A neurodynamic approach to convex optimization problems with general constraint

Sitian Qin^a, Yadong Liu^a, Xiaoping Xue^{b,*}, Fuqiang Wang^c

^aDepartment of Mathematics, Harbin Institute of Technology, Weihai, PR China

^bDepartment of Mathematics, Harbin Institute of Technology, Harbin, PR China

^cSchool of Automobile Engineering, Harbin Institute of Technology, Weihai, PR China.

Abstract

This paper presents a neurodynamic approach with a recurrent neural network for solving convex optimization problems with general constraint. It is proved that for any initial point, the state of the proposed neural network reaches the constraint set in finite time, and converges to an optimal solution of the convex optimization problem finally. In contrast to the existing related neural networks, the convergence rate of the state of the proposed neural network can be calculated quantitatively via the Lojasiewicz exponent under some mild assumptions. As applications, we estimate explicitly some Lojasiewicz exponents to show the convergence rate of the state of the proposed neural network for solving convex quadratic optimization problems. And some numerical examples are given to demonstrate the effectiveness of the proposed neural network.

Keywords: Nonsmooth convex optimization, neurodynamic approach, Lojasiewicz inequality, convergence in finite time.

1. Introduction

Consider the following nonlinear convex optimization programming

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \end{array} \quad (1)$$

where $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ is the vector of decision variables, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function, Ω is a nonempty close convex bounded subset of $\mathbb{R}^n$. Without loss of generality, in this paper, we assume that nonlinear convex programming (1) has at least an optimal solution.

*Corresponding author

Email addresses: qinsitian@163.com (Sitian Qin), hitliuyadong1994@163.com (Yadong Liu), xiaopingxue@263.net (Xiaoping Xue)

¹This research is supported by the National Science Foundation of China (61403101, 11271099, 11471088) and Weihai Science and technology Development Plan Project (2014DXGJ07).

Nonlinear programming (1) arises in all kinds of scientific and engineering applications, such as optimal control, structure design, signal processing, electrical network, robot control, power system planning, and so on (see M. S. Bazaraa & Shetty (1993)). In particular, most of existed nonlinear programming can be considered as the special cases of (1). Recently, numerous numerical algorithms are proposed to solve nonlinear programming (1). However, these numerical algorithms become inefficient for real-time online solutions for nonlinear programming (1) with increasing number of decision variables. Then, one possible and promising approach to solve nonlinear programming (1) with high dimension and complex structure in real time is to employ recurrent neural networks based on circuit implementation (Cochocki & Unbehauen (1993); Liao et al. (2004)). Since the pioneer work in Tank & Hopfield (1986), several recurrent neural networks arose for solving all kinds of nonlinear programming (1), such as neural network with a finite penalty parameter in Kennedy & Chua (1988), Lagrangian neural network in Zhang & Constantinides (1992), projection-type neural network in Xia & Wang (2005), and deterministic annealing neural network in Wang (1994). The aforementioned neural network models require the gradient information for functions appeared in the nonlinear optimization problems. In general, these models cannot immediately solve nonlinear nonsmooth optimization problems. Since the important applications in many engineering problems, the research on recurrent neural networks for nonsmooth or generalized optimization problems becomes very active in the neurodynamic optimization field (Li et al. (2014)). Recently, some simplified and generalized neural networks were proposed for solving nonlinear nonsmooth programming problems (see Forti et al. (2004); Bian & Xue (2009); Cheng et al. (2011); Gao & Liao (2010); Liu & Wang (2008a,b); Xue & Bian (2008); Hosseini (2016); Liu et al. (2016)).

It is well known that the stability of the neural network plays a key role in practical applications, and most researchers aim at proposing a neural network with global convergence property. In most cases, we hope the proposed neural networks converge to the equilibrium point as soon as possible. However, as far as we know, there are less conclusions concerning the convergence rate of the neural network for nonsmooth or generalized optimization problems. In Lojasiewicz (1963), Lojasiewicz proved a fundamental inequality defined by the gradient of an analytic function. The fundamental inequality is called classical Lojasiewicz inequality later. The classical Lojasiewicz inequality improves previous stability results via the Lyapunov approach. The classical Lojasiewicz inequality has been used to study the convergence rate of some neural networks (see Forti & Tesi (2004, 2006); Forti (2008)). But the classical Lojasiewicz inequality demands that the function be analytic, and cannot be used to deal with nonsmooth gradient dynamical systems.

Recently, authors in Bolte et al. (2007) proved the related nonsmooth Lojasiewicz inequality. The nonsmooth Lojasiewicz inequality permits to prove that each forward trajectory of the sub-gradient dynamical systems has finite length, and as a consequence it converges toward a singleton. The nonsmooth Lojasiewicz inequality involves a number belonging to $(0, 1)$, which depends on the critical

point, and is known as the Łojasiewicz exponent. Moreover, by means of a quantitative evaluation of the Łojasiewicz exponent at the equilibrium points, we can explicitly estimate the convergence rate of the solution for the non-smooth sub-gradient dynamical systems, such as exponential convergence, or convergence in finite time. The important nonsmooth Łojasiewicz inequality in Bolte et al. (2007) has been used to analyze the convergence of some neural networks. For example, based on the nonsmooth Łojasiewicz inequality, authors in Forti et al. (2006) studied the convergence of trajectories for the generalized gradient vector field describing the neural network dynamics for solving linear programming (LP) problems, convex quadratic programming (QP) problems, and nonconvex QP problems. Meanwhile, authors in Lu & Wang (2008); Qin & Xue (2010) studied the convergence of the states for two generalized neural networks based on the nonsmooth Łojasiewicz inequality. Furthermore, in Lu & Wang (2008); Qin & Xue (2010), they also calculated some Łojasiewicz exponents at the equilibrium points to show the convergence rate of the solutions.

Inspired by the above discussions, in this paper, we will propose a generalized neural network for solving more general nonlinear convex programming (1). The remainder of this paper is organized as follows. In Section 2, we present some preliminaries which are needed in the remainder of this paper. In Section 3, we propose a neural network for solving the optimization problem (1) and study the state convergence. In Section 4, we study the convergence rate of the proposed neural network by evaluating the Łojasiewicz exponent at the equilibrium point and some numerical examples are introduced to illustrate our results. Finally, Section 5 concludes this paper.

2. Preliminaries

For the convenience of later discussions, we present some definitions and properties, which are needed in the remainder of this paper. We refer readers to F.Clarke (1983) for more thorough discussions. In this paper, we let $\mathbb{R}^n$ be an n -dimensional real Euclidean space with inner product $\langle \cdot, \cdot \rangle$, and induced norm $\| \cdot \|$.

2.1. Nonsmooth analysis

For more thorough preliminaries about nonsmooth analysis, we refer readers to Aubin & Cellina (1984); Qin & Xue (2015). Here, we only introduce some necessary preliminaries.

Definition 2.1. (1) Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. The convex subdifferential of f at $x_0 \in \mathbb{R}^n$ is defined as

$$\partial f(x_0) = \{\xi \in \mathbb{R}^n : f(x_0) - f(x) \leq \langle \xi, x_0 - x \rangle, \forall x \in \mathbb{R}^n\}.$$

(2) Let f be Lipschitz near a given point $x_0 \in \mathbb{R}^n$. f is said to be regular at x_0 , if for all $v \in \mathbb{R}^n$, the usual one-sided directional derivative $f'(x_0; v)$ exists and $f'(x_0; v) = \limsup_{y \rightarrow x_0, t \downarrow 0} \frac{f(y+tv) - f(y)}{t}$. ■

Convex function has the following properties.

Proposition 2.1. (*F. Clarke (1983)*) If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function, then f is a regular function, and $\partial f : \mathbb{R}^n \rightarrow \mathbb{R}$ is an upper semi-continuous (U.S.C.) set-valued map with nonempty compact convex values. Moreover, $\partial f(\cdot)$ is maximal monotone, i.e.,

$$\langle x - x_0, v - v_0 \rangle \geq 0$$

for any $x, x_0 \in \mathbb{R}^n$, $v \in \partial f(x)$ and $v_0 \in \partial f(x_0)$.

Regular function employs a very important property (i.e., chain rule), which has been used in many papers, such as Forti et al. (2004); Bian & Xue (2009); Cheng et al. (2011); Gao & Liao (2010); Liu & Wang (2008a,b); Xue & Bian (2008); Forti et al. (2006).

Lemma 2.1. (*Chain Rule, Forti et al. (2006)*) Suppose that $V(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is regular in $\mathbb{R}^n$ and $x(t) : [0, +\infty) \rightarrow \mathbb{R}^n$ is absolutely continuous on any compact interval of $[0, +\infty)$. Then, $x(t)$ and $V(x(t))$ are differentiable for almost everywhere (a.e.) $t \in [0, +\infty)$ and we have

$$\frac{d}{dt}V(x(t)) = \langle \zeta, \dot{x}(t) \rangle \quad \forall \zeta \in \partial V(x(t)).$$

2.2. Distance function

Let $K \subset \mathbb{R}^n$ be a nonempty closed convex set. For any $x \in \mathbb{R}^n$, we define the distance function from x to the set K as follows,

$$d_K(x) = \inf_{y \in K} \|x - y\|.$$

Proposition 2.2. (*F. Clarke (1983) Proposition 2.4.2*) If $K \subset \mathbb{R}^n$ is a nonempty closed convex set, $d_K(x)$ is global Lipschitz and convex.

If $K \subset \mathbb{R}^n$ is a nonempty closed convex set, then there exists a unique point $\pi_K(x) \in K$ satisfying

$$\|x - \pi_K(x)\| = d_K(x) = \min_{y \in K} \|x - y\| \quad (2)$$

The operator $\pi_K(x)$ is called the Euclidean projection of x on K . In particular, by $m(K) = \pi_K(0)$, we denote the vector in K with the smallest norm. According to Theorem 10 in Borwein & Giles (1987) and Proposition 2.5.4 in F. Clarke (1983), we have the following proposition.

Proposition 2.3. If $K \subset \mathbb{R}^n$ is a nonempty closed convex set, then for any $x \in \mathbb{R}^n \setminus K$, $d_K(x)$ is differentiable and

$$\nabla d_K(x) = \frac{x - \pi_K(x)}{\|x - \pi_K(x)\|}. \quad (3)$$

In some cases, we can explicitly compute the projection $\pi_K(x)$ and the distance function $d_K(x)$. For example, from the section 8.1.1 of Boyd & Vandenberghe (2004), we have the following lemma.

