Neural Networks 61 (2015) 10–21

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A one-layer recurrent neural network for constrained nonconvex optimization☆

Guocheng Li [a], Zheng Yan [b], Jun Wang [b,c,*]

[a] Department of Mathematics, Beijing Information Science and Technology University, Beijing, China

[b] Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

[c] School of Control Science and Engineering, Dalian University of Technology, Dalian, Liaoning, China

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form 22 August 2014
Accepted 18 September 2014
Available online 28 September 2014

Keywords:
Nonconvex optimization
Exact penalty function
Recurrent neural network
Finite time convergence

ABSTRACT

In this paper, a one-layer recurrent neural network is proposed for solving nonconvex optimization problems subject to general inequality constraints, designed based on an exact penalty function method. It is proved herein that any neuron state of the proposed neural network is convergent to the feasible region in finite time and stays there thereafter, provided that the penalty parameter is sufficiently large. The lower bounds of the penalty parameter and convergence time are also estimated. In addition, any neural state of the proposed neural network is convergent to its equilibrium point set which satisfies the Karush–Kuhn–Tucker conditions of the optimization problem. Moreover, the equilibrium point set is equivalent to the optimal solution to the nonconvex optimization problem if the objective function and constraints satisfy given conditions. Four numerical examples are provided to illustrate the performances of the proposed neural network.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In this paper, the following constrained nonconvex minimization problem is considered:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i \in I = \{1, 2, \ldots, m\}, \end{array} \tag{1}$$

where $x \in R^n$ is the decision vector; f and $g_i, : R^n \to R$ $(i \in I)$ are continuously differentiable functions, but not necessarily convex. The feasible region

$$\mathcal{F} = \{x \in R^n : g_i(x) \leq 0,\ i \in I\}$$

is assumed to be a nonempty set. We denote by $\mathcal{G}$ the set of global solutions of problem (1) as,

$$\mathcal{G} = \{x \in \mathcal{F} : f(y) \geq f(x),\ \forall\, y \in \mathcal{F}\}.$$

Many problems in engineering applications can be formulated as dynamic optimization problems such as kinematic control of redundant robot manipulators (Wang, Hu, & Jiang, 1999), nonlinear model predictive control (Piche, Sayyar-Rodsari, Johnson, & Gerules, 2000; Yan & Wang, 2012), hierarchical control of interconnected dynamic systems (Hou, Gupta, Nikiforuk, Tan, & Cheng, 2007), compressed sensing in adaptive signal processing (Balavoine, Romberg, & Rozell, 2012), and so on. For example, real-time motion planning and control of redundant robot manipulators can be formulated as constrained dynamic optimization problems with nonconvex objective functions for simultaneously minimizing kinetic energy and maximizing manipulability. Similarly, in nonlinear and robust model predictive control, optimal control commands have to be computed with a moving time window by repetitively solving constrained optimization problems with nonconvex objective functions for error and control variation minimization, and robustness maximization. The difficulty of dynamic optimization is significantly amplified when the optimal solutions have to be obtained in real time, especially in the presence of uncertainty. In such applications, compared with traditional numerical optimization algorithms, neurodynamic optimization approaches based on recurrent neural networks have several unique advantages. Recurrent neural networks can be physically implemented in designated hardware/firmware, such as very-large-scale integration (VLSI) reconfigurable analog chips, optical chips,

☆ The work described in the paper was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China under Grant CUHK416812E; and by the National Natural Science Foundation of China under Grant 61273307.

* Corresponding author at: Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. Tel.: +852 39438472.

E-mail addresses: xyliguocheng@sohu.com (G. Li), zyan@mae.cuhk.edu.hk (Z. Yan), jwang@mae.cuhk.edu.hk (J. Wang).

http://dx.doi.org/10.1016/j.neunet.2014.09.009
0893-6080/© 2014 Elsevier Ltd. All rights reserved.

graphic processing units (GPU), field programmable gate array (FPGA), digit signal processor (DSP), and so on. Recent technological advances make the design and implementation of neural networks more feasible at a more reasonable cost (Asai, Kanazawa, & Amemiya, 2003).

Since the pioneering work of Hopfield neural networks (Hopfield & Tank, 1985; Tank & Hopfield, 1986), neurodynamic optimization has achieved great success in the past three decades. For example, a deterministic annealing neural network was proposed for solving convex programming problems (Wang, 1994), a Lagrangian network was developed for solving convex optimization problems with linear equality constraints based on the Lagrangian optimality conditions (Xia, 2003), the primal–dual network (Xia, 1996), the dual network (Xia, Feng, & Wang, 2004), and the simplified dual network (Liu & Wang, 2006) were developed for solving convex optimization problems based on the Karush–Kuhn–Tucker optimality conditions, projection neural networks were developed for constrained optimization problems based on the projection method (Gao, 2004; Hu & Wang, 2007; Liu, Cao, & Chen, 2010; Xia, Leung, & Wang, 2002). In recent years, neurodynamic optimization approaches have been extended to nonconvex and generalized convex optimization problems. For example, a Lagrangian neural network was proposed for nonsmooth convex optimization by using the Lagrangian saddle-point theorem (Cheng et al., 2011), a recurrent neural network with global attractivity was proposed for solving nonsmooth convex optimization problems (Bian & Xue, 2013), several neural networks were developed for nonsmooth pseudoconvex or quasiconvex optimization using the Clarke's generalized gradient (Guo, Liu, & Wang, 2011; Hosseini, Wang, & Hosseini, 2013; Hu & Wang, 2006; Liu, Guo, & Wang, 2012; Liu & Wang, 2013). In addition, various neural networks with finite-time convergence property were developed (Bian & Xue, 2009; Forti, Nistri, & Quincampoix, 2004, 2006; Xue & Bian, 2008).

Despite the enormous success, neurodynamic optimization approaches would reach their solvability limits at constrained optimization problems with unimodal objective functions and are important for global optimization with general nonconvex objective functions. Little progress has been made on nonconvex optimization in the neural network community. Instead of seeking global optimal solutions, a more attainable and meaningful goal is to design neural networks for searching critical points (e.g., Karush–Kuhn–Tucker points) of nonconvex optimization problems. Xia, Feng, and Wang (2008) proposed a neural network for solving nonconvex optimization problems with inequality constraints, whose equilibrium points correspond to the KKT points. But the condition that the Hessian matrix of the associated Lagrangian function is positive semidefinite for the global convergence is too strong. In this paper, a one-layer recurrent neural network based on an exact penalty function method is proposed for searching KKT points of nonconvex optimization problems with inequality constraints. The contribution of this paper can be summarized as follows. (1) State of the proposed neural network is convergent to the feasible region in finite time and stays there thereafter, with a sufficiently large penalty parameter; (2) the proposed neural network is convergent to its equilibrium point set; (3) any equilibrium point x^* of the proposed neural network corresponds to a KKT twofold (x^*, λ^*) of the nonconvex problem and vice versa; (4) if the objective function and the constraint functions meet one of the following conditions: (a) the objective function and the constraint functions are convex; (b) the objective function is pseudoconvex and the constraint functions are quasiconvex, then the state of the proposed network converges to the global optimal solution. If the objective function and the constraint functions are invex with respect to the same kernel, then the state of the proposed network converges to optimal solution set. Hence, the results presented in Li, Yan, and Wang (2014) can be viewed as special cases of this paper.

The remainder of this paper is organized as follows. Section 2 introduces some definitions and preliminary results. Section 3 discusses an exact penalty function. Section 4 presented a neural network model and analyzed its convergent properties. Section 5 provides simulation results. Finally, Section 6 concludes this paper.

2. Preliminaries

In this section, we present definitions and properties concerning the set-valued analysis, nonsmooth analysis, and the generalized convex function which are needed in the remainder of the paper. We refer readers to Aubin and Cellina (1984), Cambini and Martein (2009), Clarke (1969), Filippov (1988) and Pardalos (2008) for a more thorough research.

Let R^n be real Euclidean space with the scalar product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$, $x,\ y \in R^n$ and its related norm $\|x\| = [\sum_{i=1}^n x_i^2]^{\frac{1}{2}}$. Let $x \in R^n$ and $A \subset R^n$, $\mathrm{dist}(x, A) = \inf_{y\in A} \|x - y\|$ is the distance of x from A.

Definition 1. $F : R^n \hookrightarrow R^n$ is called a set-valued map, if to each point $x \in R^n$, there corresponds to a nonempty closed set $F(x) \subset R^n$.

Definition 2. Let F be a set-valued map. F is said to be upper semicontinuous at $x_0 \in R^n$ if $\forall\,\varepsilon > 0$, $\exists\,\delta > 0$ such that $\forall x \in (x_0 + \delta\mathcal{B})$, $F(x) \subset F(x_0) + \varepsilon\mathcal{B}$, where $\mathcal{B} = \mathcal{B}(0, 1)$ is the ball centered at the origin with radius 1. F is upper semicontinuous if it is so at every $x_0 \in R^n$.

A solution $x(t)$ of a differential inclusion is an absolutely continuous function, the derivative $\dot{x}(t)$ is only defined almost everywhere, so that its limit when $t \to \infty$ is not well defined. The concepts of limit and cluster points to a measurable function should be defined. Let $\mu(A)$ denote the Lebesgue measure of a measurable subset $A \subset R$.

Definition 3. Let $x : [0, \infty) \to R^n$ be a measurable function. $x^* \in R^n$ is the almost limit of $x(\cdot)$ if when $t \to \infty\ \forall\,\varepsilon > 0, \exists\,T > 0$ such that

$$\mu\left\{t : \|x(t) - x^*\| > \varepsilon,\ t \in [0, \infty)\right\} < \varepsilon.$$

It can be written as $x^* = \mu - \lim_{t\to\infty} x(t)$. x^* is an almost cluster point of $x(\cdot)$ if when $t \to \infty\ \forall\,\varepsilon > 0$,

$$\mu\left\{t : \|x(t) - x^*\| \le \varepsilon,\ t \in [0, \infty)\right\} = \infty.$$

The following propositions show that the usual concepts of limit and cluster are particular cases of almost limit and almost cluster point (Aubin & Cellina, 1984).

