the proof of Proposition 1.

Now, we propose an NN for solving the problem in (4) whose model is described by the following dynamic system:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = -\mathbf{C}(\mathbf{C}^T \mathbf{x} - \mathbf{d} - \mathbf{z}) - \mathbf{A}^T \mathbf{P}_\Omega(\mathbf{y} + \mathbf{Ax} - \mathbf{b}) \\ \frac{d\mathbf{y}}{dt} = -(\mathbf{AA}^T + \mathbf{I})\mathbf{y} + \mathbf{P}_\Omega(\mathbf{y} + \mathbf{Ax} - \mathbf{b}) + \mathbf{ACz} \\ \frac{d\mathbf{z}}{dt} = \mathbf{d} - \mathbf{C}^T [\mathbf{x} - \mathbf{A}^T \mathbf{y} + \mathbf{Cz}] \end{cases} \quad (5)$$

where $\mathbf{x} \in \mathbf{R}^N$, $\mathbf{y} \in \mathbf{R}^M$, $\mathbf{z} \in \mathbf{R}^P$, $\Omega = \{\omega \in \mathbf{R}^M \mid \|\omega_i\| \leq 1, i = 1, 2, \dots, M\} \subset \mathbf{R}^M$, $\mathbf{A} = [a_{ij}] \in \mathbf{R}^{M \times N}$, $\mathbf{C} = [c_{ij}] \in \mathbf{R}^{N \times P}$, $\mathbf{b} \in \mathbf{R}^M$, and $\mathbf{d} \in \mathbf{R}^P$. $\mathbf{I}$ is a unit matrix of $\mathbf{R}^M \times M$. For any given

vector $\xi = [\xi_1, \xi_2, \dots, \xi_M]^T \in \mathbf{R}^M$, $\mathbf{P}_\Omega(\xi) = [P_\Omega(\xi_1), P_\Omega(\xi_2), \dots, P_\Omega(\xi_M)]^T \in \Omega$, where $P_\Omega(\cdot)$ is a projection operator and for $i = 1, 2, \dots, M$, $P_\Omega(\xi_i)$ is defined as

$$P_\Omega(\xi_i) = \frac{1}{2}(|\xi_i + 1| - |\xi_i - 1|) = \begin{cases} 1, & \text{if } \xi_i > 1 \\ \xi_i, & \text{otherwise.} \\ -1, & \text{if } \xi_i < -1 \end{cases} \quad (6)$$

Since the NN described in (5) is a continuous-time network governed by a set of ordinary differential equations, it can be real-time implemented via very large scale integration (VLSI) complimentary metal–oxide–semiconductor (CMOS) circuits. In such a circuit implementation, the projection operator of $P_\Omega(\cdot)$ is actually a simple limiter with a unit threshold. The matrix or vector multiplications are actually the synaptic weighting and summing operations, and hence, they can be implemented via a number of adders with a weighting function [1], [2], [11]–[13], [20]–[22], [31]–[33]. The rest are a number of simple integrators. The most important thing is that the architecture of the proposed network, as shown in Fig. 1, does not contain any control unit circuits or variable parameters to be tuned, which are usually found in the literature, e.g., in [1], [16], and [17], and thus its implementation becomes much easier. To conduct numerical validations, Fig. 2(a) illustrates the simulation circuit of the proposed network under MATLAB SIMULINK. In the following, we will prove that the NN described in (5) globally converges to the exact solution to problem (4), or equivalently, to problem (3).

Theorem 1—Suppose that the problem in (4) has a solution, then the NN described by (5) is Lyapunov stable and globally converges to the exact solution to problem (4).

Theorem 1 is proven as follows. First, we begin with the proofs of a number of propositions.

Proposition 2—Let $\Omega \subset \mathbf{R}^N$ denote a closed convex set and let $\mathbf{v} = \mathbf{P}_\Omega(\mathbf{u})$ denote the projection of $\mathbf{u}$ on Ω , i.e., for each $\mathbf{u} \in \mathbf{R}^N$, there is a unique $\mathbf{v} \in \Omega$ such that

$$\|\mathbf{u} - \mathbf{v}\|_2 = \|\mathbf{u} - \mathbf{P}_\Omega(\mathbf{u})\|_2 = \inf_{\mathbf{u}^* \in \Omega} \|\mathbf{u} - \mathbf{u}^*\|_2$$

if and only if for all $\mathbf{u}^* \in \Omega$, $(\mathbf{u} - \mathbf{P}_\Omega(\mathbf{u}))^T (\mathbf{P}_\Omega(\mathbf{u}) - \mathbf{u}^*) \geq 0$.

Proof: See [11].

Proposition 3—Let $\Omega \subset \mathbf{R}^N$ denote a closed convex set and let $\mathbf{v} = \mathbf{P}_\Omega(\mathbf{u}) \in \Omega$ denote the projection of $\mathbf{u} \in \mathbf{R}^N$ on Ω . Then, the projection operator $\mathbf{P}_\Omega(\cdot)$ satisfies

$$\|\mathbf{P}_\Omega(\mathbf{u}) - \mathbf{P}_\Omega(\bar{\mathbf{u}})\|_2 \leq \|\mathbf{u} - \bar{\mathbf{u}}\|_2 \text{ for } \mathbf{u}, \bar{\mathbf{u}} \in \mathbf{R}^N.$$

That is, $\mathbf{P}_\Omega(\cdot)$ is Lipschitz continuous.

Proof: See [11].

Proposition 4—Let $\Omega \subset \mathbf{R}^N$ be compact and convex, $F : \Omega \subset \mathbf{R}^N \rightarrow \mathbf{R}^N$ be a continuous mapping, and $\mathbf{P}_\Omega(\zeta)$ be the projection of $\zeta \in \mathbf{R}^N$ on Ω . Then, there exists a $\mathbf{u}^* \in \Omega$ so that $(\mathbf{u} - \mathbf{u}^*)^T F(\mathbf{u}^*) \geq 0$ for all $\mathbf{u} \in \Omega$ if and only if $\mathbf{u}^* = \mathbf{P}_\Omega(\mathbf{u}^* - F(\mathbf{u}^*))$.

Proof: See [11].

Proposition 5—If $\mathbf{x}^* \in \mathbf{R}^N$ is a solution to the problem in (4), then there exist $\mathbf{y}^* \in \Omega$ and $\mathbf{z}^* \in \mathbf{R}^P$ such that $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$ satisfies

$$\begin{cases} \mathbf{C}^T \mathbf{x}^* = \mathbf{d} \\ \mathbf{y}^* = \mathbf{P}_\Omega(\mathbf{y}^* + \mathbf{A} \mathbf{x}^* - \mathbf{b}) \\ \mathbf{C} \mathbf{z}^* = \mathbf{A}^T \mathbf{y}^* \end{cases} \quad (7)$$

where matrices $\mathbf{A}$ and $\mathbf{C}$, vectors $\mathbf{b}$ and $\mathbf{d}$, compact and convex set Ω , and projection operator $\mathbf{P}_\Omega(\cdot)$ have identical definitions as given in (6).