Lemma 2.2. (i) *The Euclidean projection of x_0 on a hyperplane $K = \{x | a^T x = b\}$ is given by*

$$\pi_K(x_0) = x_0 + (b - a^T x_0) \frac{a}{\|a\|^2}, \quad (4)$$

and then

$$d_K(x_0) = \frac{|b - a^T x_0|}{\|a\|}. \quad (5)$$

(ii) *The Euclidean projection of x_0 on a halfspace $K = \{x | a^T x \leq b\}$ is given by*

$$\pi_K(x_0) = \begin{cases} x_0 + (b - a^T x_0) \frac{a}{\|a\|^2}, & a^T x_0 > b \\ x_0, & a^T x_0 \leq b \end{cases} \quad (6)$$

Consequently,

$$d_K(x_0) = \begin{cases} \frac{|b - a^T x_0|}{\|a\|}, & a^T x_0 > b \\ 0, & a^T x_0 \leq b \end{cases} \quad (7)$$

2.3. Subanalytic sets and nonsmooth Lojasiewicz inequality

We recall the following definitions, which are from the reference Bolte et al. (2007).

Definition 2.2. (1) *A subset A of $\mathbb{R}^n$ is called semianalytic if each point of $\mathbb{R}^n$ admits a neighborhood V for which $A \cap V$ assumes the following form:*

$$\bigcup_{i=1}^p \bigcap_{j=1}^q \{x \in V : f_{ij}(x) = 0, g_{ij}(x) > 0\}$$

where the functions $f_{ij}, g_{ij} : V \rightarrow \mathbb{R}$ are real analytic for all $1 \leq i \leq p, 1 \leq j \leq q$.

(2) *The subset A of $\mathbb{R}^n$ is called subanalytic if each point of $\mathbb{R}^n$ admits a neighborhood V such that*

$$A \cap V = \{x \in \mathbb{R}^n : (x, y) \in B\}$$

where B is a bounded semianalytic subset of $\mathbb{R}^n \times \mathbb{R}^m$ for some $m \geq 1$.

(3) *$f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is called subanalytic if its graph is a subanalytic subset of $\mathbb{R}^n \times \mathbb{R}$. ■*

For each $n \in N$, we let $C_n = (-1, 1)^n$ and define τ_n by

$$\tau_n(x_1, x_2, \dots, x_n) = \left(\frac{x_1}{1 + x_1^2}, \frac{x_2}{1 + x_2^2}, \dots, \frac{x_n}{1 + x_n^2} \right) \in C_n.$$

A subset A of $\mathbb{R}^n$ is called globally subanalytic if its image under τ_n is a subanalytic subset of $\mathbb{R}^n$. An extended-real-valued function is called globally subanalytic if its graph is globally subanalytic. In fact, globally subanalytic sets are subanalytic, and conversely any bounded subanalytic set is globally subanalytic. More generally, the image or the preimage of a globally subanalytic set by a globally subanalytic function is globally subanalytic.

For a given subanalytic set S , the distance $d_S(x) = \inf\{\|x - a\| : a \in S\}$ is a subanalytic function.

Definition 2.3. (Rockafellar & Wets (2009) Definition 8.3). (1) The Fréchet subdifferential $\hat{\partial}f(x)$ of a lower semicontinuous f at $x \in \mathbb{R}^n$ is given by

$$\hat{\partial}f(x) = \{x^* \in \mathbb{R}^n : \liminf_{y \rightarrow x, y \neq x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} \geq 0\}$$

whenever $x \in \text{dom}f$, and by $\hat{\partial}f(x) = \emptyset$ otherwise.

(2) The limiting subdifferential at $x \in \mathbb{R}^n$, denoted by $\partial_l f(x)$, is the set of all cluster points of sequence $\{x_n^*\}_{n \geq 1}$ such that $x_n^* \in \hat{\partial}f(x_n)$ and $(x_n, f(x_n)) \rightarrow (x, f(x))$ as $n \rightarrow +\infty$.

The following lemma presents the nonsmooth Łojasiewicz inequality with a subanalytic function.

Lemma 2.3. (Bolte et al. (2007)) Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a subanalytic function with closed domain, and assume that $f|_{\text{dom}f}$ is continuous. Let $a \in \mathbb{R}^n$ be a critical point of f , i.e., $0 \in \partial_l f(a)$. Then there exist $\theta \in [0, 1)$, $c > 0$, $\delta > 0$ such that

$$|f(x) - f(a)|^\theta \leq c \|m(\partial_l f(x))\| \quad (8)$$

for all $x \in B(a, \delta)$, where $m(\partial_l f(x)) \in \partial_l f(x)$ and $\|m(\partial_l f(x))\| = \inf\{\|\xi\| : \xi \in \partial_l f(x)\}$.

We let

$$\theta_a = \inf\{\theta \in [0, 1) : \text{the Łojasiewicz inequality (8) is satisfied}\} \quad (9)$$

and θ_a is called Łojasiewicz exponent of f at a .

Lemma 2.4. (Rockafellar & Wets (2009), Th. 6.29 and 8.30) If $f(x)$ is finite and convex on $\mathbb{R}^n$, then we have

$$\partial_l f(x) = \partial f(x).$$

3. Problem formulation and neural network model

In this section, we propose a generalized neural network for solving the optimization problem (1) and study its state convergence.

Lemma 3.1. x^* is an optimal solution of the optimization problem (1), if and only if there exists a positive constant λ_0 (depends on f and Ω), such that for any $\lambda > \lambda_0$, we have

$$\begin{cases} 0 \in \partial f(x^*) + \lambda \partial d_\Omega(x^*), \\ x^* \in \Omega. \end{cases} \quad (10)$$

PROOF. Suppose that x^* is an optimal solution of the optimization problem (1). Since Ω is a compact subset of $\mathbb{R}^n$, $\overline{B}(\Omega, 1) = \{x \in \mathbb{R}^n : d_\Omega(x) \leq 1\}$ is also compact. Meanwhile, f is locally Lipschitz continuous on $\mathbb{R}^n$. By Heine-Borel finite cover theorem, f is Lipschitz continuous on $\overline{B}(\Omega, 1)$. That is, there exists $\lambda_0 > 0$ (depends on f and Ω), such that

$$|f(x_1) - f(x_2)| \leq \lambda_0 \|x_1 - x_2\|$$

for all $x_1, x_2 \in \overline{B}(\Omega, 1)$. For any $\lambda > \lambda_0$, we let

$$F_\lambda(x) = f(x) + \lambda d_\Omega(x). \quad (11)$$

It is obvious that $F_\lambda(x)$ is a convex function. We next prove that x^* is a minimum point of the function $F_\lambda(x) = f(x) + \lambda d_\Omega(x)$ over $B(x^*, 1/3) = \{x \in \mathbb{R}^n : \|x - x^*\| < 1/3\}$. In fact, by the definition of $d_\Omega(x)$, for any $x \in B(x^*, 1/3)$ and a sufficiently small $\epsilon \in (0, 1)$, there exists $x_\epsilon \in \Omega$, such that $\|x - x_\epsilon\| \leq d_\Omega(x)(1 + \epsilon)$. Then, we have

$$\|x^* - x_\epsilon\| \leq \|x^* - x\| + \|x - x_\epsilon\| \leq \|x^* - x\| + (1 + \epsilon)\|x^* - x\| < 1.$$

That is, $x_\epsilon \in B(x^*, 1) \subseteq \overline{B}(\Omega, 1)$. So,

$$f(x_\epsilon) \leq f(x) + \lambda \|x - x_\epsilon\| \leq f(x) + \lambda d_\Omega(x)(1 + \epsilon).$$

Since x^* is an optimal solution of the optimization problem (1), we have

$$F_\lambda(x^*) = f(x^*) + \lambda d_\Omega(x^*) = f(x^*) \leq f(x_\epsilon) \leq f(x) + \lambda d_\Omega(x)(1 + \epsilon) = F_\lambda(x) + \epsilon \lambda d_\Omega(x).$$

Letting $\epsilon \rightarrow 0^+$, it follows that $F_\lambda(x^*) \leq F_\lambda(x)$ for any $x \in B(x^*, 1/3)$. Thus, x^* is the local minimum point of $F_\lambda(x) = f(x) + \lambda d_\Omega(x)$ over $B(x^*, 1/3)$, which implies that $0 \in \partial f(x^*) + \lambda \partial d_\Omega(x^*)$.

On the other hand, if there exists a positive constant λ_0 , such that for any $\lambda > \lambda_0$, we have $0 \in \partial f(x^*) + \lambda \partial d_\Omega(x^*) = \partial F_\lambda(x^*)$. From Definition 2.1, for any $x \in \mathbb{R}^n$, we have $F_\lambda(x) - F_\lambda(x^*) \geq \langle 0, x - x^* \rangle = 0$. In particular, for any $x \in \Omega$,

$$f(x) - f(x^*) = f(x) + \lambda d_\Omega(x) - (f(x^*) + \lambda d_\Omega(x^*)) = F_\lambda(x) - F_\lambda(x^*) \geq 0.$$

Since $x^* \in \Omega$, then x^* is an optimal solution of the optimization problem (1).

Remark 1. (i) From the above proof, λ_0 can be taken as $\lambda_0 > \sup\{\|\xi\| : \xi \in \partial f(x), x \in \overline{B}(\Omega, 1)\}$. In particular, if f is differentiable on the bounded set $\overline{B}(\Omega, 1)$, then λ_0 can be taken as $\max\{\|\nabla f(x)\| : x \in \overline{B}(\Omega, 1)\}$.

(ii) If we let

$$\begin{aligned} E_\lambda &= \{x_e \in \mathbb{R}^n : 0 \in \partial F_\lambda(x_e)\} \\ M &= \{x^* \in \Omega : f(x^*) = \min_{x \in \Omega} f(x)\} \end{aligned} \quad (12)$$

then, based on the Lemma 3.1, there exists a positive constant λ_0 , such that for any $\lambda > \lambda_0$, we obtain that

$$E_\lambda \bigcap \Omega = M. \quad (13)$$

Based on the Lemma 3.1, the dynamic equations of the proposed neural network model for solving problem (1) are described as follows,

$$\sigma \dot{x}(t) \in -\partial F_\lambda(x(t)) = -\partial f(x(t)) - \lambda \partial d_\Omega(x(t)) \quad (14)$$

where σ is a positive scaling constant.

Definition 3.1. A vector function $x(\cdot)$ is said to be a state of neural network (14) on $[0, T)$, if $x(\cdot)$ is absolutely continuous on $[0, T)$ with initial condition $x(0) = x_0$ and for almost everywhere (a.e.) $t \in [0, T)$,

$$\sigma \dot{x}(t) \in -\partial F_\lambda(x(t)).$$

Equivalently, there exist measurable functions $\gamma(t) \in \partial f(x(t))$, $\eta(t) \in \partial d_\Omega(x(t))$, such that

$$\sigma \dot{x}(t) = -\gamma(t) - \lambda \eta(t)$$

for a.e. $t \in [0, T)$.