Proposition 1. *The limit x^* of $x : [0,\ \infty) \to R^n$ is an almost limit point. If $x(\cdot)$ is uniformly continuous, any cluster point x^* of $x(\cdot)$ is an almost cluster point.*

Proposition 2. *An almost limit x^* of a measurable function $x : [0,\ \infty) \to R^n$ is a unique almost cluster point. If $x(\cdot)$ has a unique almost cluster point x^* and $\{x(t) : t \in [0,\ \infty)\}$ is a bounded subset of R^n, $\mu - \lim_{t\to\infty} x(t) = x^*$.*

Proposition 3. *Let K be a compact subset of R^n and $x : [0,\ \infty) \to K$ be a measurable function, there exists an almost cluster $x^* \in K$ of $x(\cdot)$ when $t \to \infty$.*

Definition 4. Function $f : R^n \to R$ is said to be Lipschitz near $x \in R^n$ if there exist positive number k and ε such that $|f(x_2) - f(x_1)| \le k\|x_2 - x_1\|$, for all $x_1,\ x_2 \in x + \varepsilon\mathcal{B}$. If f is Lipschitz near any point of its domain, then it is said to be locally Lipschitz.

Suppose that f is Lipschitz near $x \in R^n$, the generalized directional derivative of f at x in the direction $v \in R^n$ is given by

$$f^0(x; v) = \lim_{y \to x} \sup_{t \to 0^+} \frac{f(y + tv) - f(y)}{t}.$$

The quantity $f^0(x; v)$ is well defined and finite. Furthermore, Clarke's generalized gradient of f at x is defined as

$$\partial f(x) = \{\xi \in R^n : f^0(x; v) \geq \langle v, \xi \rangle, \text{ for all } v \in R^n\}.$$

By accounting for the properties of f^0, some properties of $\partial f(x)$ can be obtained as follows.

Proposition 4 (Clarke, 1969). *Let $f : R^n \to R$ be Lipschitz near x with positive number k, then:*

(a) *$\partial f(x)$ is a nonempty, convex, compact subset of R^n, and $|\partial f(x)| = \sup\{\|\xi\| : \xi \in \partial f(x)\} \leq k$.*
(b) *$x \hookrightarrow \partial f(x)$ is upper semicontinuous*
(c) *For every $v \in R^n$,*

$$f^0(x; v) = \max\{\langle v, \xi \rangle : \xi \in \partial f(x)\}.$$

When f is strictly differentiable at x, $\partial f(x)$ reduces to the gradient $\nabla f(x)$. When f is convex, $\partial f(x)$ coincides with classical subdifferential of convex analysis, which is

$$\xi \in \partial f(x) \Leftrightarrow \forall y \in R^n, \quad f(y) \geq f(x) + \langle \xi, y - x \rangle.$$

The concept of a regular function is defined as follows (Clarke, 1969).

Definition 5. A function $f : R^n \to R$ which is Lipschitz near x is said to be regular at x provided the following conditions hold:

(a) For all $v \in R^n$, the usual one-sided directional derivative $f'(x; v) = \lim_{t \to 0^+}[f(x + tv) - f(x)]/t$ exists;
(b) For every $v \in R^n, f'(x; v) = f^0(x; v)$.

Regular functions form a rather wide set, and several classes of them are presented in Clarke (1969) as follows.

Proposition 5. *Let $f : R^n \to R$ be Lipschitz near x.*

(a) *If f is strictly differentiable at x, then f is regular at x;*
(b) *If f is convex function, then f is regular at x.*

Proposition 6. *Let $f : R^n \to R$ be continuously differentiable. Then $\max\{0, f(x)\}$ is a regular function, its Clarke's generalized gradient as follows:*

$$\partial \max\{0, f(x)\} = \begin{cases} \nabla f(x), & \text{for } f(x) > 0; \\ [0, 1]\, \nabla f(x), & \text{for } f(x) = 0; \\ 0, & \text{for } f(x) < 0. \end{cases}$$

Proposition 7. *If $f : R^n \to R$ is regular at $x(t)$ and $x : R \to R^n$ is differentiable at t and Lipschitz near t, then*

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

Proposition 8. *Let $f_i : R^n \to R$ be regular near x $(i = 1, 2, \ldots, m)$, then*

$$\partial \left(\sum_{i=1}^{m} f_i \right)(x) = \sum_{i=1}^{m} \partial f_i(x).$$

Definition 6. A differentiable function f, defined on an open convex set $\mathcal{D} \subset R^n$, is called quasiconvex if

$$x_1, \ x_2 \in \mathcal{D}, \quad f(x_1) \geq f(x_2) \Rightarrow \nabla f(x_1)^T (x_2 - x_1) \leq 0.$$

Definition 7. A differentiable function f, defined on an open convex set $\mathcal{D} \subset R^n$, is called pseudoconvex if

$$x_1, \ x_2 \in \mathcal{D}, \quad f(x_1) > f(x_2) \Rightarrow \nabla f(x_1)^T (x_2 - x_1) < 0.$$

Definition 8. The differentiable function f defined on an open set $\mathcal{D} \subset R^n$ is invex if there exists a vector function $\eta(x, y)$ defined on $\mathcal{D} \times \mathcal{D}$ such that

$$f(x) - f(y) \geq \eta^T(x, y) \nabla f(y), \quad \forall x, \ y \in \mathcal{D},$$

where $\eta(x, y)$ is called a kernel of f.

Proposition 9 (Pardalos, 2008). *Let $f : R^n \to R$ be a differentiable invex function, x is a global minimizer of f if and only if $\nabla f(x) = 0$.*

Proposition 10 (Pardalos, 2008). *Let $f_1, f_2, \ldots, f_m : R^n \to R$ be differentiable invex functions with respect to the same kernel then a linear combination of $f_1, f_2, \ldots, f_m$, with nonnegative coefficients is invex with respect to the same kernel.*

Definition 9. A function $f : R^n \to R$ is a radially unbounded function if

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

Proposition 11. *A function $f : R^n \to R$ is radially unbounded if and only if $\forall \alpha \in R$, level set $L(\alpha) = \{x \in R^n : f(x) \leq \alpha\}$ is bounded.*

Let $\mathcal{A} \subset R^n$, we denote the closure of $\mathcal{A}$ by cl$\mathcal{A}$, the interior of $\mathcal{A}$ by int$\mathcal{A}$, the border of $\mathcal{A}$ by bd$\mathcal{A}$, and the complement of $\mathcal{A}$ by $\mathcal{A}^c$.

3. Exact penalty function

In this section, an appropriate neighborhood of the feasible region $\mathcal{F}$ is given and an exact penalty function is defined and analyzed on this neighborhood.

Consider the following function:

$$V(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\}.$$

For any $x \in R^n$, we define the index sets:

$$I_0(x) = \{i : g_i(x) = 0, \ i \in I\},$$
$$I_+(x) = \{i : g_i(x) > 0, \ i \in I\},$$
$$I_-(x) = \{i : g_i(x) < 0, \ i \in I\}.$$

By Propositions 6 and 8, the Clarke's generalized gradient of $V(x)$ is as follows:

$$\partial V(x) = \sum_{i \in I_+(x)} \nabla g_i(x) + \sum_{i \in I_0(x)} [0, 1] \nabla g_i(x). \tag{2}$$

To ensure exact penalty, the following assumptions on constraint functions in problem (1) are needed.

Assumption 1. For any $z \in R^n$ and

$$\{i_1, \ i_2, \ldots, i_q\} \subseteq I, \quad i_1 < i_2 < \cdots < i_q,$$

if

$$g_{i_1}(z) = g_{i_2}(z) = \cdots = g_{i_q}(z) = 0,$$

then $\{\nabla g_{i_1}(z), \ \nabla g_{i_2}(z), \ldots, \nabla g_{i_q}(z)\}$ is linearly independent.

Assumption 2. $\mathcal{F} = \mathrm{cl}(\mathrm{int}\mathcal{F})$.

Assumption 3. $f(x) + V(x)$ is radially unbounded.

The following lemma plays an important role in convergence analysis.

Lemma 1. *Suppose that (1) satisfies Assumptions 1 and 2.* $\exists R > 0$ *and* $m_g > 0$ *such that*

$$\min_{x\in(\mathcal{F}+R\mathcal{B})\setminus \mathrm{int}\mathcal{F}} dist(0, \partial V(x)) \geq m_g > 0. \tag{3}$$

Proof. First, we prove that $\forall\, x \in \mathrm{bd}\mathcal{F}$, $0 \notin \partial V(x)$. If not, then $\exists\, \tilde{x} \in \mathrm{bd}\mathcal{F}$ such that, $\mathrm{dist}(0, \partial V(\tilde{x})) = 0$. Since $x \to \mathrm{dist}(0, \partial V(x))$ is a lower semicontinuous real-valued function and $\mathcal{F} = \mathrm{cl}(\mathrm{int}\mathcal{F})$, we can take $\{x_k\} \subset \mathcal{F}^c$, $x_k \to \tilde{x}$ and $\eta_k \in \partial V(x_k)$ such that $\lim_{k\to\infty} \|\eta_k\| = 0$. There exist $\alpha_{k_i} \in [0, 1]$ $(i \in I_0(x_k))$ such that

$$\eta_k = \sum_{i\in I_+(x_k)} \nabla g_i(x_k) + \sum_{i\in I_0(x_k)} \alpha_{k_i} \nabla g_i(x_k). \tag{4}$$

Extracting a subsequence and re-indexing, we assume without loss of generality that, for all natural numbers k, $I_+(x_k) = I_+(x_1)$, $I_0(x_k) = I_0(x_1)$ and $\lim_{k\to\infty} \alpha_{k_i} = \alpha_i$. Taking $k \to \infty$ in (4), then

$$\sum_{i\in I_+(x_1)} \nabla g_i(\tilde{x}) + \sum_{i\in I_0(x_1)} \alpha_i \nabla g_i(\tilde{x}) = 0. \tag{5}$$

Note that $I_+(x_1) \neq \emptyset$ and $g_i(\tilde{x}) = 0$ $(i \in I_+(x_1) \cup I_0(x_1))$, (5) is a contradiction to Assumption 1, thus $\forall\, x \in \mathrm{bd}\mathcal{F}$, $0 \notin \partial V(x)$.

Next, by Proposition 4.1 in Nistri and Quincampoix (2005), there exist $R > 0$ and $m_g > 0$ such that

$$\min_{x\in(\mathcal{F}+R\mathcal{B})\setminus \mathrm{int}\mathcal{F}} \mathrm{dist}(0, \partial V(x)) \geq m_g > 0. \quad \square$$

Since $V(x)$ is continuous, $\exists\, r > 0$ such that

$$\mathcal{D} := \{x \in R^n : V(x) < r\} \subseteq \mathcal{F} + R\mathcal{B}.$$

For problem (1), a penalty function is commonly defined as follows:

$$E_\sigma(x) = f(x) + \frac{1}{\sigma} V(x),$$

where $\sigma > 0$ is penalty parameter. In view of Assumption 3, the penalty function $E_\sigma(x)$ is continuous and radially unbounded.