Proof: Let $L(\mathbf{x}, \mathbf{y}, \mathbf{z})$ denote the Lagrange function of Max–Min optimization model in (4); then

$$L(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{y}^T (\mathbf{A} \mathbf{x} - \mathbf{b}) + \mathbf{z}^T (\mathbf{d} - \mathbf{C}^T \mathbf{x}). \quad (8)$$

According to the Kuhn–Tucker (K–T) theorem [2], [26], we know that $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$ is a solution to convex optimization model (4) if and only if there exists a saddle point of model (4), $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \in \mathbf{R}^N \times \Omega \times \mathbf{R}^P$ which satisfies

$$L(\mathbf{x}^*, \mathbf{y}, \mathbf{z}^*) \leq L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \leq L(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) \quad \forall \mathbf{x} \in \mathbf{R}^N \quad \forall \mathbf{z} \in \mathbf{R}^P, \mathbf{y} \in \Omega \quad (9a)$$

$$L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}) \leq L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \leq L(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*) \quad \forall \mathbf{x} \in \mathbf{R}^N \quad \forall \mathbf{z} \in \mathbf{R}^P, \mathbf{y} \in \Omega. \quad (9b)$$

$$L(\mathbf{x}^*, \mathbf{y}, \mathbf{z}^*) \leq L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \text{ yields } \mathbf{y}^T (\mathbf{A} \mathbf{x}^* - \mathbf{b}) \leq (\mathbf{y}^*)^T (\mathbf{A} \mathbf{x}^* - \mathbf{b}) \quad (10)$$

which is equivalent to

$$(\mathbf{y} - \mathbf{y}^*)^T (\mathbf{b} - \mathbf{A} \mathbf{x}^*) \geq 0 \quad \forall \mathbf{y} \in \Omega. \quad (11)$$

In Proposition 4, let $\mathbf{u} = \mathbf{y}$, $\mathbf{u}^* = \mathbf{y}^*$, and $F(\mathbf{u}^*) = \mathbf{b} - \mathbf{A} \mathbf{x}^*$; we immediately obtain

$$\mathbf{y}^* = \mathbf{P}_\Omega(\mathbf{y}^* + \mathbf{A} \mathbf{x}^* - \mathbf{b}). \quad (12)$$

$L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}) \leq L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$ yields

$$\mathbf{z}^T(\mathbf{d}-\mathbf{C}^T\mathbf{x}^*) \leq (\mathbf{z}^*)^T(\mathbf{d}-\mathbf{C}^T\mathbf{x}^*). \quad (13)$$

Utilizing the fact that inequality (13) must hold for all $\mathbf{z} \in \mathbf{R}^P$, we obtain

$$\mathbf{C}^T\mathbf{x}^*=\mathbf{d}. \quad (14)$$

$L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \leq L(\mathbf{x}, \mathbf{y}^*, \mathbf{z}^*)$ yields

$$(\mathbf{A}^T\mathbf{y}^*-\mathbf{C}\mathbf{z}^*)^T\mathbf{x}^* \leq (\mathbf{A}^T\mathbf{y}^*-\mathbf{C}\mathbf{z}^*)^T\mathbf{x}. \quad (15)$$

Utilizing the fact that inequality (15) must hold for all $\mathbf{x} \in \mathbf{R}^N$, we obtain

$$\mathbf{C}\mathbf{z}^*=\mathbf{A}^T\mathbf{y}^*. \quad (16)$$

Conversely, if $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \in \mathbf{R}^N \times \Omega \times \mathbf{R}^P$ satisfies (7), then (9a) and (9b) hold for all $\mathbf{x} \in \mathbf{R}^N$, and $\mathbf{y} \in \Omega$ and $\mathbf{z} \in \mathbf{R}^P$. This completes the proof of Proposition 5.

Corollary: For any $\mathbf{y} \in \mathbf{R}^M$ and $P_\Omega(\xi) = 1/2 (|\xi + 1| - |\xi - 1|)$, the following inequality holds:

$$(\mathbf{P}_\Omega(\mathbf{y}+\mathbf{A}\mathbf{x}-\mathbf{b})-\mathbf{y}^*)^T(\mathbf{b}-\mathbf{A}\mathbf{x}^*) \geq 0. \quad (17)$$

Proof: Let $\mathbf{u} = (\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}) \in \mathbf{R}^M$, then $\mathbf{P}_\Omega(\mathbf{u}) \in \Omega$, i.e., $|P_\Omega(u_i)| \leq 1$ for $i = 1, 2, \dots, M$. Since (11) holds for all $\mathbf{y} \in \Omega$ ($|y_i| \leq 1, i = 1, 2, \dots, M$), thus (17) obviously holds. [Note: (17) holds for all $\mathbf{y} \in \mathbf{R}^M$, whereas (11) holds only for all $\mathbf{y} \in \Omega$.]

For convenience of proving Theorem 1, we rewrite dynamic system (5) into the following form:

$$\left(\frac{d\mathbf{x}}{dt} \frac{d\mathbf{y}}{dt} \frac{d\mathbf{z}}{dt} \right)^T = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$$

where

$$\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \begin{pmatrix} \mathbf{C}(\mathbf{C}^T\mathbf{x}-\mathbf{d}-\mathbf{z})+\mathbf{A}^T\mathbf{P}_\Omega(\mathbf{y}+\mathbf{A}\mathbf{x}-\mathbf{b}) \\ (\mathbf{A}\mathbf{A}^T+\mathbf{I})\mathbf{y}-\mathbf{P}_\Omega(\mathbf{y}+\mathbf{A}\mathbf{x}-\mathbf{b})-\mathbf{A}\mathbf{C}\mathbf{z} \\ \mathbf{C}^T[\mathbf{x}-\mathbf{A}^T\mathbf{y}+\mathbf{C}\mathbf{z}]-\mathbf{d} \end{pmatrix}. \quad (18)$$

The solution set of $\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{0}$ is just the equilibrium point set of dynamic system (5). In Theorem 2, we will give the relationship between the solution set of model (4) and the equilibrium point set of dynamic system (5).

Theorem 2—Let $\mathfrak{R}^* = \{(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \in \mathbf{R}^N \times \Omega \times \mathbf{R}^P\}$ denote the solution set of model (4) and $\mathfrak{R}^0 = \{(\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0) \in \mathbf{R}^N \times \Omega \times \mathbf{R}^P\}$ denote the equilibrium point set of dynamic system (5), i.e., the solution set of $\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{0}$, then $\mathfrak{R}^* = \mathfrak{R}^0$.