Definition 3.2. x^* is called an equilibrium point of neural network (14) if

$$0 \in \partial F_\lambda(x^*) = \partial f(x^*) + \lambda \partial d_\Omega(x^*),$$

i.e., there exist $\gamma^* \in \partial f(x^*)$, $\eta^* \in \partial d_\Omega(x^*)$ such that

$$0 = \gamma^* + \lambda \eta^*. \quad \blacksquare$$

Since $f(x)$ and $d_\Omega(x)$ are convex functions, $\partial F_\lambda(x) = \partial f(x) + \lambda \partial d_\Omega(x)$ is a U.S.C and maximal monotone set-valued map with nonempty compact convex values. Then, by Theorem 1 on page 147 in Aubin & Cellina (1984), we have the following conclusions.

Theorem 3.1. For any initial point $x(0) = x_0 \in \mathbb{R}^n$, there exists a unique solution $x(t)$ of neural network (14), which is defined on $[0, +\infty)$. Moreover, the unique solution $x(t)$ satisfies

$$\sigma \dot{x}(t) = -m(\partial F_\lambda(x(t))) \quad (15)$$

for a.e. $t \in [0, +\infty)$. Here $m(\cdot)$ is defined in (8). $\blacksquare$

Since Ω is bounded, we can introduce the following assumption.

Assumption 1 : There exists $R > 0$ such that

$$\Omega \subset B(x^*, R)$$

where x^* is an optimal solution of the problem (1) and $B(x^*, R) = \{x : \|x - x^*\| < R\}$.

We let

$$M_f = \sup\{\|\xi\| : \xi \in \partial f(x), x \in B(x^*, R)\}. \quad (16)$$

Since $\partial f(x)$ is a U.S.C and maximal monotone set-valued map with nonempty compact convex values, we have $0 \leq M_f < +\infty$.

Theorem 3.2. *Suppose that **Assumption 1** holds and $\lambda > M_f$ in (14). For any initial point $x(0) = x_0 \in B(x^*, R)$, there exists $T > 0$, such that $x(t) \in \Omega$ for any $t > T$.*

The proof of Theorem 3.2 is shown in the Appendix 6.1.

Remark 2. *From Theorem 3.2, the solution of neural network (14) reaches feasible region Ω in finite time and stays there thereafter. From definition 3.2, the equilibrium point is a special solution of neural network. Thus, from Theorem 3.2 and (ii) in Remark 1, if **Assumption 1** holds, we have*

$$E_\lambda = M. \quad (17)$$

Theorem 3.3. *Suppose that **Assumption 1** holds and $\lambda > M_f$. For any initial point $x(0) = x_0 \in B(x^*, R)$, the state $x(t)$ of neural network (14) converges to an optimal solution of the nonlinear convex programming (1).*

PROOF. We construct the following Lyapunov function,

$$V(x) = f(x) - f(x^*) + \lambda d_\Omega(x) - \lambda d_\Omega(x^*) + \frac{1}{2} \|x - x^*\|^2. \quad (18)$$

It's clear that $V(x) > 0$ for all $x \in \Omega \setminus \{x^*\}$. And the subdifferential of V can be calculated as follows,

$$\partial V(x) = \partial f(x) + \lambda \partial d_\Omega(x) + x - x^*. \quad (19)$$

Since x^* is an optimal solution of the nonlinear convex programming (1), by Lemma 3.1 and Definition 3.2, for $\lambda > M_f$, there exist $\gamma^* \in \partial f(x^*)$, $\eta^* \in \partial d_\Omega(x^*)$ such that

$$0 = \gamma^* + \lambda \eta^* \in \partial f(x^*) + \lambda \partial d_\Omega(x^*). \quad (20)$$

For any initial point $x(0) = x_0 \in B(x^*, R)$, by Definition 3.1 and Theorem 3.1, there exist measurable functions $\gamma(t) \in \partial f(x(t))$, $\eta(t) \in \partial d_\Omega(x(t))$, such that

$$\sigma \dot{x}(t) = -\gamma(t) - \lambda \eta(t) \quad (21)$$

for a.e. $t \in [0, +\infty)$. From (19) and (21), we obtain that

$$-\sigma \dot{x}(t) + x(t) - x^* \in \partial V(x(t)). \quad (22)$$

Hence, by Chain Rule, for a.e. $t \in [0, +\infty)$, we have

$$\begin{aligned} \frac{d}{dt} V(x(t)) &= \langle \gamma(t) + \lambda \eta(t) + x(t) - x^*, \dot{x}(t) \rangle \\ &= \frac{1}{\sigma} \langle \gamma(t) + \lambda \eta(t) + x(t) - x^*, -\gamma(t) - \lambda \eta(t) \rangle \\ &= -\frac{1}{\sigma} \|\gamma(t) + \lambda \eta(t)\|^2 + \frac{1}{\sigma} \langle x(t) - x^*, -\gamma(t) - \lambda \eta(t) \rangle. \end{aligned} \quad (23)$$

From Proposition 2.1 and (20),

$$\frac{1}{\sigma} \langle x(t) - x^*, -\gamma(t) - \lambda \eta(t) \rangle = -\frac{1}{\sigma} \langle x^* - x(t), \gamma^* + \lambda \eta^* - \gamma(t) - \lambda \eta(t) \rangle \leq 0. \quad (24)$$

Hence, by (23),

$$\frac{d}{dt} V(x(t)) \leq -\frac{1}{\sigma} \|\gamma(t) + \lambda \eta(t)\|^2. \quad (25)$$

We let

$$H(x) = \frac{1}{\sigma} \min\{\|\gamma + \lambda \eta\|^2 : \gamma \in \partial f(x), \eta \in \partial d_\Omega(x)\}.$$

Then, $H(x) = 0$ if and only if x is an equilibrium point of neural network (14). And by (25), we have

$$\frac{d}{dt} V(x(t)) \leq -H(x(t)). \quad (26)$$

Since $x(t)$ is bounded by (65), then there exist a monotone increasing sequence $\{t_n\}$ and $\tilde{x}$, such that $\lim_{n \rightarrow +\infty} x(t_n) = \tilde{x}$ with $\lim_{n \rightarrow +\infty} t_n = +\infty$. From Theorem 3.2, $\tilde{x} \in \Omega$. Meanwhile, there exists a positive constant $M > 0$ such that

$$\|\dot{x}(t)\| \leq M$$

for a.e. $t > 0$. Similar to the proof in Li et al. (2006), in the following, we prove that $H(\tilde{x}) = 0$ by contradiction. If $H(\tilde{x}) \neq 0$, we have $H(\tilde{x}) > 0$. According to Theorem 5 on page 52 in Aubin & Cellina (1984), H is lower semicontinuous. Hence, by the assumption that $H(\tilde{x}) > 0$, there exist $\varepsilon > 0$ and $\delta > 0$, such that

$$H(x) > \varepsilon \quad (27)$$

for all $x \in B(\tilde{x}, \delta)$. Since $\lim_{n \rightarrow +\infty} x(t_n) = \tilde{x}$, there exists a positive integer N , such that for any $n > N$,

$$x(t_n) \in B(\tilde{x}, \delta/2). \quad (28)$$

For any $t \in [t_n - \frac{\delta}{4M}, t_n + \frac{\delta}{4M}]$ with $n > N$,

$$\|x(t) - \tilde{x}\| \leq \|x(t) - x(t_n)\| + \|x(t_n) - \tilde{x}\| \leq M|t - t_n| + \frac{\delta}{2} \leq \delta \quad (29)$$

which means that

$$H(x(t)) > \varepsilon \quad (30)$$

for any $t \in [t_n - \frac{\delta}{4M}, t_n + \frac{\delta}{4M}]$ with $n > N$. Hence

$$\begin{aligned} \int_0^{+\infty} H(x(t))dt &\geq \int_{\bigcup_{n \geq N} [t_n - \frac{\delta}{4M}, t_n + \frac{\delta}{4M}]} H(x(t))dt \\ &\geq \int_{\bigcup_{n \geq N} [t_n - \frac{\delta}{4M}, t_n + \frac{\delta}{4M}]} \varepsilon dt = \sum_{n \geq N} \frac{\delta}{2M} \varepsilon = +\infty. \end{aligned} \quad (31)$$

On the other hand, according to $V(x(t)) > 0$ and (25), there exists a positive constant V_0 such that $\lim_{t \rightarrow +\infty} V(x(t)) = V_0$. Based on (26), we have

$$\begin{aligned} \int_0^{+\infty} H(x(t))dt &= \lim_{k \rightarrow +\infty} \int_0^k H(x(t))dt \leq - \lim_{k \rightarrow +\infty} \int_0^k \dot{V}(x(t))dt \\ &= - \lim_{k \rightarrow +\infty} [V(x(k)) - V(x(0))] \\ &= V(x(0)) - V_0 < +\infty \end{aligned} \quad (32)$$

which contradicts with (31). Thus, we have $H(\tilde{x}) = 0$, i.e, $\tilde{x}$ is an equilibrium point of neural network (14). Then, by Remark 2, $\tilde{x}$ is an optimal solution of nonlinear convex programming (1). We next prove that $\lim_{t \rightarrow +\infty} x(t) = \tilde{x}$, which completes the proof of the theorem.

Let us define another Lyapunov function

$$\tilde{V}(x) = f(x) - f(\tilde{x}) + \lambda d_{\Omega}(x) - \lambda d_{\Omega}(\tilde{x}) + \frac{1}{2} \|x - \tilde{x}\|^2.$$

Similar to the above proof, we have

$$\frac{d}{dt} \tilde{V}(x(t)) \leq 0 \quad \text{and} \quad \tilde{V}(x(t)) > 0 \quad (33)$$

for all $x \in \Omega \setminus \{\tilde{x}\}$. By the continuity of $\tilde{V}(x)$, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\tilde{V}(x) < \varepsilon \quad (34)$$

when $\|x - \tilde{x}\| < \delta$. Since $\tilde{V}(x(t))$ is monotonically non-increasing on $[0, +\infty)$ and $\lim_{n \rightarrow +\infty} x(t_n) = \tilde{x}$, there exists a positive integer K , such that $\|x(t_K) - \tilde{x}\| < \delta$. Hence

$$0 < \frac{1}{2} \|x(t) - \tilde{x}\|^2 \leq \tilde{V}(x(t)) \leq \tilde{V}(x(t_K)) < \varepsilon \quad (35)$$

for all $t > t_K$. That is,

$$\lim_{t \rightarrow +\infty} x(t) = \tilde{x}. \quad (36)$$

4. Convex subanalytic programming problems and numerical examples

In this section, based on Łojasiewicz inequality, we will study the convergence rate of the state of the neural network (14) for solving convex subanalytic programming as follows,

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && x \in \Omega \end{aligned} \quad (37)$$

where $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ is the vector of decision variables, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex and subanalytic function, Ω is a nonempty close convex and subanalytic subset of $\mathbb{R}^n$.