Consider the following problem:

$$\min E_\sigma(x), \quad x \in \mathcal{D}. \tag{6}$$

The Clarke's generalized gradient of $E_\sigma(x)$ is as follows:

$$\partial E_\sigma(x) = \nabla f(x) + \frac{1}{\sigma} \partial V(x).$$

Since $\mathcal{D}$ is an open set, any local solution of problem (6), provided it exists, is unconstrained; thus problem (6) can be considered as an essentially unconstrained problem. The sets of global solutions of problem (6) is denoted by $\mathcal{G}(\sigma)$:

$$\mathcal{G}(\sigma) = \{x \in \mathcal{D} : E_\sigma(y) \geq E_\sigma(x),\ \forall y \in \mathcal{D}\}.$$

Definition 10. The function $E_\sigma(x)$ is an exact penalty function for problem (1) with respect to the set $\mathcal{D}$ if there exists an $\sigma^* > 0$ such that for all $\sigma \in (0, \sigma^*]$, $\mathcal{G}(\sigma) = \mathcal{G}$.

The next theorem establishes a sufficient condition for $E_\sigma(x)$ to be an exact penalty function in the sense of Definition 9.

Theorem 1. $E_\sigma(x)$ *is an exact penalty function for (1) with respect to the set* $\mathcal{D}$.

Proof. By the compactness of $\mathrm{cl}\mathcal{D}$ and the continuity of $E_\sigma(x)$, for all $\sigma > 0$, $E_\sigma(x)$ admits a global minimum point on $\mathrm{cl}\mathcal{D}$. We show first that there exists a $\sigma_1^* > 0$ such that, for all $\sigma \in (0, \sigma_1^*]$ we have $\mathcal{G}(\sigma) \neq \emptyset$. Suppose that this assertion is false. Then, for any integer k there must exist a $\sigma_{1k} \leq 1/k$ and $y^{(k)} \in \mathrm{bd}\mathcal{D}$ such that

$$E_{\sigma_{1k}}(y^{(k)}) = \inf_{x\in \mathrm{cl}\mathcal{D}} E_{\sigma_1 k}(x).$$

There exists a convergent subsequence, which we relabel $\{y^{(k)}\}$ such that, $\lim_{k\to\infty} y^{(k)} = \bar{y}$, and $\bar{y} \in \mathrm{bd}\mathcal{D}$. For each natural number k,

$$\begin{aligned}\inf_{x\in\mathcal{F}} f(x) &= \inf_{x\in\mathcal{F}} E_{\sigma_{1k}}(x) \geq \inf_{x\in \mathrm{cl}\mathcal{D}} E_{\sigma_{1k}}(x) \\ &= E_{\sigma_{1k}}(y^{(k)}) = f(y^{(k)}) + \frac{1}{\sigma_{1k}} V(y^{(k)}) \\ &\geq \inf_{x\in \mathrm{cl}\mathcal{D}} f(x) + \frac{1}{\sigma_{1k}} V(y^{(k)}).\end{aligned} \tag{7}$$

It follows from (7) that

$$\begin{aligned}\max\{g_i(\bar{y}), 0\} &= \lim_{k\to\infty} \max\{g_i(y^{(k)}), 0\} \\ &\leq \limsup_{k\to\infty} \sigma_{1k}[\inf_{x\in\mathcal{F}} f(x) - \inf_{x\in \mathrm{cl}\mathcal{D}} f(x)] = 0.\end{aligned}$$

Therefore,

$$g_i(\bar{y}) \leq 0, \quad \text{for all } i \in I. \tag{8}$$

From (8), we have $\bar{y} \in \mathcal{F} \subset \mathcal{D}$, which is in contradiction with $\bar{y} \in \mathrm{bd}\mathcal{D}$. Therefore, there exists a $\sigma_1^* > 0$ such that, for all $\sigma \in (0, \sigma_1^*]$ we have $\mathcal{G}(\sigma) \neq \emptyset$.

Next, we will prove that there exists a $\sigma^* > 0$ $(0 < \sigma^* \leq \sigma_1^*)$ such that, for all $\sigma \in (0, \sigma^*]$ we have $\mathcal{G}(\sigma) \subseteq \mathcal{G}$. Namely, there exists $\sigma^* > 0$, such that if $\sigma \in (0, \sigma^*]$, and $z \in \mathcal{D}$ satisfies

$$E_\sigma(z) = \inf_{x\in\mathcal{D}} E_\sigma(x),$$

then $z \in \mathcal{F}$, and $f(z) = \inf_{x\in\mathcal{F}} f(x)$.

Let us assume the converse. Then there exist a sequence $\{\sigma_k\}_{k=1}^\infty \subset (0, \sigma_1^*]$ and a sequence $\{z^{(k)}\}_{i=1}^\infty \subset \mathcal{D}$ such that for all natural numbers k,

$$\sigma_k \leq \frac{1}{k}, \quad E_{\sigma_k}(z^{(k)}) = \inf_{x\in\mathcal{D}} E_{\sigma_k}(x),\ z^{(k)} \notin \mathcal{F}. \tag{9}$$

For each natural number k,

$$\begin{aligned}\inf_{x\in\mathcal{F}} f(x) &= \inf_{x\in\mathcal{F}} E_{\sigma_k}(x) \geq \inf_{x\in\mathcal{D}} E_{\sigma_k}(x) \\ &= E_{\sigma_k}(z^{(k)}) = f(z^{(k)}) + \frac{1}{\sigma_k} V(z^{(k)}) \\ &\leq \inf_{x\in\mathcal{D}} f(x) + \frac{1}{\sigma_k} V(z^{(k)}).\end{aligned} \tag{10}$$

Extracting a subsequence and re-indexing, we may assume without loss of generality that there exists

$$\lim_{k\to\infty} z^{(k)} = \bar{z}. \tag{11}$$

By (10) and (11), similar to the proof of (8), we have $\bar{z} \in \mathcal{F}$.

Since $z^{(k)}$ is a global minimizer of $E_{\sigma_k}(x)$ on $\mathcal{D}$, then for each integer $k \geq 1$,

$$0 \in \partial E_\sigma(z^{(k)}) = \nabla f(z^{(k)}) + \frac{1}{\sigma_k} \partial V(z^{(k)}),$$

$$0 \in \nabla f(z^{(k)}) + \frac{1}{\sigma_k} \left[\sum_{i\in I_+(z^{(k)})} \nabla g_i(z^{(k)}) + \sum_{i\in I_0(z^{(k)})} [0, 1] \nabla g_i(z^{(k)}) \right]. \tag{12}$$

For each integer $k \geq 1$, $\exists\, \alpha_{ki} \in [0, 1]$ such that

$$0 = \sigma_k \nabla f(z^{(k)}) + \sum_{i \in I_+(z^{(k)})} \nabla g_i(z^{(k)}) + \sum_{i \in I_0(z^{(k)})} \alpha_{ki} \nabla g_i(z^{(k)}). \tag{13}$$

By (9), for all integers $k \geq 1$,

$$I_+(z^{(k)}) \neq \emptyset. \tag{14}$$

Extracting a subsequence and re-indexing, we may assume without loss of generality that, for all natural numbers k, $I_+(z^{(1)}) = I_+(z^{(k)})$, $I_-(z^{(1)}) = I_-(z^{(k)})$, and for each $i \in I_0(z^{(1)})$,

$$\lim_{k\to\infty} \alpha_{ki} = \alpha_i. \tag{15}$$

Set $k \to \infty$ in (13), it follows from (9), (11) and (15) that

$$0 = \sum_{i \in I_+(z^{(1)})} \nabla g_i(\bar{z}) + \sum_{i \in I_0(z^{(1)})} \alpha_i \nabla g_i(\bar{z}). \tag{16}$$

By $\bar{z} \in \mathcal{F}$,

$$g_i(\bar{z}) = 0, \quad i \in I_+(z^{(1)}) \cup I_0(z^{(1)}). \tag{17}$$

Therefore, (16) is in contradiction with Assumption 1. The contradiction proves that there exists $\sigma^* > 0$, such that if $\sigma \in (0, \sigma^*]$, then $\mathcal{G}(\sigma) \subseteq \mathcal{G}$.

Let $\tilde{x}$ be a global minimizer of problem (1) and $x_\sigma \in \mathcal{G}(\sigma)$ ($\sigma \in (0, \sigma^*]$), then

$$f(x_\sigma) = E_\sigma(x_\sigma), \qquad f(\tilde{x}) = E_\sigma(\tilde{x}). \tag{18}$$

Therefore, as $f(x_\sigma) = f(\tilde{x})$, (18) implies that $E_\sigma(\tilde{x}) = E_\sigma(x_\sigma)$ and this proves that $\tilde{x}$ is a global solution to problem (6). $\square$

4. Model analysis

Based on the exact penalty property of $E_\sigma(x)$, the following recurrent neural network is proposed for solving the optimization problem (1):

$$\dot{x}(t) \in -\nabla f(x) - \frac{1}{\sigma} \partial \sum_{i=1}^{m} \max\{0, g_i(x)\}. \tag{19}$$

Definition 11. $\bar{x} \in \mathcal{D}$ is said to be an equilibrium point of system (19), if $0 \in -\partial E_\sigma(\bar{x})$. We denote by $\mathcal{E}(\sigma)$ the set of equilibrium point of (19).

Definition 12. A state $x(\cdot)$ of (19) on $[0, t_1]$ is an absolutely continuous function satisfying $x(0) = x_0$ and $\dot{x}(t) \in -\partial E_\sigma(x(t))$ almost all on $[0, t_1]$.

Since $\partial E_\sigma(\cdot)$ is an upper semicontinuous set-valued map with nonempty compact convex set values, then existence of state of (19) is a consequence of Aubin and Cellina (1984).

The Lagrangian function associated with problem (1) is the function $L : R^n \times R^m \to R$ defined by

$$L(x, \lambda) = f(x) + \lambda^T g(x),$$

where

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T,$$
$$g(x) = (g_1(x), g_2(x), \ldots, g_m(x))^T.$$

A KKT twofold for problem (1) is a twofold $(x^*, \lambda^*) \in R^n \times R^m$, such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0,$$
$$\lambda_i^* \geq 0, \quad i = 1, 2, \ldots, m,$$
$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, 2, \ldots, m.$$

It is also well known that relations above become sufficient for optimality if some convexity or generalized convexity assumption is made on f and g_i, $i \in I$ (Cambini & Martein, 2009; Pardalos, 2008).