Proof: Let $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \in \mathfrak{R}^*$, by Proposition 5

$$\begin{cases} \mathbf{C}^T \mathbf{x}^* = \mathbf{d} \\ \mathbf{y}^* = \mathbf{P}_\Omega(\mathbf{y}^* + \mathbf{A}\mathbf{x}^* - \mathbf{b}) \\ \mathbf{C}\mathbf{z}^* = \mathbf{A}^T \mathbf{y}^* \end{cases}$$

thus, $\mathbf{F}(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) = \mathbf{0}$. On the other hand, let $\mathbf{E}_1 = (\mathbf{C}^T \mathbf{x} - \mathbf{d}) \in \mathbf{R}^P$, $\mathbf{E}_2 = [\mathbf{y} - \mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b})] \in \mathbf{R}^M$, and $\mathbf{E}_3 = [\mathbf{A}^T \mathbf{y} - \mathbf{C}\mathbf{z}] \in \mathbf{R}^N$, then

$$\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \begin{pmatrix} \mathbf{C}\mathbf{E}_1 - \mathbf{A}^T \mathbf{E}_2 + \mathbf{E}_3 \\ \mathbf{E}_2 + \mathbf{A}\mathbf{E}_3 \\ \mathbf{E}_1 - \mathbf{C}^T \mathbf{E}_3 \end{pmatrix}. \quad (19)$$

If there exists $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0) \in \mathfrak{R}^0$ so that $\mathbf{F}(\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0) = \mathbf{0}$, then

$$\begin{aligned} \mathbf{C}\mathbf{E}_1 - \mathbf{A}^T \mathbf{E}_2 + \mathbf{E}_3 &= \mathbf{0} \\ \mathbf{E}_2 + \mathbf{A}\mathbf{E}_3 &= \mathbf{0} \\ \mathbf{E}_1 - \mathbf{C}^T \mathbf{E}_3 &= \mathbf{0}. \end{aligned} \quad (20)$$

According to (20), we have $(\mathbf{A}^T \mathbf{A} + \mathbf{C}\mathbf{C}^T + \mathbf{I})\mathbf{E}_3 = \mathbf{0}$, where $\mathbf{I}$ is an $\mathbf{R}^{N \times N}$ unit matrix. Since $(\mathbf{A}^T \mathbf{A} + \mathbf{C}\mathbf{C}^T + \mathbf{I})$ is positive definite, thus $\mathbf{E}_3 = \mathbf{0}$ and, from (20), we also have $\mathbf{E}_1 = \mathbf{0}$ and $\mathbf{E}_2 = \mathbf{0}$. That is, $\mathbf{C}^T \mathbf{x}^0 = \mathbf{d}$, $\mathbf{y}^0 = \mathbf{P}_\Omega(\mathbf{y}^0 + \mathbf{A}\mathbf{x}^0 - \mathbf{b})$, and $\mathbf{C}\mathbf{z}^0 = \mathbf{A}^T \mathbf{y}^0$. Based on Proposition 5, $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{z}^0)$ is the solution to model (4). This completes the proof of Theorem 2. Theorem 2 guarantees that if dynamic system (5) converges, it will converge to the exact solution as model (4). In the following, we need to further prove that system (5) does converge.

Proposition 6—Suppose that model (4) has a solution $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) \in \mathbf{R}^N \times \Omega \times \mathbf{R}^P$, which satisfies (7). Then, for all $\mathbf{x} \in \mathbf{R}^N$, $\mathbf{y} \in \mathbf{R}^M$, and $\mathbf{z} \in \mathbf{R}^P$, we have

$$\begin{pmatrix} \mathbf{x} - \mathbf{x}^* \\ \mathbf{y} - \mathbf{y}^* \\ \mathbf{z} - \mathbf{z}^* \end{pmatrix}^T \mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \geq \|\mathbf{E}_1\|_2^2 + \|\mathbf{E}_2\|_2^2 + \|\mathbf{E}_3\|_2^2 \quad (21)$$

where $\|\cdot\|_2$ denotes L_2 -norm. The definitions of $\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$, $\mathbf{E}_1$, $\mathbf{E}_2$, and $\mathbf{E}_3$ are the same as given previously.

Proof: Let $\mathbf{u}^* = \mathbf{y}^* \in \Omega$ and $\mathbf{u} = (\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}) \in \mathbf{R}^M$. By Proposition 2, we have

$$((\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}) - \mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}))^T \times (\mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}) - \mathbf{y}^*) \geq 0. \quad (22)$$

We rewrite the corollary in (17)

$$(\mathbf{P}_\Omega(\mathbf{y}+\mathbf{Ax}-\mathbf{b})-\mathbf{y}^*)^T(\mathbf{b}-\mathbf{Ax}^*) \geq 0. \quad (23)$$

By (22) and (23), we have

$$(\mathbf{y}+\mathbf{Ax}-\mathbf{Ax}^*-\mathbf{P}_\Omega(\mathbf{y}+\mathbf{Ax}-\mathbf{b}))^T \times (\mathbf{P}_\Omega(\mathbf{y}+\mathbf{Ax}-\mathbf{b})-\mathbf{y}^*) \geq 0$$

or

$$(\mathbf{A}(\mathbf{x}-\mathbf{x}^*)+\mathbf{E}_2)^T((\mathbf{y}-\mathbf{y}^*)-\mathbf{E}_2) \geq 0. \quad (24a)$$

That is

$$(\mathbf{x}-\mathbf{x}^*)^T(\mathbf{A}^T\mathbf{y}-\mathbf{A}^T\mathbf{y}^*)+\mathbf{E}_2^T(\mathbf{y}-\mathbf{y}^*)-(\mathbf{x}-\mathbf{x}^*)^T\mathbf{A}^T\mathbf{E}_2 \geq \|\mathbf{E}_2\|_2^2. \quad (24b)$$

Since $\mathbf{A}^T\mathbf{y}^* = \mathbf{Cz}^*$, substituting $\mathbf{A}^T\mathbf{y} - \mathbf{A}^T\mathbf{y}^* = (\mathbf{A}^T\mathbf{y} - \mathbf{Cz}) + (\mathbf{Cz} - \mathbf{Cz}^*) = \mathbf{E}_3 + \mathbf{C}(\mathbf{z} - \mathbf{z}^*)$ into (24b), we have

$$(\mathbf{x}-\mathbf{x}^*)^T\mathbf{E}_3+(\mathbf{x}-\mathbf{x}^*)^T\mathbf{C}(\mathbf{z}-\mathbf{z}^*)+(\mathbf{y}-\mathbf{y}^*)^T\mathbf{E}_2-(\mathbf{x}-\mathbf{x}^*)^T\mathbf{A}^T\mathbf{E}_2 \geq \|\mathbf{E}_2\|_2^2 \quad (24c)$$

where $\|\mathbf{E}_2\|_2^2 = \mathbf{E}_2^T\mathbf{E}_2$ and $\|\cdot\|_2$ denotes L_2 norm.