Based on the nonsmooth Łojasiewicz inequality, we have the following theorem.

Theorem 4.1. *Suppose that **Assumption 1** holds and $\lambda > M_f$.*

(1) *For any initial point $x(0) = x_0 \in \mathbb{R}^n$, the state $x(t)$ of neural network (14) converges to an optimal solution $\tilde{x}$ of the nonlinear convex subanalytic programming (37).*

(2) *Let $\theta_{\tilde{x}}$ be the Łojasiewicz exponent at $\tilde{x}$ defined in Lemma 2.3. Then there exist $k > 0$, $k' > 0$ and $t_0 > 0$ such that for any $t \geq t_0$, the following estimates hold:*

- i) if $\theta_{\tilde{x}} \in [0, 1/2)$, then $x(t)$ converges to $\tilde{x}$ in finite time;*
- ii) if $\theta_{\tilde{x}} = 1/2$, then $\|x(t) - \tilde{x}\| \leq ke^{k't}$;*
- iii) if $\theta_{\tilde{x}} \in (1/2, 1)$, then $\|x(t) - \tilde{x}\| \leq k(t+1)^{(1-\theta)/(1-2\theta)}$.*

The proof of Theorem 4.1 is shown in the Appendix 6.2.

As an application of Theorem 4.1, we will study the convergence rate of the state of the neural network (14) for solving some convex subanalytic programming.

4.1. Convex quadratic programming problems with linear inequality constraints

Consider the following convex quadratic programming,

$$\begin{aligned} & \text{minimize} && f(x) = x^T Q x + b^T x \\ & \text{subject to} && g_j(x) = A_j^T x - v_j \geq 0, \quad j = 1, \dots, q \end{aligned} \quad (38)$$

where $Q \in \mathbb{R}^n \times \mathbb{R}^n$ is a positive semi-definite matrix, $x, b \in \mathbb{R}^n$, $0 \neq A_j \in \mathbb{R}^n$, and $v_j \in \mathbb{R}$, $j = 1, \dots, q$.

Similar to the **Assumption 1**, we let

$$\Omega = \{x \in \mathbb{R}^n : A_j^T x - v_j \geq 0, \quad j = 1, \dots, q\}$$

and suppose that Ω is a bounded set. According to the Example 4.4 in Bolte et al. (2007), Ω is a nonempty close convex and subanalytic set. It is obvious that f is subanalytic. Hence, based on Theorem 4.1, neural network (14) can solve the convex quadratic programming (38). In this case, neural work (14) can be explicitly expressed as follows,

$$\sigma \dot{x} \in -\partial F_\lambda(x) = -2Qx - b - \lambda \partial d_\Omega(x). \quad (39)$$

Remark 3. In general, we believe that it is not an easy work to calculate $\partial d_\Omega(x)$. However, in some cases, we can calculate $\partial d_\Omega(x)$ explicitly. Next, we propose some examples to explain it.

(I) When $\Omega = \{x \in \mathbb{R}^n : A_j^T x - v_j \geq 0, j = 1, \dots, q\}$, by Lemma 2.2, we can calculate $\partial d_\Omega(x)$ explicitly. In Figure 1, we let

$$l_i = \{x \in \mathbb{R}^2 | A_i^T x = v_i\} \cap \Omega, \quad i = 1, 2, 3, 4$$

and $\Omega = \bigcap_{i=1}^4 \{x \in \mathbb{R}^2 | A_i^T x \geq v_i\}$. Then, $\mathbb{R}^2$ is divided into 9 disjoint subset regions as follows,

$$\mathbb{R}^2 = \Omega \cup \left(\bigcup_{k=1}^8 \Omega_k \right).$$

From Lemma 2.2, we can calculate explicitly $d_\Omega(x)$ and $\partial d_\Omega(x)$ in different regions, such as

$$\begin{aligned} d_\Omega(x) &= \begin{cases} |v_1 - A_1^T x| / \|A_1\|, & x \in \Omega_1 \\ \|x - p_1\|, & x \in \Omega_2 \\ 0, & x \in l_1 \end{cases} \\ \partial d_\Omega(x) &= \begin{cases} -A_1 / \|A_1\|, & x \in \Omega_1 \\ (x - p_1) / \|x - p_1\|, & x \in \Omega_2 \\ -[0, 1]A_1 / \|A_1\|, & x \in l_1 \setminus \{p_1, p_4\} \\ -\overline{co}\{[0, 1] \frac{A_1}{\|A_1\|}, [0, 1] \frac{A_2}{\|A_2\|}\}, & x = p_1 \end{cases} \end{aligned} \quad (40)$$

where $p_1 = l_1 \cap l_2$. More details, please see Figure 1.

(II) When $\Omega = \overline{B}(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}$, we can also calculate $\partial d_\Omega(x)$ explicitly as follows,

$$d_\Omega(x) = \begin{cases} \|x - x_0\| - r, & x \notin \Omega \\ 0, & x \in \Omega \end{cases}, \quad \partial d_\Omega(x) = \begin{cases} \frac{x - x_0}{\|x - x_0\|}, & x \notin \Omega \\ [0, 1] \frac{x - x_0}{\|x - x_0\|}, & x \in bd(\Omega) \\ 0, & x \in int(\Omega) \end{cases} \quad (41)$$

(III) When Ω is an intersection between some half-planes and ball, $\partial d_\Omega(x)$ can also be calculated explicitly. For example, in Figure 2,

$$\Omega = \{x \in \mathbb{R}^2 | A_1^T x \geq v_1\} \cap \{x \in \mathbb{R}^2 | A_2^T x \geq v_2\} \cap \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}.$$

Here, we calculate $d_\Omega(x)$ and $\partial d_\Omega(x)$ in Ω_1 , Ω_2 and Ω_3 as follows,

Figure 1: Illustration in Remark 3 (I) for 9 disjoint polyhedral regions in $\mathbb{R}^2$.

$$d_{\Omega}(x) = \begin{cases} |v_1 - A_1^T x| / \|A_1\|, & x \in \Omega_1 \\ \|x - p_1\|, & x \in \Omega_2 \\ \|x - x_0\| - r, & x \in \Omega_3 \end{cases}$$

$$\partial d_{\Omega}(x) = \begin{cases} -A_1 / \|A_1\|, & x \in \Omega_1 \\ \overline{\text{co}}\{-\frac{A_1}{\|A_1\|}, (x - p_1) / \|x - p_1\|\}, & x \in \Omega_1 \cap \Omega_2 \\ (x - p_1) / \|x - p_1\|, & x \in \Omega_2 \\ \overline{\text{co}}\{(x - p_1) / \|x - p_1\|, \frac{x - x_0}{\|x - x_0\|}\}, & x \in \Omega_2 \cap \Omega_3 \\ \frac{x - x_0}{\|x - x_0\|}, & x \in \Omega_3 \end{cases} \quad (42)$$

Next, in order to study the convergence rate of the state, we will calculate the Łojasiewicz exponent at the equilibrium point of neural network (39) for solving the convex quadratic programming (38). Let $\tilde{x}$ be an equilibrium point of the neural network (39), from Definition 3.2, there exists $\eta_{\tilde{x}} \in \partial d_{\Omega}(\tilde{x})$ such that

$$0 = -2Q\tilde{x} - b - \lambda\tilde{\eta} \in -2Q\tilde{x} - b - \lambda\partial d_{\Omega}(\tilde{x}). \quad (43)$$

Figure 2: Illustration in Remark 3 (III) for disjoint regions in $\mathbb{R}^2$.

For simplicity, we let

$$\Gamma_\theta(x) = \frac{|F_\lambda(x) - F_\lambda(\tilde{x})|^\theta}{\|m(\partial F_\lambda(x))\|}. \quad (44)$$

Theorem 4.2. *Let $\tilde{x}$ be an equilibrium point of the neural network (39), and $\theta_{\tilde{x}}$ be the Lojasiewicz exponent of F_λ at point $\tilde{x}$. Then we have,*

- 1) *if $\tilde{x} \in \text{int}(\Omega)$, then $\theta_{\tilde{x}} = 1/2$.*
- 2) *if $\tilde{x} \in \text{bd}(\Omega)$, and $\eta_{\tilde{x}} \in \text{int}(\partial d_\Omega(\tilde{x}))$, then $\theta_{\tilde{x}} = 0$.*
- 3) *if $\tilde{x} \in \text{bd}(\Omega)$, and $\eta_{\tilde{x}} \in \text{bd}(\partial d_\Omega(\tilde{x}))$, then $\theta_{\tilde{x}} = 1/2$.*

PROOF. Firstly, if $x \in E_\lambda$, then by the definition of E_λ in (12), x is an equilibrium point of the neural network (39). Since F_λ is a convex function on $\mathbb{R}^n$, we have $F_\lambda(x) = F_\lambda(\tilde{x})$, which means for any $\theta \in (0, 1)$, the Lojasiewicz inequality (8) holds for any $x \in E_\lambda$. Next, we only consider the case that $x \notin E_\lambda$.