We denote by $\mathcal{K}$ the set

$$\mathcal{K} = \{x^* \in R^n : \text{there exist } \lambda^* \text{ such that } (x^*, \lambda^*) \text{ is a KKT twofold for problem (1)}\}.$$

Proposition 12. *Any $x^* \in \mathcal{K}$ is an optimal solution of (1), if one of the following conditions hold:*

(a) *$f(x)$ and $g_i(x)$, $i \in I$ are convex functions;*
(b) *$f(x)$ is a pseudoconvex function and $g_i(x)$, $i \in I$ are quasiconvex functions;*
(c) *$f(x)$ is a pseudoconvex function and ${\lambda^*}^T g(x)$ is a quasiconvex function;*
(d) *$f(x) + {\lambda^*}^T g(x)$ is a pseudoconvex function;*
(e) *$f(x)$ and $g_i(x)$, $i \in I$ are invex functions with respect to the same kernel.*

The following two propositions show the correspondence between equilibrium points of (19) and KKT twofold of problem (1) (Pillo & Grippo, 1989).

Proposition 13 (*Pillo & Grippo, 1989*). *Let $\bar{x} \in \mathcal{F}$. If $\bar{x} \in \mathcal{E}(\sigma)$, then $\bar{x} \in \mathcal{K}$. Moreover, if Assumption 1 holds, there exists a $\sigma^* > 0$, such that for all $\sigma \in (0, \sigma^*]$, if $x_\sigma \in \mathcal{E}(\sigma)$, then $x_\sigma \in \mathcal{K}$.*

Proposition 14. *Let $x^* \in \mathcal{K}$, then $x^* \in \mathcal{E}_\sigma$ for all $\sigma > 0$ such that $\lambda_i^* \leq 1/\sigma$, $i \in I_0(x^*)$.*

Since $\nabla f(x)$ is continuous and $\partial V(x)$ is upper semicontinuous with nonempty compact convex values, $\mathrm{cl}\mathcal{D} \setminus \mathrm{int}\mathcal{F}$ is a compact subset, there exist $L_f > 0$ and $L_V > 0$ such that, for $x \in \mathrm{cl}\mathcal{D} \setminus \mathrm{int}\mathcal{F}$, $\|\nabla f(x)\| \leq L_f$ and $|\partial V(x)| = \max\{\|\eta\| : \eta \in \partial V(x)\} \leq L_V$.

The following corollary shows that any equilibrium point of (19) corresponds to a KKT twofold when the penalty parameter is sufficiently small.

Corollary 1. *If $\sigma^* = m_g/2L_f$ and $\sigma \in (0, \sigma^*]$, then $\mathcal{E}(\sigma) \subseteq \mathcal{K}$.*

Proof. Note that for all $x \in \mathcal{D} \setminus \mathcal{F}$ and $\sigma \in (0, \sigma^*]$, $\exists\, \nu_0 \in \partial V(x)$ such that

$$\begin{aligned} |\partial E_\sigma(x)| &:= \min\{\|\nabla f(x) + (1/\sigma)\nu\| : \nu \in \partial V(x)\} \\ &= \|\nabla f(x) + (1/\sigma)\nu_0\| \\ &\geq (1/\sigma)\|\nu_0\| - \|\nabla f(x)\| \\ &\geq (1/\sigma)m_g - L_f > 0. \end{aligned}$$

Therefore, if $x \in \mathcal{E}(\sigma)$, then $x \in \mathcal{F}$. By Proposition 13, $x \in \mathcal{K}$. $\square$

The following theorem gives properties of state of system (19) and shows that for a sufficiently large penalty parameter, the state of (19) is convergent to $\mathcal{K}$.

Theorem 2. *If $x(t)$ is a state trajectory of (19), $\sigma^* = m_g/2L_f$ and $\sigma \in (0, \sigma^*]$, then*

(i) *for almost all $t \geq 0$,*

$$\frac{d}{dt} E_\sigma(x(t)) = -\|\dot{x}(t)\|^2.$$

(ii) *$x(t)$ is bounded and defined on $[0, +\infty)$.*
(iii) *$\dot{x}(t)$ almost converges to 0; i.e.,*

$$\mu - \lim_{t\to+\infty} \dot{x}(t) = 0.$$

(iv) *any almost cluster point of $x(t)$ is an equilibrium point of (19).*
(v) $\lim_{t\to+\infty} \mathit{dist}(x(t), \mathcal{K}) = 0$.

Proof. (i) From Proposition 7, for almost all $t \geq 0$

$$\frac{d}{dt}E_\sigma(x(t)) = \langle \eta(t), \dot{x}(t)\rangle, \quad \forall \eta(t) \in \partial E_\sigma(x(t)). \tag{20}$$

Replacing $\eta(t)$ with $-\dot{x}(t)$ in (20),

$$\frac{d}{dt}E_\sigma(x(t)) = -\|\dot{x}(t)\|^2. \tag{21}$$

$E_\sigma(x(t))$ is a non-increasing function.

(ii) Integrating (21) over $[0, t]$, we have

$$E_\sigma(x(t)) - E_\sigma(x_0) \leq -\int_0^t \|\dot{x}(\tau)\|^2 d\tau \leq 0, \tag{22}$$

then $E_\sigma(x(t)) \leq E_\sigma(x_0)$. Since $E_\sigma(x)$ is radially unbounded, it follows from Proposition 11 that $x(t)$ is bounded, thus $x(t)$ is defined $[0, +\infty)$.

(iii) From (21) and (22), we have $\int_0^{+\infty} \|\dot{x}(t)\|dt < +\infty$. If $\dot{x}(t)$ does not almost converge to 0, then $\exists\, \varepsilon_0 > 0$ such that for $\forall\, T > 0$,

$$\mu\{t : \|\dot{x}(t)\| > \varepsilon_0, t \in [T, +\infty)\} \geq \varepsilon_0,$$

then $\int_T^{+\infty} \|\dot{x}(\tau)\|^2 d\tau \geq \varepsilon_0^2$ which is a contradiction.

(iv) Let $\dot{x}(t) = -\nabla f(x(t)) - (1/\sigma)\eta(t)$, $\eta(t) \in \partial V(x(t))$, and $t \in [0, +\infty) \setminus N$, where $\mu(N) = 0$. Since $x(\cdot)$ on $[0, +\infty) \setminus N$ is bounded, by Proposition 3 there exists an almost cluster point x^* of $x(t)$ when $t \to +\infty$. Next we shall show that x^* is an equilibrium point of (19). Note that $\dot{x}(t)$ almost converges to 0 when $t \to +\infty$, then there exists an increasing sequence $t_k \uparrow +\infty$ in $[0, +\infty)$ such that

$$\lim_{k\to\infty} x(t_k) = x^*, \qquad \lim_{k\to\infty} \dot{x}(t_k) = 0. \tag{23}$$

Since $\nabla f(x)$ is continuous and $-\partial V(x)$ is upper semicontinuous at x^*, for every closed neighborhood U of the origin in R^n, $\exists k_0$ such that for $k > k_0$, we have

$$\begin{aligned} \dot{x}(t_k) &\in -\nabla f(x(t_k)) - (1/\sigma)V(x(t_k)) \\ &\subset -\nabla f(x^*) - (1/\sigma)V(x^*) + U. \end{aligned} \tag{24}$$

Taking $k \to \infty$ in (24), it follows from (23) that

$$0 \in -\nabla f(x^*) - (1/\sigma)V(x^*) + U.$$

Since U is arbitrary, then $0 \in -\nabla f(x^*) - (1/\sigma)V(x^*)$, and x^* is an equilibrium point of (19).

(v) If $\lim_{t\to+\infty} \mathrm{dist}(x(t), \mathcal{E}(\sigma)) = 0$ does not hold, then there exist $\varepsilon_0 > 0$ and $\{t_k\}_{k=1}^\infty \uparrow +\infty$, such that

$$\mathrm{dist}(x(t_k), \mathcal{E}(\sigma)) \geq \varepsilon_0.$$

Since $\{x(t_k)\}_{k=1}^\infty$ is bounded, we may assume without loss of generality that there exists

$$\lim_{k\to\infty} x(t_k) = \bar{x}. \tag{25}$$

Next, we prove that $\bar{x} \in \mathcal{E}(\sigma)$, i.e. $0 \in \partial E_\sigma(\bar{x})$. If it does not hold, that is $\mathrm{dist}(0, \partial E_\sigma(\bar{x})) = d > 0$. Since $x \hookrightarrow \partial E_\sigma(x)$ is upper semicontinuous, then $x \to \mathrm{dist}(0, \partial E_\sigma(x))$ is l.s.c. Therefore, $\forall\, \varepsilon > 0, \exists\, \delta > 0$, such that for all $x \in \bar{x} + \delta\mathcal{B}$,

$$d = \mathrm{dist}(0, \partial E_\sigma(\bar{x})) < \mathrm{dist}(0, \partial E_\sigma(x)) + \varepsilon.$$

Taking $\varepsilon = \frac{d}{2}$, when $x \in \bar{x} + \delta\mathcal{B}$, $\mathrm{dist}(0, \partial E_\sigma(x)) > \frac{d}{2}$.

Since $\lim_{k\to\infty} x(t_k) = \bar{x}$, there exists a positive integer k_0, such that for all $k > k_0$, $x(t_k) \in \bar{x} + \frac{\delta}{2}\mathcal{B}$.

Note that $x \hookrightarrow \partial E_\sigma(x)$ is upper semicontinuous and $x(t)$ is bounded, then $\partial E_\sigma(x(t))$ is bounded, set

$$\sup_{\eta\in\partial E_\sigma(x(t))} \|\eta\| \leq L, \quad t \in [0, +\infty).$$

Therefore for each $t_1, t_2 \in [0, \infty]$,

$$\|x(t_2) - x(t_1)\| \leq \int_{t_1}^{t_2} \|\dot{x}(t)\|dt \leq L|t_2 - t_1|.$$

When $t \in [t_k - \delta/(4L), t_k + \delta/(4L)]$ and $k \geq k_0$,

$$\begin{aligned} \|x(t) - \bar{x}\| &\leq \|x(t) - x(t_k)\| + \|x(t_k) - \bar{x}\| \\ &\leq L|t - t_k| + \frac{\delta}{2} < \delta. \end{aligned}$$

It follows that $\mathrm{dist}(0, \partial E_\sigma(x(t))) > \frac{d}{2}$ for all $t \in [t_k - \delta/(4L), t_k + \delta/(4L)]$.