By Proposition 5, $\mathbf{d} - \mathbf{C}^T\mathbf{x}^* = (\mathbf{d} - \mathbf{C}^T\mathbf{x}) + (\mathbf{C}^T\mathbf{x} - \mathbf{C}^T\mathbf{x}^*) = -\mathbf{E}_1 + \mathbf{C}^T(\mathbf{x} - \mathbf{x}^*) \equiv \mathbf{0}$, we can write the following identical equation:

$$[\mathbf{E}_1-(\mathbf{z}-\mathbf{z}^*)]^T[-\mathbf{E}_1+\mathbf{C}^T(\mathbf{x}-\mathbf{x}^*)] \equiv 0$$

or

$$(\mathbf{x}-\mathbf{x}^*)^T\mathbf{CE}_1+(\mathbf{z}-\mathbf{z}^*)^T\mathbf{E}_1-(\mathbf{x}-\mathbf{x}^*)^T\mathbf{C}(\mathbf{z}-\mathbf{z}^*) \equiv \|\mathbf{E}_1\|_2^2. \quad (25b)$$

Again, by Proposition 5, $\mathbf{A}^T\mathbf{y}^* - \mathbf{Cz}^* \equiv \mathbf{0}$ we have

$$\mathbf{A}^T\mathbf{y}^* - \mathbf{Cz}^* = (\mathbf{A}^T\mathbf{y} - \mathbf{Cz}) - \mathbf{A}^T(\mathbf{y} - \mathbf{y}^*) + \mathbf{C}(\mathbf{z} - \mathbf{z}^*) \equiv \mathbf{0}$$

or

$$\mathbf{E}_3 - \mathbf{A}^T(\mathbf{y} - \mathbf{y}^*) + \mathbf{C}(\mathbf{z} - \mathbf{z}^*) \equiv 0. \quad (26a)$$

Therefore

$$(-\mathbf{E}_3)^T [\mathbf{E}_3 - \mathbf{A}^T(\mathbf{y} - \mathbf{y}^*) + \mathbf{C}(\mathbf{z} - \mathbf{z}^*)] \equiv 0$$

or

$$(\mathbf{y} - \mathbf{y}^*)^T \mathbf{A} \mathbf{E}_3 - (\mathbf{z} - \mathbf{z}^*)^T \mathbf{C}^T \mathbf{E}_3 \equiv \|\mathbf{E}_3\|_2^2. \quad (26b)$$

By (24c), (25b), and (26b), we obtain

$$(\mathbf{x} - \mathbf{x}^*)^T (\mathbf{C} \mathbf{E}_1 - \mathbf{A}^T \mathbf{E}_2 + \mathbf{E}_3) + (\mathbf{y} - \mathbf{y}^*)^T (\mathbf{E}_2 + \mathbf{A} \mathbf{E}_3) + (\mathbf{z} - \mathbf{z}^*)^T (\mathbf{E}_1 - \mathbf{C}^T \mathbf{E}_3) \geq \|\mathbf{E}_1\|_2^2 + \|\mathbf{E}_2\|_2^2 + \|\mathbf{E}_3\|_2^2$$

or

$$\begin{pmatrix} \mathbf{x} - \mathbf{x}^* \\ \mathbf{y} - \mathbf{y}^* \\ \mathbf{z} - \mathbf{z}^* \end{pmatrix}^T \begin{pmatrix} \mathbf{C} \mathbf{E}_1 - \mathbf{A}^T \mathbf{E}_2 + \mathbf{E}_3 \\ \mathbf{E}_2 + \mathbf{A} \mathbf{E}_3 \\ \mathbf{E}_1 - \mathbf{C}^T \mathbf{E}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{x} - \mathbf{x}^* \\ \mathbf{y} - \mathbf{y}^* \\ \mathbf{z} - \mathbf{z}^* \end{pmatrix}^T \mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \geq \|\mathbf{E}_1\|_2^2 + \|\mathbf{E}_2\|_2^2 + \|\mathbf{E}_3\|_2^2.$$

This completes the proof of Proposition 6.

Now, Let us finish the final proof of Theorem 1.

Let $\mathbf{u}(t) = (\mathbf{x}(t), \mathbf{y}(t), \mathbf{z}(t))^T \in \mathbf{R}^N \times \mathbf{R}^M \times \mathbf{R}^P$. Since by Proposition 3, the projection function $\mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b})$ is Lipschitz continuous, then by Lemma 2, the function $\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is also Lipschitz continuous. Thus, system (5), i.e., $(d\mathbf{u}/dt) = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is Lipschitz continuous. From the existence theory of ordinary differential equations (ODEs) [25], $(d\mathbf{u}/dt) = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ has a unique and continuous solution $\mathbf{u}(t)$ on some interval $[t_0, +\infty)$ ($t_0 \geq 0$).

On the other hand let $\mathbf{F}(\mathbf{u}^*) = \mathbf{F}(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*) = \mathbf{0}$ be the equilibrium point of system $(d\mathbf{u}/dt) = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$, i.e., $(d\mathbf{u}/dt) = \mathbf{0}$ when $\mathbf{u}(t) = \mathbf{u}^*$. Let $V(\mathbf{u}) = \|\mathbf{u} - \mathbf{u}^*\|_2^2$ denote a scalar function. Then, $V(\mathbf{u}) > 0$ when $\mathbf{u} - \mathbf{u}^* \neq \mathbf{0}$. By Proposition 6, we have

$$\begin{aligned} \frac{dV}{dt} &= \frac{d}{dt} \|\mathbf{u}(t) - \mathbf{u}^*\|_2^2 = 2(\mathbf{u}(t) - \mathbf{u}^*)^T \frac{d\mathbf{u}}{dt} \\ &= -2(\mathbf{u}(t) - \mathbf{u}^*)^T \mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \\ &\leq -2(\|\mathbf{E}_1\|_2^2 + \|\mathbf{E}_2\|_2^2 + \|\mathbf{E}_3\|_2^2) \leq 0. \end{aligned}$$