1) If $\tilde{x} \in \text{int}(\Omega)$, then $d_\Omega(\tilde{x}) = 0$ and $\partial d_\Omega(\tilde{x}) = \{0\}$. In this case, the neural network (39) reduces to

$$\sigma \dot{x} = -2Qx - b \quad (45)$$

and $-2Q\tilde{x} - b = 0$. Hence, by (44)

$$\begin{aligned} \Gamma_\theta(x) &= \frac{|x^T Qx + b^T x - \tilde{x}^T Q\tilde{x} - b^T \tilde{x}|^\theta}{\|2Qx + b\|} \\ &= \frac{|(x - \tilde{x})^T Q(x - \tilde{x}) + x^T Q\tilde{x} + \tilde{x}^T Qx - 2\tilde{x}^T Q\tilde{x} + b^T(x - \tilde{x})|^\theta}{\|2Qx - 2Q\tilde{x}\|} \\ &= \frac{|(x - \tilde{x})^T Q(x - \tilde{x})|^\theta}{\|2Q(x - \tilde{x})\|}. \end{aligned} \quad (46)$$

Let $\mathfrak{R} = \ker(Q)$ and $\mathfrak{R}^\perp = \ker^\perp(Q)$. Then, we have $(x - \tilde{x}) = \pi_{\mathfrak{R}}(x - \tilde{x}) + \pi_{\mathfrak{R}^\perp}(x - \tilde{x})$. If $x \notin E_\lambda$, we have $2Q(x - \tilde{x}) \neq 0$ and $\pi_{\mathfrak{R}^\perp}(x - \tilde{x}) \neq 0$. Therefore,

we obtain

$$|(x - \tilde{x})^T Q(x - \tilde{x})| = |(\pi_{\mathbb{R}^\perp}(x - \tilde{x}))^T Q \pi_{\mathbb{R}^\perp}(x - \tilde{x})| \leq \lambda_M \|\pi_{\mathbb{R}^\perp}(x - \tilde{x})\|^2$$

where $\lambda_M = \max\{\lambda_i : \lambda_i \text{ is an eigenvalue of } Q\}$. In addition, we also have

$$\|2Q(x - \tilde{x})\| = \|2Q\pi_{\mathbb{R}^\perp}(x - \tilde{x})\| \geq 2\lambda_m \|\pi_{\mathbb{R}^\perp}(x - \tilde{x})\| > 0.$$

for all $x \notin E_\lambda$, where $\lambda_m = \min\{\lambda_i : \lambda_i \neq 0 \text{ is an eigenvalue of } Q\}$. Note that $\lambda_m > 0$ since Q is a non-zero positive semi-definite matrix. Hence, for any $\theta \in (1/2, 1]$ and $x \notin E_\lambda$, we have

$$\Gamma_\theta(x) = \frac{|(x - \tilde{x})^T Q(x - \tilde{x})|^\theta}{\|2Q(x - \tilde{x})\|} \leq \frac{(\lambda_M)^\theta}{2\lambda_m} \|\pi_{\mathbb{R}^\perp}(x - \tilde{x})\|^{2\theta-1} < +\infty \quad (47)$$

which means that $\theta_{\tilde{x}} \leq 1/2$. On the other hand, if $\theta < 1/2$, similar to the above proof,

$$\Gamma_\theta(x) = \frac{|(x - \tilde{x})^T Q(x - \tilde{x})|^\theta}{\|2Q(x - \tilde{x})\|} \geq \frac{(\lambda_m)^\theta}{2\lambda_M} \|\pi_{\mathbb{R}^\perp}(x - \tilde{x})\|^{2\theta-1} \quad (48)$$

is unbounded in any neighborhood of $\tilde{x}$.

Hence, $\theta_{\tilde{x}} = 1/2$.

2) Suppose that $\tilde{x} \in bd(\Omega)$ and $\eta_{\tilde{x}} \in \text{int}(\partial d_\Omega(\tilde{x}))$. From the construction of Ω , we know that $\tilde{x}$ must be a vertex of Ω . By (43), $-2Q\tilde{x} - b - \lambda\eta_{\tilde{x}} = 0$, then $-2Q\tilde{x} - b = \lambda\eta_{\tilde{x}} \in \text{int}(\lambda\partial d_\Omega(\tilde{x}))$. So

$$d_{bd(\lambda\partial d_\Omega(\tilde{x}))}(-2Q\tilde{x} - b) > 0$$

which means there exists $r > 0$, such that for all $x \in B(\tilde{x}, r) \cap \Omega \setminus E_\lambda$,

$$d_{bd(\lambda\partial d_\Omega(\tilde{x}))}(-2Qx - b) \geq \frac{d_{bd(\lambda\partial d_\Omega(\tilde{x}))}(-2Q\tilde{x} - b)}{2} \triangleq \delta > 0.$$

Moreover, for any $\xi \in \partial d_\Omega(x)$, we have

$$\| -2Qx - b - \lambda\xi \| \geq d_{bd(\lambda\partial d_\Omega(x))}(-2Qx - b) \geq d_{bd(\lambda\partial d_\Omega(\tilde{x}))}(-2Qx - b) \geq \delta.$$

Therefore

$$\Gamma_\theta(x) \leq \frac{|(x - \tilde{x})^T Q(x - \tilde{x}) - \lambda(x - \tilde{x})^T \eta_{\tilde{x}}|^\theta}{\delta}.$$

Then, for any $\theta \in [0, 1)$, $\Gamma_\theta(x)$ is bounded in a neighborhood of $\tilde{x}$. Hence, $\theta_{\tilde{x}} = 0$.

3) From Remark 3, when $\tilde{x} \in bd(\Omega)$, $\partial d_\Omega(\tilde{x})$ is a closed and convex hull of some $[0, 1]A_i/\|A_i\|$. For any $x \in bd(\Omega)$, we take $\eta_x \in \partial d_\Omega(x)$ which satisfies

$$\|2Qx + b + \lambda\eta_x\| = \min\{\|2Qx + b + \lambda\xi\| : \xi \in \partial d_\Omega(x)\}. \quad (49)$$

Hence, by the definition of Euclidean projection in (2),

$$\eta_x = \pi_{\partial d_\Omega(x)}((-2Qx - b)/\lambda). \quad (50)$$

In the following, we first consider the case that $\eta_{\tilde{x}} \neq 0$.

(3-1) If $\tilde{x} \in bd(\Omega)$ with a unique j_0 satisfying $A_{j_0}^T \tilde{x} - v_{j_0} = 0$, then there exists a neighborhood $U_{\tilde{x}}$ of $\tilde{x}$, such that

$$\partial d_\Omega(x) = \partial d_\Omega(\tilde{x}) = [0, 1] \frac{A_{j_0}}{\|A_{j_0}\|} \triangleq H_0 \quad (51)$$

for any $x \in U_{\tilde{x}} \cap bd(\Omega)$. From (50), we have

$$\eta_x = \pi_{\partial d_\Omega(x)}((-2Qx - b)/\lambda) = \pi_{H_0}((-2Qx - b)/\lambda). \quad (52)$$

Now, for any $x \in U_{\tilde{x}} \cap bd(\Omega)$, we have $\eta_{\tilde{x}} \in \partial d_\Omega(x)$, which means that $x - \tilde{x}$ is orthogonal to $\eta_{\tilde{x}}$, i.e., $(x - \tilde{x})^T \eta_{\tilde{x}} = 0$. Hence, from (43,44,50),

$$\Gamma_\theta(x) = \frac{|(x - \tilde{x})^T Q(x - \tilde{x})|^\theta}{\|2Qx + b + \lambda\eta_x\|} \quad (53)$$

for any $x \in U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda$.

Proposition 4.1. (i) For $\theta \in [1/2, 1)$, $\Gamma_\theta(x)$ defined in (53) is bounded in $U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda$. And for $\theta \in [0, 1/2)$, $\Gamma_\theta(x)$ defined in (53) is unbounded in $U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda$.

(ii) There is a neighborhood U_0 of $\tilde{x}$, such that for any $\theta \in [0, 1)$, $\Gamma_\theta(x)$ defined in (53) is bounded in $int(\Omega) \cap U_0$. ■

Proposition 4.1 will be proved in Appendix 6.3. From Proposition 4.1, if $\tilde{x} \in bd(\Omega)$ with a unique j_0 satisfying $A_{j_0}^T \tilde{x} - v_{j_0} = 0$, then $\theta_{\tilde{x}} = 1/2$.

(3-2) If $\tilde{x} \in bd(\Omega)$ with multiple j satisfying $A_j^T \tilde{x} - v_j = 0$, without loss of generality, we suppose

$$\tilde{x} \in \bigcap_{j=1}^k \{x \in bd(\Omega) : A_j^T x - v_j = 0\} \triangleq L. \quad (54)$$

We introduce the following denotations,

$$\begin{aligned} V &\triangleq \bigcup_{j=1}^k \{x \in bd(\Omega) : A_j^T x - v_j = 0\}, \\ \Omega_1 &\triangleq int(\Omega) \cup \{x \in bd(\Omega) : \eta_{\tilde{x}} \in \partial d_\Omega(x)\}, \\ \Omega_2 &\triangleq \{x \in bd(\Omega) : \eta_{\tilde{x}} \notin \partial d_\Omega(x)\}, \\ \Omega &= \Omega_1 \cup \Omega_2. \end{aligned} \quad (55)$$

From Proposition 4.1, there exists a neighborhood U of $\tilde{x}$, such that $\Gamma_\theta(x)$ is bounded in $U \cap \Omega_1 \setminus E_\lambda$ for any $\theta \in [1/2, 1)$. And $\Gamma_\theta(x)$ is unbounded in $U \cap \Omega_1 \setminus E_\lambda$ for any $\theta \in [0, 1/2)$.

If $x \in \Omega_2$, then $x \notin L$, that is $\Omega_2 \cap L = \emptyset$. By (54), there exists a neighborhood U of $\tilde{x}$, such that

$$\Omega_2 \cap U = (V \cap U) \setminus L = \bigcup_{j=1}^k \{x \in bd(\Omega) \cap U \setminus L : A_j^T x - v_j = 0\}. \quad (56)$$

For any $j = 1, 2, \dots, k$, and $x \in \{x \in bd(\Omega) \cap U \setminus L : A_j^T x - v_j = 0\}$, we have that $\partial d_\Omega(x) = [0, 1] \frac{A_j}{\|A_j\|}$ is a constant set.

Now,

$$\begin{aligned} m(\partial F_\lambda(x)) &= 2Qx + b + \lambda\eta_x \\ &= 2Q(x - \tilde{x}) - \lambda\eta_{\tilde{x}} + \lambda\eta_x \\ &= 2Q(x - \tilde{x}) - \lambda\eta_{\tilde{x}} + \pi_{\partial d_\Omega(x)}(\lambda\eta_{\tilde{x}}) - \pi_{\partial d_\Omega(x)}(\lambda\eta_{\tilde{x}}) + \lambda\eta_x. \end{aligned} \quad (57)$$

Since $\eta_{\tilde{x}} \notin \partial d_\Omega(x)$, it follows that, for any $x \in \{x \in bd(\Omega) \cap U \setminus L : A_j^T x - v_j = 0\}$, we have

$$\lambda\eta_{\tilde{x}} - \pi_{\partial d_\Omega(x)}(\lambda\eta_{\tilde{x}}) = \lambda\eta_{\tilde{x}} - \pi_{[0,1] \frac{A_j}{\|A_j\|}}(\lambda\eta_{\tilde{x}}) \neq 0. \quad (58)$$

Therefore, from the properties of the projection on closed convex set (see Aubin & Cellina (1984), p.24), we have

$$\begin{aligned} \|\pi_{\partial d_\Omega(x)}(\lambda\eta_{\tilde{x}}) - \lambda\eta_x\| &= \|\pi_{\partial d_\Omega(x)}(\lambda\eta_{\tilde{x}}) - \pi_{\partial d_\Omega(x)}(-2Qx + b)\| \\ &\leq \|2Q(x - \tilde{x})\|. \end{aligned} \quad (59)$$

Hence, by (57, 58, 59), when U is sufficiently small, we have

$$\inf \{\|m(\partial F_\lambda(x))\| : x \in bd(\Omega) \cap U \setminus L, A_j^T x - v_j = 0\} > 0.$$

That is, for any $\theta \in [0, 1]$, $\Gamma_\theta(x)$ is bounded for any $x \in \{x \in bd(\Omega) \cap U \setminus L : A_j^T x - v_j = 0\}$. From (56), $\Gamma_\theta(x)$ is also bounded in $\Omega_2 \cap U$. Thus, if $\tilde{x} \in bd(\Omega)$ with multiple j satisfying $A_j^T \tilde{x} - v_j = 0$, then $\theta_{\tilde{x}} = 1/2$.