Since

$$\mu\left(\bigcup_{k>k_0} [t_k - \delta/(4L), t + \delta/(4L)]\right) = +\infty,$$

then

$$\int_0^{+\infty} \mathrm{dist}^2(0, \partial E_\sigma(x(\tau)))d\tau = +\infty. \tag{26}$$

Note that

$$\int_0^{+\infty} \mathrm{dist}^2(0, \partial E_\sigma(x(\tau)))d\tau \leq \int_0^{+\infty} \|\dot{x}(\tau)\|^2 d\tau < +\infty$$

which contradicts (26). Therefore, $\bar{x} \in \mathcal{E}(\sigma)$. Since

$$t \to \mathrm{dist}(x(t), \mathcal{E}(\sigma))$$

is continuous, then

$$\lim_{k\to\infty} \mathrm{dist}(x(t_k), \mathcal{E}(\sigma)) = \mathrm{dist}(\bar{x}, \mathcal{E}(\sigma)) = 0,$$

which is a contradiction. The contradiction proves

$$\lim_{t\to+\infty} \mathrm{dist}(x(t), \mathcal{E}(\sigma)) = 0.$$

By Corollary 1, $\mathcal{E}(\sigma) \subseteq \mathcal{K}$, thus

$$\mathrm{dist}(x(t), \mathcal{K}) \leq \mathrm{dist}(x(t), \mathcal{E}(\sigma)).$$

Therefore, (v) holds. □

By Proposition 12 and Theorem 2, the following corollary holds:

Corollary 2. *If one of the assumptions* (a)–(e) *in Proposition 12 holds, then state of (19) is convergent to the optimal solution set of Problem (1).*

Next, we will show that a state of (19) can reach the feasible region in finite time and stay there thereafter. The following lemma is very useful in finite-time convergence analysis.

Lemma 2 (Chong, Hui, & Zak, 1999)**.** *Suppose that there exists $V : R^n \to R$ such that $V(x(t))$ is absolutely continuous on $[0, +\infty)$, and there exists $\rho > 0$ such that for almost all $t \in [0, +\infty)$ for which $x(t) \in \{x \in R^n : V(x) > 0\}$, and*

$$\frac{d}{dt}V(x(t)) \leq -\rho.$$

Then, the state $x(t)$ is convergent to $\{x \in R^n : V(x) \leq 0\}$ in finite time and stay there thereafter.

Theorem 3. *If $\sigma^* = m_g/2L_f$ and $\sigma \in (0, \sigma^*]$, then any state of (19) is guaranteed to be convergent to the feasible region in finite time and stay there thereafter.*

Proof. Consider the following set of time:

$$\Gamma = \{t : V(x(t)) > 0\}.$$

If Γ is bounded, then Theorem 3 naturally holds. If Γ is unbounded, we will prove

$$\inf_{t\in\Gamma} \min_{\eta\in\partial V(x(t))} \|\eta\| = l > 0. \tag{27}$$

If (27) does not hold, i.e. $l = 0$, then there exists $\{t_k\}_{k=1}^{\infty} \subset \Gamma$ which satisfies $t_k \to +\infty$ and $\eta_k \in \partial V(x(t_k))$ such that

$$\|\eta_k\| < \frac{1}{k}. \tag{28}$$

From (2), $\exists\, \alpha_{k_i} \in [0, 1]$ such that

$$\eta_k = \sum_{i\in I_+(x(t_k))} \nabla g_i(x(t_k)) + \sum_{i\in I_0(x(t_k))} \alpha_{k_i} \nabla g_i(x(t_k)). \tag{29}$$

Since $\{x(t_k)\}_{k=1}^{\infty}$ is bounded, we may assume without loss of generality that there exists

$$\lim_{k\to\infty} x(t_k) = \bar{z}.$$

Similarly to the proof of (iv) in Theorem 2 and Corollary 1, we have $\bar{z} \in \mathcal{F}$, thus $g_i(\bar{z}) \le 0,\ i = 1, 2, \ldots, m$.

We assume without loss of generality that, for all natural numbers k, $I_+(x(t_k)) = I_+(x(t_1)) \neq \emptyset$. Since $g_i(x(_k)) > 0$ $(i \in I_+(x(t_k)))$, then

$$\lim_{k\to+\infty} g_i(x(t_k)) = g_i(\bar{z}) \ge 0, \quad i \in I_+(x(t_1)).$$

Therefore

$$g_i(\bar{z}) = 0, \quad i \in I^+(x(t_1)). \tag{30}$$

We may assume without loss of generality $\lim_{k\to\infty} \alpha_{k_i} = \bar{\alpha_i}$. Taking $k \to \infty$ in (29),

$$\sum_{i\in I_+(x(t_1))} \nabla g_i(\bar{z}) + \sum_{i\in I_0(x(t_1))} \bar{\alpha_i} \nabla g_i(\bar{z}) = 0. \tag{31}$$

(31) is a contradiction to Assumption 1. The contradiction proves (27).

According to Proposition 7, for almost all $t \in \Gamma$,

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

From (19), there exists $\bar{\eta} \in \partial V(x(t))$ such that

$$\begin{aligned}\frac{d}{dt} V(x(t)) &= \langle \bar{\eta}(t), -\nabla f(x(t)) - (1/\sigma)\bar{\eta}(t)\rangle \\ &\le \|\nabla f(x(t))\| \cdot \|\bar{\eta}(t)\| - (1/\sigma)\|\bar{\eta}(t)\|^2.\end{aligned} \tag{32}$$

Since $x(t)$ is bounded, $\nabla f(x(t))$ is continuous and $\partial V_\gamma(x(t))$ is an upper semicontinuous map with nonempty compact convex values, then $\exists\, l_1,\ l_2 > 0$ such that $\|\nabla f(x(t))\| \le l_1$ and $\|\eta(t)\| \le l_2$, $\forall \eta(t) \in \partial V(x(t))$. It follows from (27) and (32) that

$$\frac{d}{dt} V(x(t)) \le l_1 l_2 - (1/\sigma) l^2,$$

thus, for $\sigma < \frac{l^2}{2l_1 l_2}$, $\rho = l_1 l_2 > 0$, we obtain

$$\frac{d}{dt} V(x(t)) \le -\rho. \tag{33}$$

According to Lemma 2, the state of (19) reaches feasible in finite time. Integrating (33) on $[0, t]$,

$$V(x(t)) \le V(x_0) - \rho t.$$

Next we prove that when $t \ge V(x_0)/\rho$, the state of (19) remains inside $\mathcal{F}$ thereafter. If not so, we suppose that the state leaves $\mathcal{F}$ in finite time. Let

$$\vartheta = \inf\{t > V(x_0)/\rho : x(t) \notin \mathcal{F}\} < +\infty;$$

by the continuity of $t \to x(t)$, there exists $h > 0$ such that for any $t \in (\vartheta, \vartheta + h)$,

$$V(x(t)) > 0. \tag{34}$$

On the other hand, by (33) and $(\vartheta, \vartheta + h) \subset \Gamma$, the function $t \to V(x(t))$ is decreasing on $(\vartheta, \vartheta + h)$ and $V(x(\vartheta)) = 0$; hence, for any $t \in (\vartheta, \vartheta + h)$, $V(x(t)) < 0$, which is in contradiction with (34). □

Next, we will further study the convergence of state of (19). To this end, the following lemma is needed (Opial, 1967).

Lemma 3. *Let $x(t) : [0, \infty) \to R^n$ be a state trajectory such that there exists a nonempty subset $\mathcal{C} \subset R^n$ which satisfies the conditions below:*

(i) *All cluster points of $x(\cdot)$ are contained in $\mathcal{C}$;*
(ii) *$\forall\, x^* \in \mathcal{C}$, $\lim_{t\to\infty} \|x(t) - x^*\|$ exists.*

Then, $x(t)$ converges to an element of $\mathcal{C}$ as $t \to \infty$.

The following theorem shows that the state of (19) is convergent to an optimal solution of Problem (1), if some convexity or generalized convexity conditions on f and $g_i,\ i \in I$ are satisfied.

Theorem 4. *Any state of (19) converges to an optimal solution of Problem (1) if $\sigma^* = m_g/2L_f$, $\sigma \in (0, \sigma^*]$ and one of the following two conditions hold:*

(a) *$f(x)$ and $g_i(x),\ i \in I$ are convex functions;*
(b) *$f(x)$ is a pseudoconvex function and $g_i(x),\ i \in I$ are quasiconvex functions.*

Proof. (a) Let $x(\cdot)$ be a state of (19) with initial point $x_0 \in \mathcal{D}$. Since $f(x)$ and $g_i(x),\ i \in I$ are convex functions, then $E_\sigma(x)$ is a convex function. By Theorem 1 and Corollary 1, $\mathcal{G} = \mathcal{G}(\sigma) \subseteq \mathcal{E}(\sigma)$.

By Theorem 3, there exists $t_0 > 0$ such that for $t > t_0$, $x(t) \in \mathcal{F}$. In the following proof, we assume that $t > t_0$.

For any $x^* \in \mathcal{G}$, consider the following energy function

$$E(t, x^*) = E_\sigma(x(t)) - E_\sigma(x^*) + (1/2)\|x(t) - x^*\|^2. \tag{35}$$

By Proposition 7, for almost all $t \ge 0$,

$$(d/dt)E(t, x^*) = -\|\dot{x}(t)\|^2 + \langle x(t) - x^*, \dot{x}(t)\rangle. \tag{36}$$

By convexity of $E_\sigma(x)$,

$$\begin{aligned}&E_\sigma(x(t)) - E_\sigma(x^*) \le \langle x(t) - x^*, \eta(t)\rangle, \\ &\forall\, \eta(t) \in \partial E_\sigma(x(t)).\end{aligned} \tag{37}$$

From (19) and (37), taking $\dot{x}(t) = -\eta(t)$ in (36), we have

$$(d/dt)E(t, x^*) = -\|\dot{x}(t)\|^2 - \langle x(t) - x^*, \eta(t)\rangle \le 0. \tag{38}$$

By (38), $E(t, x^*)$ is monotonically nonincreasing with respect to t. Since $E(t, x^*) \ge 0$, then $E(t, x^*)$ converges as $t \to +\infty$. By Theorem 2, $E_\sigma(x(t))$ converges as $t \to +\infty$, then $\|x(t) - x^*\|$ is convergent as $t \to +\infty$. According to Lemma 3, $x(t)$ converges to an element of $\mathcal{G}$, which is an optimal solution of (1).

(b) By Proposition 12 and Corollary 1, $\mathcal{E}(\sigma) \subseteq \mathcal{K} = \mathcal{G}$.

Similar to the proof in (a), we only need to prove that for any $x^* \in \mathcal{E}(\sigma)$, (38) holds.

If $x(t) \in \mathcal{E}(\sigma)$, then $x(t)$ reaches $\mathcal{E}(\sigma)$ in finite time.