That means

$$\|\mathbf{u}(t) - \mathbf{u}^*\|_2 \leq k \|\mathbf{u}(t_0) - \mathbf{u}^*\|_2 \quad \forall t \in [t_0, +\infty)$$

where k is a positive constant. Therefore, the solution $\mathbf{u}(t)$ is bounded on $[t_0, +\infty)$. By Definition 2, the dynamic system in (5) or $(d\mathbf{u}/dt) = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is Lyapunov stable at its equilibrium point. Moreover, let $\Omega_l = \{\mathbf{u} \in \mathbf{R}^N \times \mathbf{R}^M \times \mathbf{R}^P \mid V(\mathbf{u}(t)) \leq V(\mathbf{u}(t_0))\}$ denote a bounded region. Then, the dynamic system in (5) or $(d\mathbf{u}/dt) = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ satisfies Lemma 3. Therefore, the solution trajectory $\mathbf{u}(t)$ converges to the largest invariant subset of the following set:

$$\mathbf{R} = \left\{ \mathbf{u} \in \Omega_l \mid \frac{dV}{dt} = 0 \right\}$$

In fact, if $(dV/dt) = 0$, from $dV/dt \leq -2(\|\mathbf{E}_1\|_2^2 + \|\mathbf{E}_2\|_2^2 + \|\mathbf{E}_3\|_2^2) \leq 0$, we have $\mathbf{E}_1 = \mathbf{0}$, $\mathbf{E}_2 = \mathbf{0}$, and $\mathbf{E}_3 = \mathbf{0}$. Thus, by Theorem 2, $\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{0}$, i.e., $(d\mathbf{u}/dt) = \mathbf{0}$. Conversely, if $(d\mathbf{u}/dt) = \mathbf{0}$, then $(dV/dt) = 2(\mathbf{u}(t) - \mathbf{u}^*)^T (d\mathbf{u}/dt) = 0$. That is, $(dV/dt) = 0$, if and only if $(d\mathbf{u}/dt) = \mathbf{0}$. Therefore

$$\mathbf{R} = \left\{ \mathbf{u} \in \Omega_l \mid \frac{d\mathbf{u}}{dt} = \mathbf{0} \right\}.$$

In fact, from $(dV/dt) = 2(\mathbf{u}(t) - \mathbf{u}^*)^T d\mathbf{u}/dt$, we also have that $(dV/dt) = 0$ if and only if $\mathbf{u}(t) = \mathbf{u}^*$ ($\mathbf{u}(t) = \mathbf{u}^* \Rightarrow (dV/dt) = 0$; conversely, $(dV/dt) = 0 \Rightarrow \mathbf{u}(t) = \mathbf{u}^*$). Furthermore,

$V(\mathbf{u}(t)) = \|\mathbf{u}(t) - \mathbf{u}^*\|_2^2$ is continuous and bounded, thus limit $\lim_{t \rightarrow \infty} V(\mathbf{u}(t))$ exists. Since $dV(\mathbf{u}(t))/dt < 0$ for any $t \in [0, +\infty)$ and $\mathbf{u}(t) \neq \mathbf{u}^*$, and $dV(\mathbf{u}(t))/dt = 0$ only when $\mathbf{u}(t) = \mathbf{u}^*$, thus $V(\mathbf{u}(t)) = \|\mathbf{u}(t) - \mathbf{u}^*\|_2^2$ is always decreasing as $t \rightarrow +\infty$ except that $\mathbf{u}(t) = \mathbf{u}^*$. Whereas $V(\mathbf{u}(t)) \geq 0$ for any $t \in [0, +\infty)$, therefore

$$\lim_{t \rightarrow +\infty} V(\mathbf{u}(t)) = \lim_{t \rightarrow +\infty} \|\mathbf{u}(t) - \mathbf{u}^*\|_2^2$$

or

$$\lim_{t \rightarrow +\infty} \mathbf{u}(t) = \mathbf{u}^*.$$

This means that $(d\mathbf{u}/dt) = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z})$ or the system in (5) globally converges to its equilibrium point set. Finally, by Theorem 2, system (5) globally converges to the exact solution to the model in (4), or equivalently, to the original LAD problem in (3). The proof of Theorem 1 is complete.

B. Theories of an NN of the LAD With Equality and Inequality Constraints

Now, we extend the LAD NN governed by the dynamic equations in (5) to a more general form to solve the general constrained LAD problem described in (1). First, we can transfer the original problem in (1) into the following equivalent problem:

$$\begin{cases} \text{Minimize} & \left\{ \max_{\mathbf{y} \in \Omega} \{\mathbf{y}^T (\mathbf{A}\mathbf{x} - \mathbf{b})\} \right\} \\ \text{Subject to} & \mathbf{C}^T \mathbf{x} \leq \mathbf{d}, -\mathbf{C}^T \mathbf{x} \leq -\mathbf{d}; \mathbf{E}^T \mathbf{x} \leq \mathbf{f} \end{cases} \quad (27a)$$

In this convert, the fact that $\mathbf{C}^T \mathbf{x} = \mathbf{d}$ is equivalent to $\mathbf{C}^T \mathbf{x} \leq \mathbf{d}$ and $-\mathbf{C}^T \mathbf{x} \leq -\mathbf{d}$ is utilized. Let $\mathbf{G} = [\mathbf{C} \ -\mathbf{C} \ \mathbf{E}] \in \mathbf{R}^{N \times K}$ and $\mathbf{h} = [\mathbf{d} \ -\mathbf{d} \ \mathbf{f}]^T \in \mathbf{R}^K$, where $k = 2P + Q$ [see (27a)] can be rewritten as

$$\begin{cases} \text{Minimize} & \left\{ \max_{\mathbf{y} \in \Omega} \{\mathbf{y}^T (\mathbf{A}\mathbf{x} - \mathbf{b})\} \right\} \\ \text{Subject to} & \mathbf{G}^T \mathbf{x} \leq \mathbf{h} \end{cases} \quad (27b)$$

Equation (27b) is almost identical to (4) except that (27b) contains an additional inequality constraint. Temporarily ignoring the inequality constraint and replacing $\mathbf{C}$ and $\mathbf{d}$ in (19) with $\mathbf{G}$ and $\mathbf{h}$ defined previously, respectively, we immediately obtain

$$\begin{aligned} \left(\frac{d\mathbf{x}}{dt} \frac{d\mathbf{y}}{dt} \frac{d\mathbf{z}}{dt} \right)^T &= -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \\ &= - \begin{pmatrix} \mathbf{G}\mathbf{E}_1 - \mathbf{A}^T \mathbf{E}_2 + \mathbf{E}_3 \\ \mathbf{E}_2 + \mathbf{A}\mathbf{E}_3 \\ \mathbf{E}_1 - \mathbf{G}^T \mathbf{E}_3 \end{pmatrix} \end{aligned} \quad (28a)$$