Finally, we consider the case that $\eta_{\tilde{x}} = 0$, thus

$$\Gamma_\theta(x) = \frac{|(x - \tilde{x})^T Q(x - \tilde{x})|^\theta}{\|2Qx + b + \lambda\eta_x\|}.$$

Since $\eta_{\tilde{x}} = 0 \in \partial d_\Omega(x)$, the case that $x \in \Omega_2$ in (3-2) cannot occur. Thus, by Proposition 4.1, $\Gamma_\theta(x)$ is bounded for any $\theta \in [1/2, 1]$ and is unbounded for any $\theta \in [0, 1/2)$. According to the proof from different cases above, we know that $\theta_{\tilde{x}} = 1/2$, which completes the proof of 3).

Remark 4. According to Theorem 4.1 and Theorem 4.2, each state of the neural network (39) is either exponentially convergent, or convergent in finite time, toward an optimal solution of the convex quadratic programming (38). In the following, we introduce an example to show the effectiveness of neural network (39).

Example 1. Consider the following convex quadratic programming:

$$\begin{aligned} \text{minimize } f(x) &= x_1^2 + x_2^2 - x_1x_2 + x_1 - x_2 \\ \text{subject to } & 2x_1 + x_2 \leq 5 \\ & x_1 + 2x_2 \leq 7 \\ & x_1, x_2 \geq -1 \end{aligned} \tag{60}$$

The feasible region $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : 2x_1 + x_2 \leq 5, x_1 + 2x_2 \leq 7, x_1 \geq -1, x_2 \geq -1\}$ is bounded. And $Q = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$, $b = (1, -1)^T$. It is easy to know that $\Omega \subset B(0, 6)$. Moreover, from (16), we obtain that

$$M_f \leq \max_{x \in S} \|\nabla f(x)\| = \max_{x \in B(0, 6)} \|2Qx + b\| < 20.$$

Here we choose $\lambda = 50 > 20$ and $\sigma = 10^{-2}$. We simulate the transient behaviors of state vector $x(t)$ of neural network (39) by using the routine ode23s of MATLAB. In Figure 3, we take eight different initial points in disjoint subset regions of $\mathbb{R}^2$. Figure 3 shows that all the states with different initial points reach Ω in finite time, and converge to the equilibrium point $x^* = (-0.33, 0.33)^T \in \text{int}(\Omega)$ finally. Meanwhile, from Theorem 4.1 and Theorem 4.2, we know that the state is convergent to x^* exponentially.

Figure 3: Transient behaviors of the states of neural network (39) for solving the problem (60) with eight different initial points.

Example 2. Consider the following convex quadratic programming:

$$\begin{aligned} & \text{minimize} && f(x) = x_1^2 + x_2^2 \\ & \text{subject to} && 0 \leq x_1 \leq 3 \\ & && 0 \leq x_2 \leq 2 \end{aligned} \tag{61}$$

The feasible region $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 3, 0 \leq x_2 \leq 2\}$ is bounded. It is easy to know that $\Omega \subset B(0, 5)$, from (16), we can obtain that

$$M_f \leq \max_{x \in B(0, 5)} \|\nabla f(x)\| < 12.$$

Here we choose $\lambda = 40 > 12$ and $\sigma = 10^{-2}$. We simulate the transient behaviors of state vector $x(t)$ of neural network (39) by using the routine `ode23s` of MATLAB. Similar to Example 1, we take eight different initial points in disjoint subset regions of $\mathbb{R}^2$ in Figure 4. The figure shows that all the states with different initial points converge to the equilibrium point $x^* = (0, 0)^T \in \text{bd}(\Omega)$ finally. Meanwhile, from Theorem 4.1 and Theorem 4.2, we know that all the states converge to x^* in finite time.

Figure 4: Transient behaviors of the states of neural network (39) for solving the problem (61) with eight different initial points.

4.2. Convex quadratic programming problems with sphere constraints

In this subsection, we consider the following convex quadratic programming problems with sphere constraints,

$$\begin{aligned} & \text{minimize} && f(x) = x^T Q x + b^T x \\ & \text{subject to} && x \in \overline{B}(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\} \end{aligned} \tag{62}$$

where $x_0 \in \mathbb{R}^n$ and $r > 0$.

It is clear that $\Omega = \overline{B}(x_0, r)$ is a bounded convex subanalytic set of $\mathbb{R}^n$. Hence, based on Theorem 4.1, the convex quadratic programming (62) can be solved by neural work (14). And neural work (14) for solving the convex quadratic programming (62) can be explicitly expressed as follows,

$$\sigma \dot{x} \in -\partial F_\lambda(x) = -2Qx - b - \lambda \partial d_\Omega(x) \quad (63)$$

where $d_\Omega(x)$ and $\partial d_\Omega(x)$ are calculated explicitly in (41).

We next calculate the Lojasiewicz exponent at the equilibrium point of neural network (63).

Theorem 4.3. *Let $\tilde{x} \in \text{int}(\Omega)$ be an equilibrium point of the neural network (63), and $\theta_{\tilde{x}}$ be the Lojasiewicz exponent of F_λ at $\tilde{x}$. Then we have $\theta_{\tilde{x}} = 1/2$.*

PROOF. Since $\tilde{x} \in \text{int}(\Omega)$, there exists $r_0 > 0$ such that $B(\tilde{x}, r_0) \subseteq \Omega$. Then for any $x \in B(\tilde{x}, r_0)$, $\partial F_\lambda(x)$ defined in (63) coincides with $\partial F_\lambda(x)$ defined in (39). Hence, similar to the proof of 1) in Theorem 4.2, we have $\theta_{\tilde{x}} = 1/2$.

Example 3. *Consider the following convex quadratic programming with sphere constraint,*

$$\begin{aligned} &\text{minimize} && f(x) = 2x_1^2 + x_2^2 + x_1 + x_2 \\ &\text{subject to} && x \in \Omega = B(0, 5) \end{aligned} \quad (64)$$

Let $\Omega \subset B(0, 10)$, from (16), we can obtain that

$$M_f \leq \max_{x \in B(0, 10)} \|\nabla f(x)\| < 22.$$

Here we choose $\lambda = 60 > 22$ and $\sigma = 10^{-2}$. We simulate the transient behaviors of state vector $x(t)$ of neural network (63) by using the routine `ode23s` of MATLAB. Figure 4 shows the transient behaviors of state vector $x(t)$ with random initial points. The simulation results show that the states are globally convergent to the unique solution $x^* = (-0.25, -0.5)^T \in \text{int}(\Omega)$. Meanwhile, from Theorem 4.1 and Theorem 4.3, we know that all the states exponentially converge to x^* .

5. Conclusions

In this paper, we propose a generalized recurrent neural network described by a differential inclusion for solving convex optimization problems with general constraints. For any initial point, the state of the proposed neural network reaches the feasible set in finite time, and converges to an optimal solution of the convex optimization problem finally. Moreover, we give a quantitative evaluation of the Lojasiewicz exponent at the equilibrium point to show the convergence rate of the state of the proposed neural network. Our future work will aim at nonconvex optimization problems.

Figure 5: Transient behaviors of $x(t)$ of neural network (63) with four different initial points in Example 3.

6. Appendix

6.1. The proof of Theorem 3.2

PROOF. Since x^* is an optimal solution of the problem (1), by Lemma 3.1 and Definition 3.2, for $\lambda > M_f$, there exist $\gamma^* \in \partial f(x^*)$, $\eta^* \in \partial d_\Omega(x^*)$ such that

$$0 = \gamma^* + \lambda \eta^* \in \partial f(x^*) + \lambda \partial d_\Omega(x^*).$$

For any initial point $x(0) = x_0 \in B(x^*, R)$, by Definition 3.1 and Theorem 3.1, there exist measurable functions $\gamma(t) \in \partial f(x(t))$, $\eta(t) \in \partial d_\Omega(x(t))$, such that

$$\sigma \dot{x}(t) = -\gamma(t) - \lambda \eta(t)$$

for a.e. $t \in [0, +\infty)$. Hence, by Chain Rule (i.e., Lemma 2.1) and Proposition 2.1, for a.e. $t \in [0, +\infty)$,

$$\begin{aligned} \frac{\sigma}{2} \frac{d}{dt} \|x(t) - x^*\|^2 &= \langle x(t) - x^*, \sigma \dot{x}(t) \rangle \\ &= \langle x(t) - x^*, -\gamma(t) - \lambda \eta(t) \rangle \\ &= \langle x(t) - x^*, -\gamma(t) + \gamma^* \rangle + \lambda \langle x(t) - x^*, -\eta(t) + \eta^* \rangle \\ &\leq 0, \end{aligned} \tag{65}$$

which implies that

$$\|x(t) - x^*\|^2 \leq \|x_0 - x^*\|^2 < R^2.$$

Thus, $x(t) \in B(x^*, R)$ for any $t \geq 0$. Hence, by the definition of M_f in (16), we have

$$\lambda > M_f > \sup\{\|\xi\| : \xi \in \partial f(x(t)), t \geq 0\}. \tag{66}$$

From the Proposition 2.3, $d_\Omega(x)$ is differentiable and $\|\nabla d_\Omega(x)\| = 1$ for $x \in \mathbb{R}^n \setminus \Omega$. For any initial point $x(0) = x_0 \in B(x^*, R) \setminus \Omega$, there exists $T_0 > 0$ such that

$$x(t) \in B(x^*, R) \setminus \Omega, \text{ for } t \in [0, T_0]. \quad (67)$$

From Definition 3.1, there exists a measurable function $\gamma(t) \in \partial f(x(t))$, such that

$$\sigma \dot{x}(t) = -\gamma(t) - \lambda \nabla d_\Omega(x(t))$$

for almost everywhere $t \in [0, T_0]$. Hence, by Lemma 2.1 and (66), for a.e. $t \in [0, T_0]$,

$$\begin{aligned} \sigma \frac{d}{dt} d_\Omega(x(t)) &= \langle \nabla d_\Omega(x(t)), \sigma \dot{x}(t) \rangle \\ &= \langle \nabla d_\Omega(x(t)), -\gamma(t) - \lambda \nabla d_\Omega(x(t)) \rangle \\ &= \|\nabla d_\Omega(x(t))\| \cdot \|\gamma(t)\| - \lambda \|\nabla d_\Omega(x(t))\|^2 \\ &= \|\gamma(t)\| - \lambda \\ &\leq M_f - \lambda < 0. \end{aligned} \quad (68)$$