If $x(t) \notin \mathcal{E}(\sigma)$, $f(x(t)) > f(x^*)$. Since $f(x)$ is pseudoconvex, then

$$\langle \nabla f((t)), x(t) - x^*\rangle > 0. \tag{39}$$

If $x(t) \in \mathrm{bd}\mathcal{F}$, $\exists\, \alpha_i \in [0, 1]$, $i \in I_0(x(t))$,

$$\dot{x}(t) = -\nabla f(x(t)) - (1/\sigma) \sum_{i \in I_0(x(t))} \alpha_i \nabla g_i(x(t)), \tag{40}$$

thus

$$(d/dt)E(t, x^*) = -\|\dot{x}(t)\|^2 - \langle \nabla f(x(t)), x(t) - x^* \rangle - (1/\sigma) \sum_{i \in I_0(x(t))} \alpha_i \langle \nabla g_i(x(t)), x(t) - x^* \rangle. \tag{41}$$

Since $g_i(x)$ $(i \in I_0(x(t)))$ are quasiconvex functions and $g_i(x(t)) \geq g_i(x^*)$, then

$$\langle \nabla g_i(x(t)), x(t) - x^* \rangle \geq 0, \quad i \in I_0(x(t)).$$

By (39) and (41), $(d/dt)E(t, x^*) \leq 0$.

If $x(t) \in \mathrm{int}\mathcal{F}$,

$$(d/dt)E(t, x^*) = -\|\dot{x}(t)\|^2 - \langle \nabla f(x(t)), x(t) - x^* \rangle. \tag{42}$$

By (39) and (42) $(d/dt)E(t, x^*) \leq 0$.

Therefore, the state of (19) either converges to an optimal solution of (1) in finite time or converges to an optimal solution of (1) asymptotically. □

Remark 1. Compared with existing results on recurrent neural networks for nonconvex optimization, such as Bian and Chen (2012), Bian and Xue (2009), and Xue and Bian (2008) the neural network model (19) has several advantages. First, it allows the feasible region to be a nonconvex set. Although the neural networks in Bian and Chen (2012), Bian and Xue (2009) and Xue and Bian (2008) can handle some nonconvex objective functions, the constraints are required to be convex functions. Second, any neural state of the proposed model is convergent to the feasible region in finite time. Third, any neural state is convergent to a KKT point of the constrained optimization problem. As a result, the proposed neural network can obtain the global optimal solution to an invex optimization problem.

5. Simulation results

In this section, simulation results on three nonconvex optimization problems are provided to illustrate the effectiveness and efficiency of the proposed recurrent neural network model (19).

Example 1. Consider a nonconvex optimization problem as follows:

$$\begin{array}{ll} \min & f(x) = 2 - x_1 x_2 \\ \text{subject to} & x_1 + 4x_2 - 1 \leq 0,\ x_1^2 + x_2^2 - 1 \leq 0, \end{array} \tag{43}$$

where objective function is nonconvex as shown in Fig. 1. The generalized gradient ∂E_σ is computed as

$$\partial E_\sigma(x) = (-x_2, -x_1)^T + (1/\sigma)\partial V(x),$$

where

$$\partial V(x) = \partial \max\{0, x_1 + 4x_2 - 1\} + \partial \max\{0, x_1^2 + x_2^2 - 1\}.$$

There are three KKT points corresponding to (43): $(-0.7071, -0.7071)$, $(0, 0)$, and $(0.5, 0.125)$ where $(-0.7071, -0.7071)$ is a local minimum solution, $(0, 0)$ is the global maximum solution, and $(0.5, 0.125)$ is the global minimum solution. Fig. 2 illustrates the transient behaviors of the proposed neural network from 30 random initial states. Let the design parameter $\sigma = 0.2$ so it satisfied $\sigma \leq \sigma^* = 5$. Fig. 3 shows the 2-dimensional phase plot from 100 random initial states. The simulation results show that the proposed neural network always converges to a KKT point

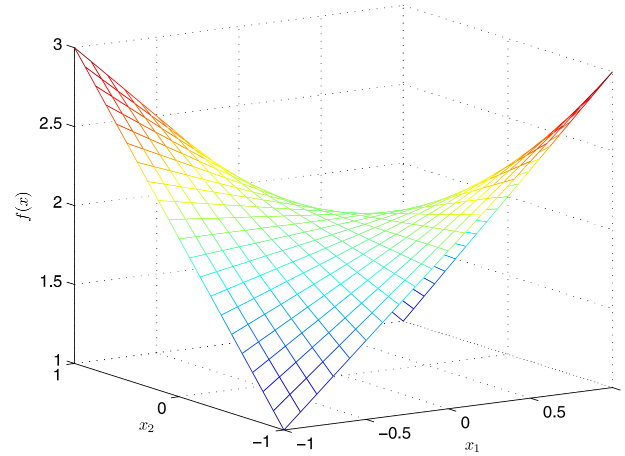

Fig. 1. Objective function in Example 1.

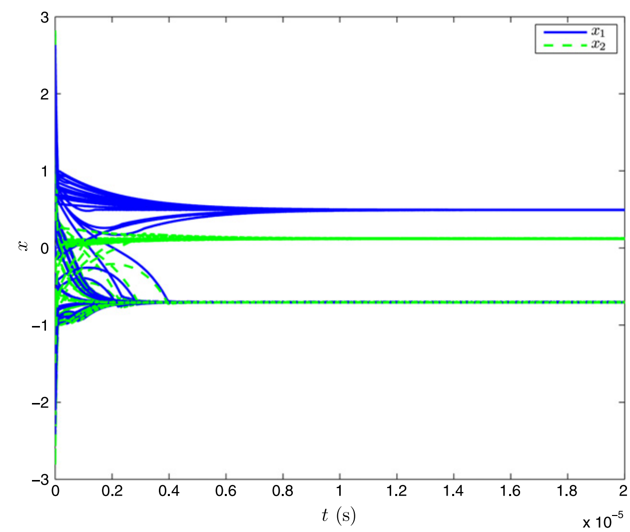

Fig. 2. Transient behaviors of the proposed neural network for Example 1.

of the nonconvex optimization problem (43). The neural network presented in Xia et al. (2008) was also applied for solving (43). As depicted in Fig. 4, the network is not convergent for some initial conditions.

Example 2. Consider a three-hump optimization problem:

$$\begin{array}{ll} \min & f(x) = 2x_1^2 - 1.05x_1^4 + \dfrac{1}{6}x_1^6 - x_1 x_2 + x_2^2 \\ \text{subject to} & x_1 + x_2 - 2 \leq 0,\ x_1^2 + x_2^2 - 4 \leq 0. \end{array} \tag{44}$$

As shown in Fig. 5, the objective function is a multimodal nonconvex function. The generalized gradient ∂E_σ is

$$\partial E_\sigma(x) = (4x_1 - 4.2x_1^3 + x_1^5 - x_2, -x_1 + 2x_2)^T + (1/\sigma)\partial V(x),$$

where

$$\partial V(x) = \partial \max\{0, x_1 + x_2 - 2\} + \partial \max\{0, x_1^2 + x_2^2 - 4\}.$$

There are three KKT points for the problem: $(-1.7476, -0.8738)$, $(-1.7476, 0.8738)$, and $(0, 0)$. The global minimum solution is $(0, 0)$. The upper bound of the design parameter is estimated as $\sigma^* = \sqrt{(2)}$. Let $\sigma = 0.2$ in the simulation. Fig. 6 illustrates the transient behaviors of the proposed neural network from 30 random initial states. Fig. 7 shows the 2-dimensional phase plot from 100 random initial states. The simulation results show that the proposed neural network always converges to a KKT point of the nonconvex optimization problem (44).

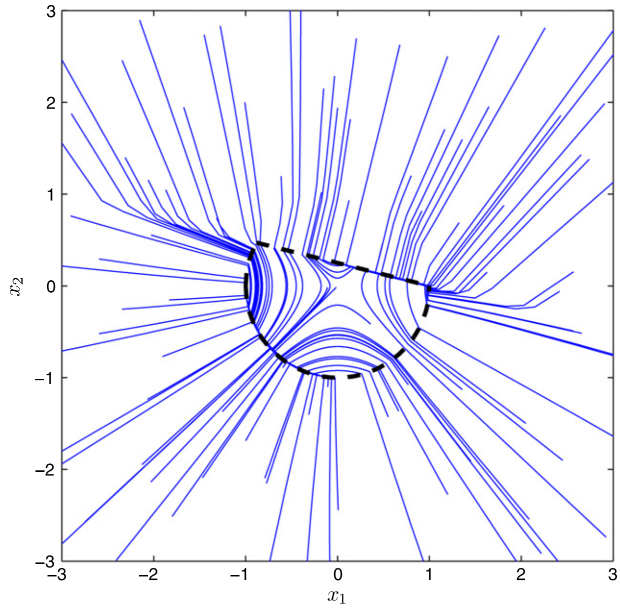

Fig. 3. Phase plot of the proposed neural network for Example 1.

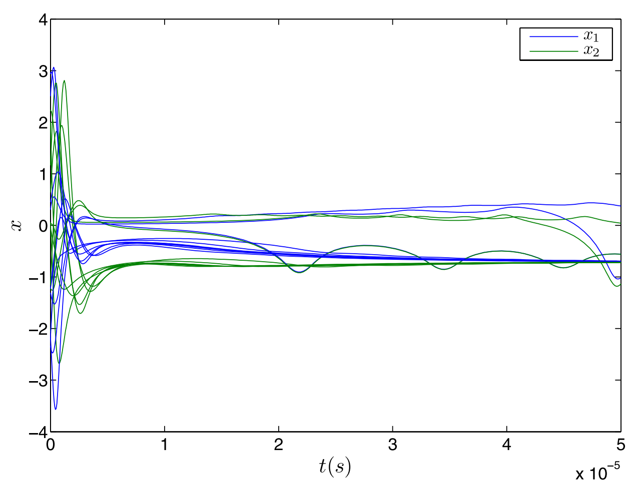

Fig. 4. Transient behaviors of network in Xia et al. (2008) for Example 1.