where $\mathbf{E}_1 = [\mathbf{G}^T \mathbf{x} - \mathbf{h}] \in \mathbf{R}^K$, $\mathbf{E}_2 = [\mathbf{y} - \mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b})] \in \mathbf{R}^M$ and $\mathbf{E}_3 = [\mathbf{A}^T \mathbf{y} - \mathbf{G}\mathbf{z}] \in \mathbf{R}^N$, where $\mathbf{x} \in \mathbf{R}^N$, $\mathbf{y} \in \mathbf{R}^M$, and $\mathbf{z} \in \mathbf{R}^K$. To consider the inequality constraint, we replace $\mathbf{E}_1 = [\mathbf{G}^T \mathbf{x} - \mathbf{h}]$ in (28a) with $\mathbf{E}_1 = \max(0, \mathbf{G}^T \mathbf{x} - \mathbf{h})$ so that the state equation given in (28a) can converge to the equilibrium point (solution) under constraint $\mathbf{G}^T \mathbf{x} \leq \mathbf{h}$. After this modification, we can rewrite (28a) as (28b), shown at the bottom of the page, where $\mathbf{D} = (\mathbf{A}\mathbf{A}^T + \mathbf{I}) \in \mathbf{R}^{M \times M}$, $\mathbf{U} = \mathbf{A}\mathbf{G} \in \mathbf{R}^{M \times K}$, and $\mathbf{V} = \mathbf{G}^T \mathbf{G} \in \mathbf{R}^{K \times K}$. The NN governed by (28b) will converge to the exact solution as the problem in (27b) or its original form in (1). The architecture of NN (28b) is illustrated in Fig. 1.

$$\left(\frac{d\mathbf{x}}{dt} \frac{d\mathbf{y}}{dt} \frac{d\mathbf{z}}{dt} \right)^T = -\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = - \begin{pmatrix} \mathbf{G}[\max\{0, \mathbf{G}^T \mathbf{x} - \mathbf{h}\} - \mathbf{z}] + \mathbf{A}^T \mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}) \\ \mathbf{D}\mathbf{y} - \mathbf{P}_\Omega(\mathbf{y} + \mathbf{A}\mathbf{x} - \mathbf{b}) - \mathbf{U}\mathbf{z} \\ \max\{0, \mathbf{G}^T \mathbf{x} - \mathbf{h}\} - \mathbf{U}^T \mathbf{y} + \mathbf{V}\mathbf{z} \end{pmatrix} \quad (28b)$$

IV. Numerical Experimental Validations

A. Experiment 1: Linear Curve-Fitting Problem

Let us consider a constrained LAD curve-fitting problem: Find the parameters (a_0, a_1) of the straight line $y(x) = a_1 x + a_0$, which fit the data $\{x_i, y_i\} = \{(0, 1), (1, 2), (2, 4), (4, 5)\}$ subject to the equality $y(5) = 6$. To use CLADNN solving this problem, we formulate this problem as follows:

$$\begin{cases} \text{Minimize} & \|\mathbf{Ax}-\mathbf{b}\|_1 \\ \text{Subject to} & \mathbf{C}^T\mathbf{x}=\mathbf{d} \end{cases} \quad (29)$$

where

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 0 & 1 & 2 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix}^T \\ \mathbf{b} &= \begin{bmatrix} 1 & 2 & 4 & 5 \end{bmatrix}^T \\ \mathbf{x} &= \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} \\ \mathbf{C} &= \begin{bmatrix} 5 \\ 1 \end{bmatrix} \\ d &= 6. \end{aligned}$$

We implement the CLADNN with equality constraint under MATLAB SIMULINK, as illustrated in Fig. 2(a). After inputting $\mathbf{A}$, $\mathbf{b}$, $\mathbf{C}$, and $\mathbf{d}$ into this CLADNN simulator, we start simulation by clicking “START” in simulation menu and we get the solution convergence curve as shown in Fig. 2(b), which gives the solution of $\mathbf{x} = [1 \ 1]^T$ or $(a_0, a_1) = (1, 1)$.

B. Experiment 2: Ill-Conditioned Problem

Let us solve the following so-called ill-conditioned LAD problem:

$$\text{Minimize } \|\mathbf{Ax}-\mathbf{b}\|_1 \quad (30)$$

where

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1.1 & 4 & 7 & -3.1 & 5.5 \\ 2 & 5.1 & 8 & -4 & 6.9 \\ 3.1 & 6 & 8.9 & -5 & 8.1 \end{bmatrix}^T \\ \mathbf{b} &= \begin{bmatrix} -2 & 1.1 & 4.2 & -0.5 & 2.1 \end{bmatrix}^T. \end{aligned}$$

The condition number of the matrix $\mathbf{A}$ is approximately 200 and the exact solution to problem (30) is $\mathbf{x}^* = [2 \ 1 \ -2]^T$. In order to solve this ill-conditioned problem, Cichocki and Unbehauen [1] used the following modified cost function:

$$C(\mathbf{x}, \alpha) = \frac{1}{2} \left(\alpha \|\mathbf{x}\|_2^2 + \|\mathbf{Ax}-\mathbf{b}\|_2^2 \right)$$

instead of the original definition of LAD optimization model. However, just as they mentioned in [1], their solutions depend crucially on the selections of the so-called regularization parameter α in cost function $C(\mathbf{x}, \alpha)$, which is unrealistic in practical applications. In [11], by using unconstrained LAD NN, we got an approximate solution $\mathbf{x} \approx [1.9968 \ 1.0056 \ -2.0026]^T$, which was close to the exact solution. If some prior information is available in this

problem, for example, one element of the exact solution is known, which is often true in real-world problems, we can use the proposed CLADNN to solve this ill-conditioned problem. Assume that $x(1)$ is known and $x(1) = 2$. We can formulate this problem into the form as described in (29) where $\mathbf{C} = [1 \ 0 \ 0]^T$ and $d = 2$. Fig. 3 demonstrates that the CLADNN is able to converge to the exact solution $\mathbf{x} = [2 \ 1 \ -2]^T$ to the ill-conditioned problem independent of initial values. Fig. 3(a)–(c) shows the convergence process of solution components x_1 , x_2 , and x_3 , respectively, whereas Fig. 3(d) illustrates the convergence process of the energy function of the CLADNN with equality constraint, $E(\mathbf{x}, \mathbf{y}, \mathbf{z})$, all under different initial values. The energy function is given by

$$E(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \|\mathbf{Cz} - \mathbf{A}^T P_{\Omega}(\mathbf{y} + \mathbf{Ax} - \mathbf{b})\|_2^2 + \|\mathbf{C}^T \mathbf{x} - \mathbf{d}\|_2^2.$$