Integrating (68) on $[0, T_0]$, we have that

$$0 \leq d_\Omega(x(T_0)) \leq d_\Omega(x_0) + \epsilon^{-1}(M_f - \lambda)T_0. \quad (69)$$

This implies

$$T_0 \leq \frac{\sigma d_\Omega(x_0)}{\lambda - M_f} \triangleq T.$$

Now we prove that $x(t) \in \Omega$ for any $t \geq T$. We argue it by contradiction. In fact, if there exists $(\alpha, \beta) \subseteq (T, +\infty)$ which satisfies $x(\alpha) \in \Omega$, and $x(t) \in B(x^*, R) \setminus \Omega$ for any $t \in (\alpha, \beta)$. Similar to the above calculation, for any $t \in (\alpha, \beta)$, the inequality (68) still holds. Then integrating (68) on (α, β) , we have that

$$0 \leq d_\Omega(x(\beta)) \leq d_\Omega(x(\alpha)) + \sigma^{-1}(M_f - \lambda)(\beta - \alpha) < 0. \quad (70)$$

Obviously, it is impossible. Therefore, $x(t) \in \Omega$ for any $t \geq T$.

6.2. The proof of Theorem 4.1

PROOF. According to Theorem 3.3, for any initial point $x(0) = x_0 \in \mathbb{R}^n$, the state $x(t)$ of neural network (14) converges to an optimal solution $\tilde{x}$ of the nonlinear programming (37). We next prove the rest of the Theorem 4.1. From Bolte et al. (2007), we know that the distance

$$d_\Omega(x) = \inf\{\|x - a\| : a \in \Omega\}.$$

is a subanalytic function since Ω is a nonempty close convex and subanalytic subset of $\mathbb{R}^n$. Hence, $F_\lambda(x) = f(x) + \lambda d_\Omega(x)$ is also a subanalytic function. Hence, by nonsmooth Łojasiewicz inequality (i.e., Lemma 2.3) and Lemma 2.4, there exist $\theta \in [0, 1)$, $c > 0$, $\delta > 0$ such that

$$|F_\lambda(x) - F_\lambda(\tilde{x})|^\theta \leq c \|m(\partial F_\lambda(x))\| \quad (71)$$

for all $x \in B(\tilde{x}, \delta)$. Since the state $x(t)$ of neural network (14) converges to the optimal solution $\tilde{x}$, there exists $t_0 > 0$ such that $x(t) \in B(\tilde{x}, \delta)$ for all $t \geq t_0$. Then, by Theorem 3.1, for a.e. $t \geq t_0$, we have

$$|F_\lambda(x(t)) - F_\lambda(\tilde{x})|^\theta \leq c \|m(\partial F_\lambda(x(t)))\| = c \|\dot{x}(t)\|. \quad (72)$$

Without loss of generality, we assume that $\theta > 0$. In fact, if $\theta = 0$, we can take some $\theta' \in (0, 1/2)$ to replace θ . By chain rule and Theorem 3.1, for a.e. $t \in [0, +\infty)$, we have

$$\begin{aligned} \frac{d}{dt}(F_\lambda(x(t)) - F_\lambda(\tilde{x})) &= \langle m(\partial F_\lambda(x(t))), \dot{x}(t) \rangle = \langle -\sigma \dot{x}(t), \dot{x}(t) \rangle \\ &= -\sigma \|\dot{x}(t)\|^2 \leq 0. \end{aligned} \quad (73)$$

The fact that $x(t) \rightarrow \tilde{x}$ implies

$$F_\lambda(x(t)) - F_\lambda(\tilde{x}) \geq \lim_{t \rightarrow +\infty} [F_\lambda(x(t)) - F_\lambda(\tilde{x})] = 0 \quad (74)$$

for $t \in [0, +\infty)$. Hence, from (72,73,74), for a.e. $t \geq t_0$, we have

$$\begin{aligned} &\sigma^{-1} \frac{d}{dt} (F_\lambda(x(t)) - F_\lambda(\tilde{x}))^{1-\theta} \\ &= \sigma^{-1} (1-\theta) (F_\lambda(x(t)) - F_\lambda(\tilde{x}))^{-\theta} \frac{d}{dt} (F_\lambda(x(t)) - F_\lambda(\tilde{x})) \\ &= - (1-\theta) (F_\lambda(x(t)) - F_\lambda(\tilde{x}))^{-\theta} \|\dot{x}(t)\|^2 \\ &\leq - (1-\theta) c^{-1} \|\dot{x}(t)\| \end{aligned} \quad (75)$$

which implies that

$$\begin{aligned} \int_t^{+\infty} \|\dot{x}(s)\| ds &\leq \frac{c}{1-\theta} \sigma^{-1} (F_\lambda(x(t)) - F_\lambda(\tilde{x}))^{1-\theta} \\ &\leq \frac{c}{1-\theta} \sigma^{-1} c^{\frac{1-\theta}{\theta}} \|\dot{x}(t)\|^{\frac{1-\theta}{\theta}} = \frac{c^{1/\theta}}{\sigma(1-\theta)} \|\dot{x}(t)\|^{\frac{1-\theta}{\theta}}. \end{aligned} \quad (76)$$

For simplicity, we let

$$\begin{aligned} \alpha &= \frac{\sigma(1-\theta)}{c^{1/\theta}}, \quad \beta = \frac{\theta}{1-\theta}, \\ \gamma(t) &= \int_t^{+\infty} \|\dot{x}(s)\| ds. \end{aligned} \quad (77)$$

By (76), it is obtained that

$$\dot{\gamma}(t) = -\|\dot{x}(t)\| \leq -\alpha^\beta \gamma(t)^\beta. \quad (78)$$

Hence, we have the following three cases.

Case 1: $\theta \in [0, 1/2)$. In this case, it is clear that $\beta \in (0, 1)$. If there exists $t' > t_0$ such that $\gamma(t) \neq 0$ for $t \in [t_0, t')$. Then, from (78),

$$\frac{d}{dt}\gamma^{1-\beta}(t) = (1-\beta)\gamma(t)^{-\beta}\dot{\gamma}(t) \leq -(1-\beta)\alpha^\beta. \quad (79)$$

By integration on $[t_0, t')$, we have

$$0 \leq \gamma^{1-\beta}(t') \leq \gamma^{1-\beta}(t_0) - (1-\beta)\alpha^\beta t' \quad (80)$$

which means that $t' \leq \gamma^{1-\beta}(t_0)/(1-\beta)\alpha^\beta$. Meanwhile, for any $t > \gamma^{1-\beta}(t_0)/(1-\beta)\alpha^\beta$, by (78), we have $\gamma(t) = 0$. That is, $\gamma(t)$ is convergent to 0 in finite time, and an upper bound of the hit time is $\gamma^{1-\beta}(t_0)/(1-\beta)\alpha^\beta$.

Case 2: $\theta = 1/2$. In this case, $\beta = 1$. Then, from (78),

$$(e^{\alpha t}\gamma(t))' = \alpha e^{\alpha t}\gamma(t) + e^{\alpha t}\dot{\gamma}(t) \leq 0 \quad (81)$$

which means $\gamma(t) \leq \gamma(t_0)e^{\alpha(t_0-t)}$ for all $t \geq t_0$.

Case 3: $\theta \in (1/2, 1)$. In this case, it is clear that $\beta > 1$. Similar to Case 1 above, if there exists $t' > t_0$ such that $\gamma(t) \neq 0$ for $t \in [t_0, t')$, then we have

$$0 \leq \gamma^{1-\beta}(t) \leq \gamma^{1-\beta}(t_0) + (\beta-1)\alpha^\beta t \quad (82)$$

which means

$$\begin{aligned} \gamma(t) &\leq [\gamma^{1-\beta}(t_0) + (\beta-1)\alpha^\beta t]^{\frac{1}{1-\beta}} = [\gamma^{1-\beta}(t_0) + (\beta-1)\alpha^\beta t]^{(1-\theta)/(1-2\theta)} \\ &\leq k(t+1)^{(1-\theta)/(1-2\theta)} \end{aligned} \quad (83)$$

where $k = \max\{\gamma^{1-\beta}(t_0), (\beta-1)\alpha^\beta\}^{(1-\theta)/(1-2\theta)}$.

On the other hand, we get that

$$\begin{aligned} \|x(t) - \tilde{x}\| &= \|x(t) - \lim_{s \rightarrow +\infty} x(s)\| \\ &= \left\| \int_t^{+\infty} \dot{x}(s) ds \right\| \leq \int_t^{+\infty} \|x(s)\| ds = \gamma(t). \end{aligned} \quad (84)$$

Then, according to (**Case 1,2,3**) and (84), we can obtain the conclusions of the theorem.

6.3. The proof of Proposition 4.1

(i) First we introduce two denotations as follows,

$$\begin{aligned} \Psi_1 &= \{x \in U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda : \eta_x = \eta_{\tilde{x}}\}, \\ \Psi_2 &= \{x \in U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda : \eta_x \neq \eta_{\tilde{x}}\}, \\ \Psi_1 \cup \Psi_2 &= U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda. \end{aligned} \quad (85)$$

where η_x is from (52), and $\eta_{\tilde{x}}$ is from (43).

If $x \in \Psi_1$, then from (43),

$$2Qx + b + \lambda\eta_x = 2Qx + b - 2Q\tilde{x} - b = 2Q(x - \tilde{x}).$$

So, by (53),

$$\Gamma_\theta(x) = \frac{|(x - \tilde{x})^T Q(x - \tilde{x})|^\theta}{\|2Q(x - \tilde{x})\|}. \quad (86)$$

Thus we can proceed as the proof of 1) in Theorem 4.2, to conclude that $\Gamma_\theta(x)$ is bounded in Ψ_1 for any $\theta \in [1/2, 1)$ and is unbounded for any $\theta \in [0, 1/2)$.