Example 3. Consider an invex optimization problem as follows:

$$\begin{array}{ll} \min & f(x) = 1 + x_1^2 - e^{-x_2^2} \\ \text{subject to} & x_1^2 - x_2 + 0.5 \leq 0,\ 2x_2 - x_1^2 - 3 \leq 0. \end{array} \tag{45}$$

As depicted in Fig. 8, the objective function $f(x)$ is a smooth invex function. The feasible region $S = \{x \in \Re^2 : x_1^2 - x_2 \leq 0, 2x_2 - x_1^2 - 3 \leq 0\}$ is not a convex set. The generalized gradient ∂E_σ is

$$\partial E_\sigma(x) = (2x_1, 2x_2 + 2x_2 e^{-x_2^2})^T + (1/\sigma)\partial V(x),$$

where

$$\partial V(x) = \partial \max\{0, x_1^2 - x_2 + 0.5\} + \partial \max\{0, 2x_2 - x_1^2 - 3\}.$$

The invex optimization problem (45) has a unique KKT point (0, 0.5), which is the global minimum solution. The upper bound of the design parameter is estimated as $\sigma^* = 0.14$. Let $\sigma = 0.067$ in the simulation. Fig. 9 illustrates the transient behaviors of the proposed neural network from 30 random initial states. Fig. 10 shows the 2-dimensional phase plot from 100 random initial states. The proposed neural network can globally convergent

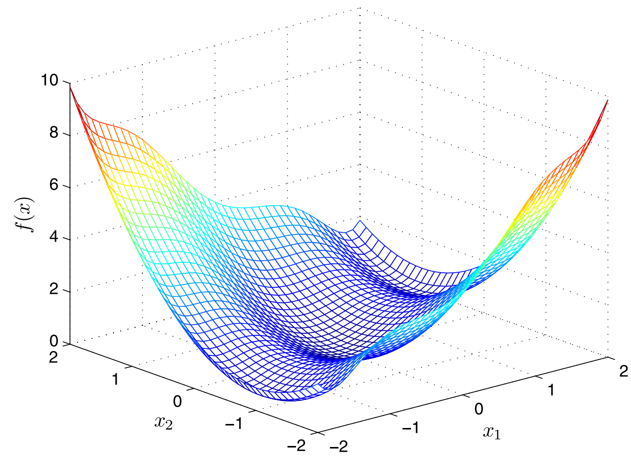

Fig. 5. Objective function in Example 2.

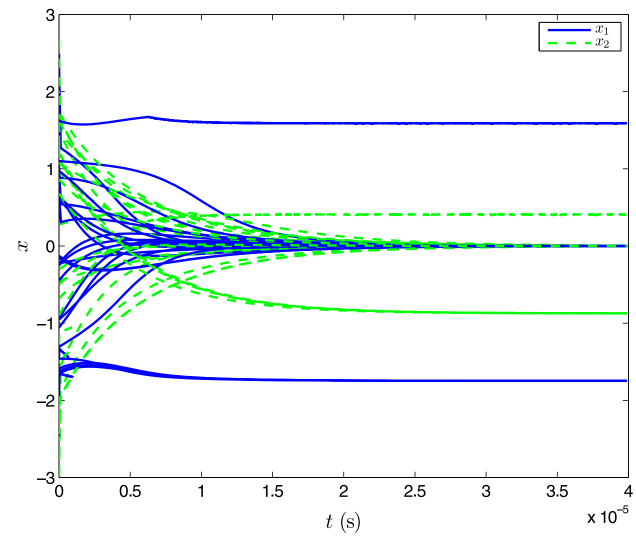

Fig. 6. Transient behaviors of the proposed neural network for Example 2.

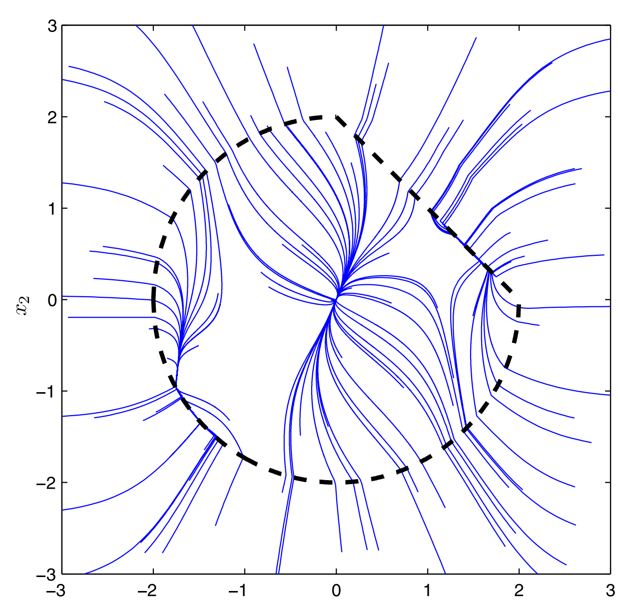

Fig. 7. Phase plot of the proposed neural network for Example 2.

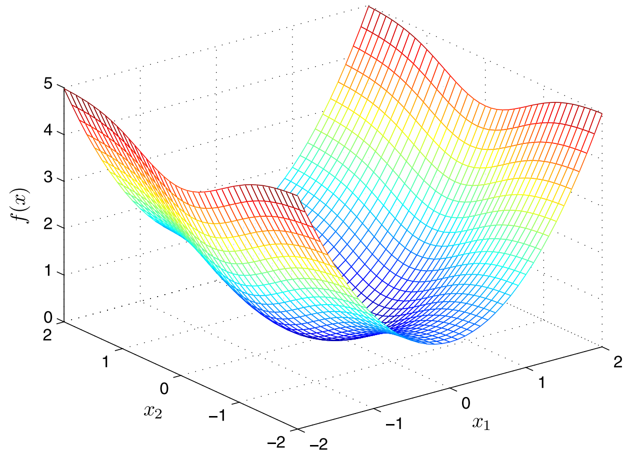

Fig. 8. Objective function in Example 3.

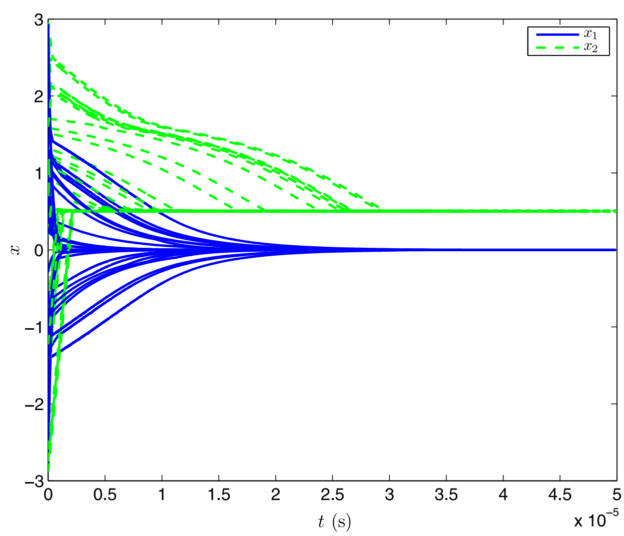

Fig. 9. Transient behaviors of the proposed neural network for Example 3.

to the global optimal solution to the invex optimization problem. In contrast, as shown in Fig. 11, the neural network in Nazemi (2012) cannot converge to the optimal solution.

Example 4. Consider a nonconvex optimization problem as follows:

$$\begin{aligned} \min \quad & f(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right) x_1^2 + x_1x_2 \\ & + (4x_2^2 - 4)x_2^2 \qquad (46) \\ \text{subject to} \quad & \frac{x_1^2}{1.9^2} + \frac{x_2^2}{1.5^2} - 1 \le 0, \ \ \frac{1}{2}x_1^2 + x_2 - 1.1 \le 0. \end{aligned}$$

As depicted in Fig. 12, $f(x)$ in (46) is a Six-Hump Camel Back function. The problem has six KKT points locating at $(-0.0898, 0.7127)$, $(0.0898, -0.7127)$, $(-1.7036, 0.7961)$, $(1.7036, -0.7961)$, $(-1.6071, -0.5687)$, and $(1.6071, 0.5687)$. Among these KKT points, the first two correspond to global minima and the other four correspond to local minima. The upper bound of the design parameter is estimated as $\sigma^* = 1.05$. Let $\sigma = 0.1$ in the simulation. Fig. 13 illustrates the transient behaviors of the proposed neural network from 30 random initial states. Fig. 14 shows the 2-dimensional phase plot from 100 random initial states. The simulation results show that the proposed neural network is

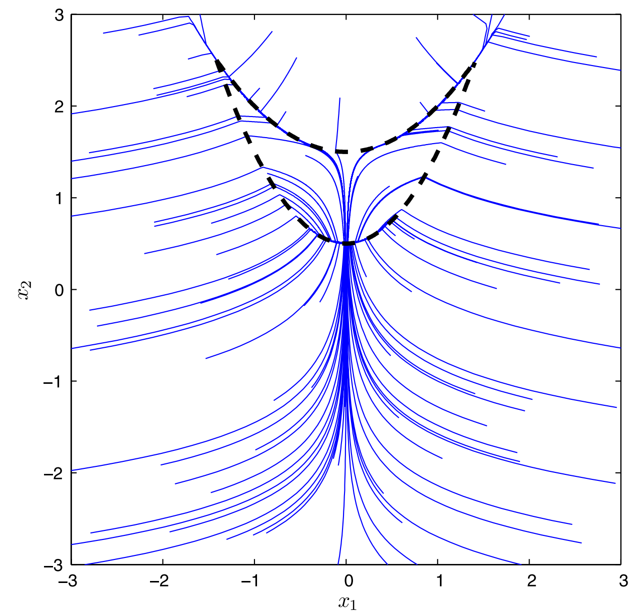

Fig. 10. Phase plot of the proposed neural network for Example 3.

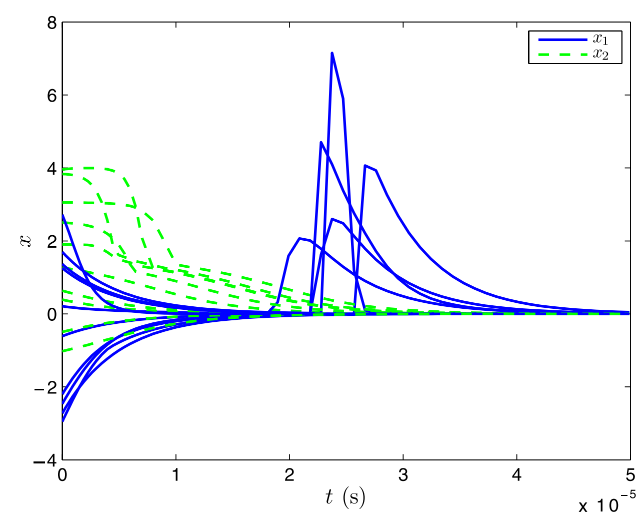

Fig. 11. Transient behaviors of neural network in Nazemi (2012) for Example 3.

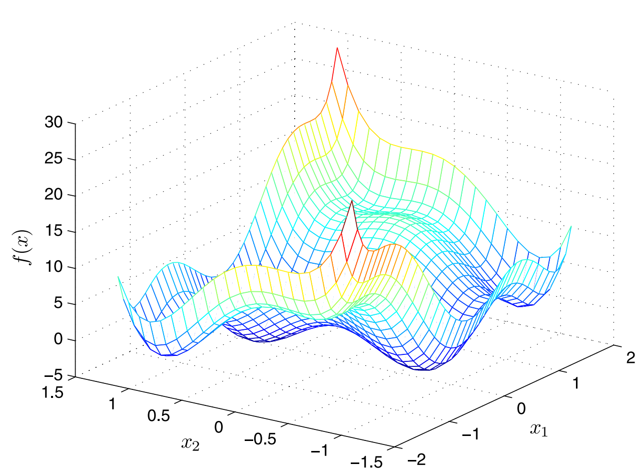

Fig. 12. Objective function in Example 4.