C. Experiment 3: Nonlinear Curve-Fitting Problem

Let us consider a constrained nonlinear LAD curve-fitting problem: Find the parameters of the combination of polynomial and exponential function, $y(x) = a_4 x^3 + a_3 x^2 + a_2 x + a_1 e^{x/2} + a_0$, which fits the data given in Table I, simultaneously satisfying the constraints $y(4.5) = -1.83$ and $y(5) \leq -3$. We can formulate this problem into the LAD problem with constraints of equality and inequality in form

$$\begin{cases} \text{Minimize} & \|\mathbf{Ax} - \mathbf{b}\|_1 \\ \text{Subject to} & \mathbf{C}^T \mathbf{x} = \mathbf{d}, \mathbf{E}^T \mathbf{x} \leq \mathbf{f}. \end{cases} \quad (31)$$

where the equation shown at the top of the next page, holds. We use the MATLAB SIMULINK simulator as shown in Fig. 2(a) to implement the CLADNN for solving the problem in (31), by replacing $\mathbf{C}$ and $\mathbf{d}$ in the network with $\mathbf{G}$ and $\mathbf{h}$, respectively, where $\mathbf{G} = [\mathbf{C} \ -\mathbf{C} \mathbf{E}]$ and $\mathbf{h} = [\mathbf{d} \ -\mathbf{d} \mathbf{f}]^T$, and adding a limiter to implement the operation: $\mathbf{E}_1 = \max(0, \mathbf{G}^T \mathbf{x} - \mathbf{h})$. By inputting the previous parameters into the simulator and starting simulations for 100 times with 100 different initial values, we get the solution converge curves as shown in Fig. 4(a), indicating that the solution is $\mathbf{x} = [1.01 \ -0.46 \ 10.09 \ -14.28 \ 5.28]^T$ and it is independent of initial values. Replacing a_4, a_3, a_2, a_1 , and a_0 with this solution in $y(x) = a_4 x^3 + a_3 x^2 + a_2 x + a_1 e^{x/2} + a_0$, we can draw the curve fitting as shown in Fig. 4(b) (solid line). By changing L_1 -norm into L_2 -norm in (31) and invoking MATLAB optimization routine, namely lsqmin, we obtain the constrained LS curve fitting, as shown in Fig. 4(b) (dashed-dotted line). It has been shown that due to outliers at points (1.5, -9) and (3, -3), LS fitting performance is degenerated severely, but the CLADNN fitting is not affected, indicating that CLADNN has a more robust performance than the constrained LS optimization.

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 0 & 0.125 & 1 & 3.375 & 8 & 15.625 & 27 & 42.875 & 64 & 91.125 & 125 \\ 0 & 0.25 & 1 & 2.25 & 4 & 6.25 & 9 & 12.25 & 16 & 20.25 & 25 \\ 0 & 0.5 & 1 & 1.5 & 2 & 2.5 & 3 & 3.5 & 4 & 4.5 & 5 \\ 1 & 1.284 & 1.649 & 2.117 & 2.718 & 3.49 & 4.48 & 5.75 & 7.389 & 9.488 & 12.183 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T \\ \mathbf{b} &= \begin{bmatrix} -9 & -8 & -7.6 & -9 & -7.1 & -6.4 & -3 & -3.8 & -2.4 & -1.8 & -3.1 \end{bmatrix}^T \quad d = -1.83 \quad f = -3 \\ \mathbf{C} &= \begin{bmatrix} 91.125 & 20.25 & 4.5 & 9.488 & 1 \end{bmatrix}^T \\ \mathbf{E} &= \begin{bmatrix} 125 & 25 & 5 & 12.183 & 1 \end{bmatrix}^T \end{aligned}$$

and

$$\mathbf{x} = \begin{bmatrix} a_4 & a_3 & a_2 & a_1 & a_0 \end{bmatrix}^T$$

V. Conclusion

A new NN, namely, CLADNN, to solve the constrained LAD problems has been proposed in this paper. It has been mathematically shown that the proposed CLADNN is Lyapunov stable and it is able to globally converge to the exact solution of a given constrained LAD problem if it does exist. We have conducted three numerical simulations using the MATLAB SIMULINK simulator of the CLADNN. It has been demonstrated that the CLADNN can be used to solve linear and nonlinear curve-fitting problems. In addition, the CLADNN is able to give an exact solution to some ill-conditioned unconstrained LAD problems if prior information is available serving as a certain constraint. This is a promising feature for real-world applications where the problems are often ill-conditioned, but some prior information may be available for the solution, for example, in typical spectra quantification of magnetic resonance spectroscopy (MRS) where the frequencies of some metabolisms may be known. As illustrated in solving the nonlinear curve-fitting problem, compared with conventional LS (L_2 -norm) optimization, the CLADNN is more robust in spiky noise environment, which is of great application potentials in modeling brain imaging data, such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) data since the scanning process often suffers severely from sudden movements of the head, especially, when the subjects are children, which make the final data contain a lot of spiky noise or outliers. Therefore, the applications of the proposed CLADNN to real-world problem deserve further exploration.

Acknowledgments

This work was supported by the National Institute of Mental Health (NIMH) under Grants MH59139, MH068318, and T32 MH 16434, the Thomas D. Klingenstein & Nancy D. Perlman Family Fund, and the Suzanne Crosby Murphy Endowment at Columbia University.

References

1. Cichocki A, Unbehauen R. Neural networks for solving systems of linear equations—Part II: Minimax and least absolute value problems. *IEEE Trans Circuit Syst Sep;1992 CAS-39(9):619–633.*
2. Cichocki, A.; Unbehauen, R. *Neural Networks for Computing, Optimization and Signal Processing.* Stuttgart, Germany: Teubner-Wiley; 1993.
3. Zala CA, Barrodale I, Kennedy JS. High-resolution signal and noise field estimation using the L_1 (least absolute values) norm. *IEEE J Ocean Eng Jan;1987 OE-12(1):253–264.*
4. Barrodale, I.; Zala, CA. L_1 and L_∞ curve fitting and linear programming algorithms and applications. In: Mohamed, JL.; Walsh, JE., editors. *Numerical Algorithms.* Vol. ch. 11. Oxford, U.K.: Oxford Univ. Press; 1986. p. 220-238.
5. Bandler JW, Kellerman W, Madsen K. A nonlinear L_1 optimization algorithm for design, modeling, and diagnosis of networks. *IEEE Trans Circuits Syst Feb;1987 CAS-34(1):174–181.*
6. Abdelmalek NN, Ostu N. Restoration of images with missing high-frequency components by minimizing the L_1 norm of the solution vector. *Appl Opt 1986;24:1415–1420.* [PubMed: 18223730]
7. Bloomfield, P.; Steiger, WL. *Least Absolute Deviations: Theory Applications and Algorithms.* Boston, MA: Birkhäuser; 1983.
8. Ward RK. An on-line adaptation for discrete L_1 linear estimation. *IEEE Trans Autom Control Jan; 1984 AC-29(1):67–71.*