On the other hand, by the fact $\eta_x \rightarrow \eta_{\tilde{x}} \neq 0$ as $x \rightarrow \tilde{x}$, there exists $\delta > 0$, such that $\eta_x \neq 0$ for any $x \in B(\tilde{x}, \delta)$.

if $x \in \Psi_2$, based on the projection properties and $[0, 1]a_i/\|a_i\|$ is a segment, we have,

$$\pi_{H_0}((-2Qx - b)/\lambda) = \pi_{\mathbb{R}H_0}((-2Qx - b)/\lambda) \quad (87)$$

where H_0 is from (51). In fact, if (87) does not hold, then we have $\eta_{\tilde{x}} = \eta_x$ or $\eta_{\tilde{x}} = 0$. Obviously, it contradicts with above assumptions. We let

$$H = \mathbb{R}H_0 = \left\{ h \frac{A_{j_0}}{\|A_{j_0}\|} : h \in (-\infty, +\infty) \right\}$$

then π_H is a linear operator. It is clear that $(\pi_H - Id)(2Q\tilde{x} + b) = 0$ since $2Q\tilde{x} + b = \lambda\eta_{\tilde{x}} \in H$. Thus, for any $x \in \Psi_2$, from (52) we have

$$\begin{aligned} 0 &\neq 2Qx + b + \lambda\eta_x \\ &= 2Qx + b + \lambda\pi_{H_0}\left(\frac{-2Qx - b}{\lambda}\right) = 2Qx + b + \pi_H(-2Qx - b) \\ &= (\pi_H - Id)(-2Qx - b) = (\pi_H - Id)(-2Qx - b + 2Q\tilde{x} + b) \\ &= (\pi_H - Id)(-2Q(x - \tilde{x})) = Q_1(x - \tilde{x}) \end{aligned} \quad (88)$$

where $Q_1 = (\pi_H - Id)(-2Q) \in \mathbb{R}^{n \times n}$ is a linear nonzero operator. Similar to the proof of 1) in Theorem 4.2, let $\mathfrak{R}_1 = \ker(Q_1)$ and $\mathfrak{R}_1^\perp = \ker^\perp(Q_1)$. It is noted that $\mathfrak{R}_1 = \ker(Q_1) \neq \mathbb{R}^n$, then we have $(x - \tilde{x}) = \pi_{\mathfrak{R}_1}(x - \tilde{x}) + \pi_{\mathfrak{R}_1^\perp}(x - \tilde{x})$, and $\pi_H(x - \tilde{x}) = 0$ since $x - \tilde{x}$ is orthogonal to $\eta_{\tilde{x}}$. By (88), we have $\pi_{\mathfrak{R}_1^\perp}(x - \tilde{x}) \neq 0$. Therefore, we obtain

$$\begin{aligned} |(x - \tilde{x})^T Q(x - \tilde{x})| &= \left| \frac{1}{2} (\pi_{\mathfrak{R}_1^\perp}(x - \tilde{x}))^T Q_1 \pi_{\mathfrak{R}_1^\perp}(x - \tilde{x}) \right| \\ &\leq \frac{1}{2} \lambda_M^1 \|\pi_{\mathfrak{R}_1^\perp}(x - \tilde{x})\|^2 \end{aligned}$$

where $\lambda_M^1 = \max\{|\lambda_i| : \lambda_i \text{ is an eigenvalue of } Q_1\}$. In addition, we also have

$$\begin{aligned} \|2Qx + b + \lambda\eta_x\| &= \|Q_1(x - \tilde{x})\| \\ &\geq \lambda_m^1 \|\pi_{\mathfrak{R}_1^\perp}(x - \tilde{x})\| > 0 \end{aligned}$$

where $\lambda_m^1 = \min\{|\lambda_i| : \lambda_i \neq 0 \text{ is an eigenvalue of } Q_1\}$. So, we have

$$\Gamma_\theta(x) \leq \frac{(1/2\lambda_M^1)^\theta}{\lambda_m^1} \|\pi_{\mathfrak{R}_1^\perp}(x - \tilde{x})\|^{2\theta-1}.$$

Therefore, $\Gamma_\theta(x)$ is bounded in Ψ_2 for any $\theta \in [1/2, 1)$.

On the other hand, when $\theta \in [0, 1/2)$, for any $x \in \Psi_2$, we have

$$\Gamma_\theta(x) \geq \frac{(1/2\lambda_m^1)^\theta}{\lambda_M^1} \|\pi_{\mathbb{R}_1^\perp}(x - \tilde{x})\|^{2\theta-1},$$

which means that $\Gamma_\theta(x)$ is unbounded in Ψ_2 for any $\theta \in [0, 1/2)$.

Hence, for $\theta \in [1/2, 1)$, $\Gamma_\theta(x)$ is bounded in $U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda$. And for $\theta \in [0, 1/2)$, $\Gamma_\theta(x)$ is unbounded in $U_{\tilde{x}} \cap bd(\Omega) \setminus E_\lambda$.

(ii) If $x \in \text{int}(\Omega)$, then $\eta_x = 0$. Therefore, from (44), we obtain

$$\Gamma_\theta(x) = \frac{|(x - \tilde{x})^T Q(x - \tilde{x}) - \lambda(x - \tilde{x})^T \eta_{\tilde{x}}|^\theta}{\|2Qx + b\|}$$

Since $2Qx + b \rightarrow 2Q\tilde{x} + b = -\lambda\eta_{\tilde{x}} \neq 0$ as $x \rightarrow \tilde{x}$, for any $\theta \in [0, 1)$, we have

$$\lim_{x \rightarrow \tilde{x}} \Gamma_\theta(x) = 0.$$

Hence, there is some neighborhood U_0 of $\tilde{x}$, such that for any $\theta \in [0, 1)$, $\Gamma_\theta(x)$ is bounded in $\text{int}(\Omega) \cap U_0$.

References

- Aubin, J. P., & Cellina, A. (1984). *Differential inclusions*. Berlin, Germany: Springer-Verlag.
- Bian, W., & Xue, X. (2009). Subgradient-based neural networks for nonsmooth nonconvex optimization problems. *IEEE Trans. Neural Networks*, 20, 1024–1038.
- Bolte, J., Daniilidis, A., & Lewis, A. (2007). The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17, 1205–1223.
- Borwein, J. M., & Giles, J. (1987). The proximal normal formula in banach space. *Transactions of the American Mathematical Society*, 302, 371–381.
- Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Cheng, L., Hou, Z.-G., Lin, Y., Tan, M., Zhang, W. C., & Wu, F.-X. (2011). Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks. *IEEE Trans. Neural Networks*, 22, 714–726.
- Cochocki, A., & Unbehauen, R. (1993). *Neural networks for optimization and signal processing*. John Wiley & Sons, Inc.
- F.Clarke (1983). *Optimization and nonsmooth analysis*,. Wiley, New York.

- Forti, M. (2008). Convergence of a subclass of Cohen–Grossberg neural networks via the Łojasiewicz inequality. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *38*, 252–257.
- Forti, M., Nistri, P., & Quincampoix, M. (2004). Generalized neural network for nonsmooth nonlinear programming problems. *IEEE Trans. Circuits and Systems I: Regular Papers*, *51*, 1741–1754.
- Forti, M., Nistri, P., & Quincampoix, M. (2006). Convergence of neural networks for programming problems via a nonsmooth Łojasiewicz inequality. *IEEE Trans. Neural Networks*, *17*, 1471–1486.
- Forti, M., & Tesi, A. (2004). Absolute stability of analytic neural networks: An approach based on finite trajectory length. *IEEE Trans. Circuits and Systems I: Regular Paper*, *51*, 2460–2469.
- Forti, M., & Tesi, A. (2006). The Łojasiewicz exponent at an equilibrium point of a standard CNN is $1/2$. *International Journal of Bifurcation and Chaos*, *16*, 2191–2205.
- Gao, X.-B., & Liao, L.-Z. (2010). A new one-layer neural network for linear and quadratic programming. *IEEE Trans. Neural Networks*, *21*, 918–929.
- Hosseini, A. (2016). A non-penalty recurrent neural network for solving a class of constrained optimization problems. *Neural Networks*, *73*, 10–25.
- Kennedy, M. P., & Chua, L. O. (1988). Neural networks for nonlinear programming. *IEEE Trans. Circuits and Systems*, *35*, 554–562.
- Li, G., Song, S., Wu, C., & Du, Z. (2006). A neural network model for non-smooth optimization over a compact convex subset. In *Advances in Neural Networks-ISNN 2006* (pp. 344–349). Springer.
- Li, G., Yan, Z., & Wang, J. (2014). A one-layer recurrent neural network for constrained nonsmooth invex optimization. *Neural Networks*, *50*, 79–89.
- Liao, L.-Z., Qi, H., & Qi, L. (2004). Neurodynamical optimization. *Journal of Global Optimization*, *28*, 175–195.
- Liu, Q., & Wang, J. (2008a). A one-layer recurrent neural network with a discontinuous activation function for linear programming. *Neural Computation*, *20*, 1366–1383.
- Liu, Q., & Wang, J. (2008b). A one-layer recurrent neural network with a discontinuous hard-limiting activation function for quadratic programming. *IEEE Trans. Neural Networks*, *19*, 558–570.
- Liu, Q., Yang, S., & Wang, J. (2016). A collective neurodynamic approach to distributed constrained optimization. *IEEE Trans. Neural Networks and Learning Systems*, *in press*.

- Lojasiewicz, S. (1963). Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 87-89.
- Lu, W., & Wang, J. (2008). Convergence analysis of a class of nonsmooth gradient systems. *IEEE Trans. Circuits and Systems I: Regular Papers*, 55, 3514–3527.
- M. S. Bazaraa, H. D. S., & Shetty, C. M. (1993). *Nonlinear Programming*. New York: Theory and Algorithms.
- Qin, S., & Xue, X. (2010). Dynamical analysis of neural networks of subgradient system. *IEEE Trans. Automatic Control*, 55, 2347–2352.
- Qin, S., & Xue, X. (2015). A two-layer recurrent neural network for nonsmooth convex optimization problems. *IEEE Trans. Neural Networks and Learning Systems*, 26, 1149–1160.
- Rockafellar, R. T., & Wets, R. J.-B. (2009). *Variational analysis* volume 317. Springer Science & Business Media.
- Tank, D., & Hopfield, J. (1986). Simple ‘neural’ optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Trans. Circuits and Systems*, 33, 533–541.
- Wang, J. (1994). A deterministic annealing neural network for convex programming. *Neural Networks*, 7, 629–641.
- Xia, Y., & Wang, J. (2005). A recurrent neural network for solving nonlinear convex programs subject to linear constraints. *IEEE Trans. Neural Networks*, 16, 379–386.
- Xue, X., & Bian, W. (2008). Subgradient-based neural networks for nonsmooth convex optimization problems. *IEEE Trans. Circuits and Systems I: Regular Paper*, 55, 2378–2391.
- Zhang, S., & Constantinides, A. (1992). Lagrange programming neural networks. *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, 39, 441–452.