9. Ruzinsky SA, Olsen ET. L_1 and L_∞ minimization via a variant of Karmarkar's algorithm. *IEEE Trans Acoust Speech Signal Process* Feb;1989 37(2):245–253.
10. Levy S, Walker C, Ulrych TJ, Fullagar PK. A linear programming approach to the estimation of the power spectra of harmonic processes. *IEEE Trans Acoust Speech Signal Process* Aug;1982 ASSP-30 (4):675–679.
11. Wang ZS, Cheung JY, Xia YS, Chen J. Neural implementation of unconstrained minimum L_1 -norm optimization—Least absolute deviation model and its application to time delay estimation. *IEEE Trans Circuits Syst II Analog Digit Signal Process* Nov;2000 47(11):1214–1226.
12. Wang ZS, He Z, Chen J. Robust time delay estimation of bioelectric signals. *IEEE Trans Biomed Eng* Mar;2005 52(3):454–462. [PubMed: 15759575]
13. Wang ZS, Cheung JY, Xia YS, Chen J. Minimum fuel neural networks and their applications to overcomplete signal representations. *IEEE Trans Circuits Syst I Fundam Theory Appl* Aug;2000 47 (8):1146–1159.
14. Hopfield JJ, Tank DW. Neural computation of decisions in optimization problems. *Biol Cybern* 1985;52:141–152. [PubMed: 4027280]
15. Tank DW, Hopfield JJ. Simple neural optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Trans Circuit Syst* May;1986 CAS-33(5):533–541.
16. Kennedy MP, Chua LO. Neural networks for nonlinear programming. *IEEE Trans Circuit Syst* May; 1988 CAS-35(5):554–562.
17. Maa CY, Shanblatt MA. A two-phase optimization neural network. *IEEE Trans Neural Netw* Nov; 1992 3(6):1003–1009. [PubMed: 18276497]
18. Sudharsannan S, Sundareshan M. Exponential stability and a systematic synthesis of a neural network for quadratic minimization. *Neural Netw* 1991;4:599–613.
19. Wu X, Xia Y, Li J, Chen WK. A high performance neural network for solving linear and quadratic programming problems. *IEEE Trans Neural Netw* May;1996 7(3):525–529. [PubMed: 18255604]
20. Xia YS. A new neural network for solving linear programming problems and its application. *IEEE Trans Neural Netw* Mar;1996 7(2):525–529. [PubMed: 18255604]
21. Xia YS, Wang J. A general methodology for designing globally convergent optimization neural networks. *IEEE Trans Neural Netw* Nov;1998 9(6):1331–1343. [PubMed: 18255813]
22. Rodriguez-Vazquez A, Domínguez-Castro R, Rueda A, Huertas JL, Sánchez-Sinencio E. Nonlinear switched-capacitor neural networks for optimization problems. *IEEE Trans Circuits Syst* Mar;1990 37(3):384–398.
23. Huang DS. A constructive approach for finding arbitrary roots of polynomials by neural networks. *IEEE Trans Neural Netw* Mar;2004 15(2):487–491.
24. Huang DS, Ip HHS, Chi Z. A neural root finder of polynomials based on root moments. *Neural Comput* 2004;16:1721–1762.
25. Miller, RK.; Michel, AN. *Ordinary Differential Equations*. New York: Academic; 1982.
26. Luenberger, DG. *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley; 1973.
27. LaSalle J. Stability theory for ordinary differential equations. *J Differ Equ* 1968;4:57–65.
28. LaSalle, J.; Lefschetz, S. *Stability by Lyapunov's Direct Method with Applications*. New York: Academic; 1961.
29. LaSalle, J. *The Stability of Dynamical Systems*. Philadelphia, PA: SIAM; 1976.
30. Kinderlehrer, D.; Stampacchia, G. *An Introduction to Variational Inequalities and Their Applications*. New York: Academic; 1980.
31. Tzividis Y, Anastassiou D. Switched-capacitor neural networks. *Electron Lett* 1987;23:958–959.
32. Cichocki A, Ramirez-Angulo J, Unbehauen R. Architectures for analog VLSI implementation of neural networks for solving linear equations with inequality constraints. *Proc IEEE Int Symp Circuits Syst* May;1992 3:1529–1532.
33. Cichocki A, Unbehauen R. Switched-capacitor neural networks of linear equations and related problems. *IEEE Trans Circuit Syst* Feb;1992 39(2):124–138.

Biographies

Zhishun Wang (M'99–SM'99) received the Ph.D. degree in signal processing from Southeast University, Nanjing, China, in 1997.

Currently, he is an Assistant Professor in Brain Imaging at the Department of Psychiatry, Columbia University, New York, NY. He is also a Research Scientist at the New York State Psychiatry Institute, New York. He has been extensively working on brain imaging and its application to the research of human cognitive process and childhood psychiatry with an emphasis on signal processing of functional magnetic resonance imaging (fMRI). His research

interests include functional and structural MRI, multidimensional signal/image processing, neural networks, chaos and fractal in the brain and the gut, independent component analysis, wavelet, and functional and effective brain connectivity.

Dr. Wang received the “Excellent Doctoral Thesis Award in Year 2000” from the Academic Committee of the State Council of China for his Ph.D. dissertation.

Bradley S. Peterson received the M.D. degree from the University of Wisconsin—Madison in 1987.

Currently, he is the Suzanne Crosby Murphy Professor in Pediatric Neuropsychiatry and Director of MRI Research at the Department of Psychiatry, Columbia University, New York, and New York State Psychiatric Institute, New York. He trained in general psychiatry at Massachusetts General Hospital and Harvard University, in child psychiatry at the Child Study Center of Yale University, and in MRI methodologies at Yale University. His research interests include neuropsychiatry and multimodality neuroimaging of childhood psychiatry.

Fig. 1.
Architecture of the CLADNN.

Fig. 2. MATLAB SIMULINK simulator of the CLADNN. (a) Simulator architecture. (b) Solution convergence process for solving problem 1 (linear fitting).

Fig. 3.

Convergence process of the CLADNN for solving ill-conditioned problem 2. (a) Convergence process of x_1 under different initial values. (b) Convergence process of x_2 under different initial values. (c) Convergence process of x_3 under different initial values. (d) Convergence process of energy function x_4 under different initial values.

Fig. 4. Solving a nonlinear-fitting problem using the CLADNN. (a) Solution convergence process for solving problem 3 (nonlinear fitting). (b) Comparison of the two nonlinear-fitting methods between the CLADNN (solid line) and the constrained LS (L_2 -norm) optimization (dashed-dotted line).

TABLE I

Nonlinear Fitting Data for Experiment 3

x	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
y	-9	-8	-7.6	-9	-7.1	-6.4	-3	-3.8	-2.4	-1.8	-3.1