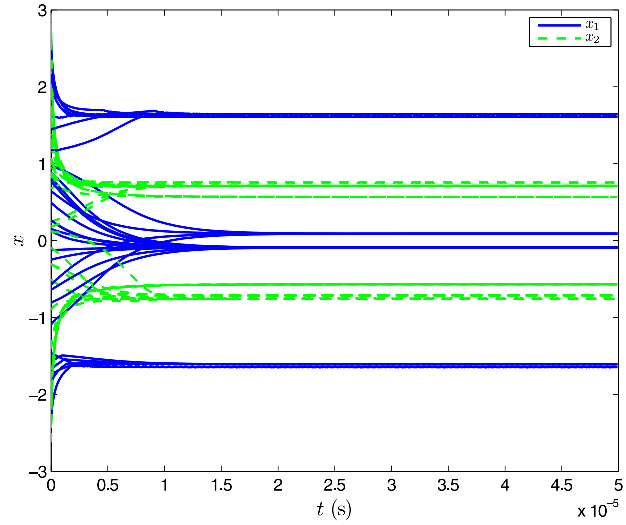

Fig. 13. Transient behaviors of the proposed neural network for Example 4.

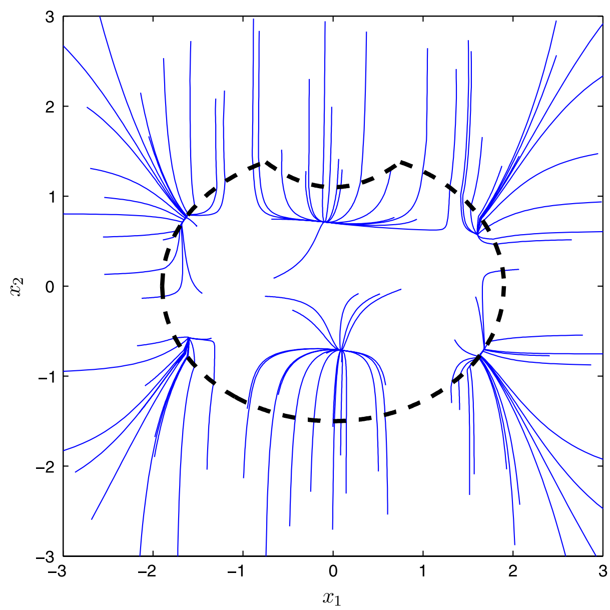

Fig. 14. Phase plot of the proposed neural network for Example 4.

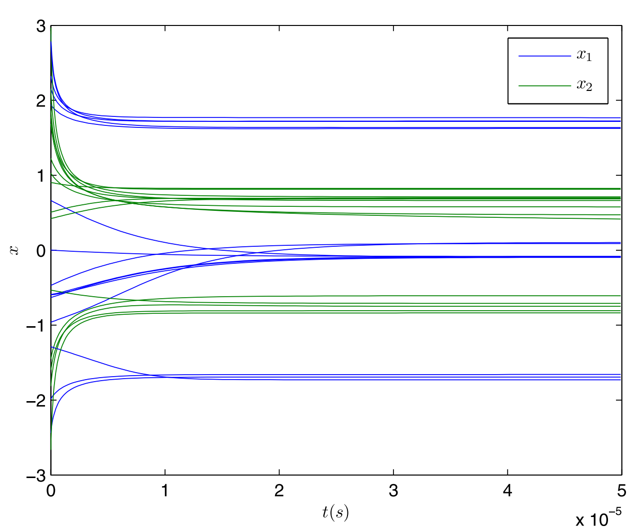

Fig. 15. Transient behaviors of network in Xia et al. (2008) for Example 4.

convergent to the KKT point set of the nonconvex optimization problem (46). When the neural network in Xia et al. (2008) is applied for solving (46), its equilibria are not in one-to-one correspondence with the KKT points. As shown in Fig. 15, the number of equilibria is more than the number of KKT points.

6. Conclusion

This paper presents a one-layer recurrent neural network for nonconvex optimization problems with inequality constraints based on an exact penalty design. The proposed neural network is proved to be convergent to its equilibrium point set and any equilibrium point of the neural network corresponds to a KKT point of the nonconvex problem. Moreover, it is proved that any state of the proposed neural network converges to the feasible region in finite time and stays there thereafter. Simulation results are discussed to substantiate the characteristics and effectiveness of the proposed neural network. Future investigations are directed to global optimization of hybrid intelligence with a combination of neurodynamic and evolutionary optimization approaches.

References

Asai, T., Kanazawa, Y., & Amemiya, Y. (2003). A subthreshold mos neuron circuit based on the Volterra system. *IEEE Transactions on Neural Networks*, *14*, 1308–1312.

Aubin, J., & Cellina, A. (1984). *Differential inclusions*. Verlin: Springer-Verlag.

Balavoine, A., Romberg, J., & Rozell, C. J. (2012). Convergence and rate analysis of neural networks for sparse approximation. *IEEE Transactions on Neural Networks and Learning Systems*, *23*, 1377–1389.

Bian, W., & Chen, X. (2012). Smoothing neural network for constrained non-Lipschitz optimization with applications. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(3), 399–411.

Bian, W., & Xue, X. (2009). Subgradient-based neural networks for nonsmooth nonconvex optimization problems. *IEEE Transactions on Neural Networks*, *20*, 1024–1038.

Bian, W., & Xue, X. (2013). Neural network for solving constrained convex optimization problems with global attractivity. *IEEE Transactions on Circuits and Systems. I. Regular Papers*, *60*(3), 710–723.

Cambini, A., & Martein, L. (2009). *Generalized convexity and optimization: theory and applications*. Berlin, Heidelberg: Springer-Verlag.

Cheng, L., Hou, Z., Lin, Y., Tan, M., Zhang, W. C., & Wu, F. (2011). Recurrent neural network for nonsmooth convex optimization problems with applications to the identification of genetic regulatory networks. *IEEE Transactions on Neural Networks*, *22*, 714–726.

Chong, E., Hui, S., & Zak, S. (1999). An analysis of a class of neural network for solving linear programming problems. *IEEE Transactions on Automatic Control*, *44*, 1995–2006.

Clarke, F. (1969). *Optimization and non-smooth analysis*. New York: Wiley.

Filippov, A. (1988). *Differential equations with discontinuous right-hand side*. Dordrecht: Kluwer Academic.

Forti, M., Nistri, P., & Quincampoix, M. (2004). Generalized neural network for nonsmooth nonlinear programming problems. *IEEE Transactions on Circuits and Systems. I*, *51*, 1741–1754.

Forti, M., Nistri, P., & Quincampoix, M. (2006). Convergence of neural network for programming problems via a nonsmooth Lojasiewicz inequality. *IEEE Transactions on Neural Networks*, *17*, 1471–1485.

Gao, X. (2004). A novel neural network for nonlinear convex programming. *IEEE Transactions on Neural Networks*, *15*, 613–621.

Guo, Z., Liu, Q., & Wang, J. (2011). A one-layer recurrent neural network for pseudoconvex optimization subject to linear equality constraints. *IEEE Transactions on Neural Networks*, *22*, 1892–1900.

Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, *52*, 141–152.

Hosseini, A., Wang, J., & Hosseini, S. M. (2013). A recurrent neural network for solving a class of generalized convex optimization problems. *Neural Networks*, *44*, 78–86.

Hou, Z.-G., Gupta, M. M., Nikiforuk, P. N., Tan, M., & Cheng, L. (2007). A recurrent neural network for hierarchical control of interconnected dynamic systems. *IEEE Transactions on Neural Networks*, *18*, 466–481.

Hu, X., & Wang, J. (2006). Solving pseudomonotone variational inequalities and pseudoconvex optimization problems using the projection neural network. *IEEE Transactions on Neural Networks*, *17*, 1487–1499.

Hu, X., & Wang, J. (2007). Design of general projection neural network for solving monotone linear variational inequalities and linear and quadratic optimization problems. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *37*, 1414–1421.

Li, G., Yan, Z., & Wang, J. (2014). A one-layer recurrent neural network for constrained nonsmooth invex optimization. *Neural Networks*, *50*, 79–89.

Liu, Q., Cao, J., & Chen, G. (2010). A novel recurrent neural network with finite-time convergence for linear programming. *Neural Computation*, *22*, 2962–2978.

Liu, Q., Guo, Z., & Wang, J. (2012). A one-layer recurrent neural network for constrained pseudoconvex optimization and its application for dynamic portfolio optimization. *Neural Networks*, *26*, 99–109.

Liu, Q., & Wang, J. (2013). A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints. *IEEE Transactions on Neural Networks and Learning Systems*, *24*, 812–824.

Liu, S., & Wang, J. (2006). A simplified dual neural network for quadratic programming with its KWTA application. *IEEE Transactions on Neural Networks*, *17*(6), 1500–1510.

Nazemi, A. R. (2012). A dynamic system model for solving convex nonlinear optimization problems. *Communications in Nonlinear Science and Numerical Simulation*, *17*, 1696–1705.

Nistri, P., & Quincampoix, M. (2005). On the dynamics of a differential inclusion built upon a nonconvex constrained minimization problem. *Journal of Optimization Theory and Applications*, *124*(3), 659–672.

Opial, Z. (1967). Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, *73*, 591–597.

Pardalos, P. 2008. Nonconvex optimization and its application. Berlin, Heidelberg.

Piche, S., Sayyar-Rodsari, B., Johnson, D., & Gerules, M. (2000). Nonlinear model predictive control using neural networks. *IEEE Control Systems*, *20*, 53–62.

Pillo, G. D., & Grippo, L. (1989). Exact penalty function in constrained optimization. *SIAM Journal on Control and Optimization*, *27*(6), 1333–1360.

Tank, D. W., & Hopfield, J. J. (1986). Simple 'neural' optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Transactions on Circuits and Systems*, *33*, 533–541.

Wang, J. (1994). A deterministic annealing neural network for convex programming. *Neural Networks*, *7*, 629–641.

Wang, J., Hu, Q., & Jiang, D. (1999). A Lagrangian network for kinematic control of redundant robot manipulators. *IEEE Transactions on Neural Networks*, *10*, 1123–1132.

Xia, Y. (1996). A new neural network for solving linear and quadratic programming problems. *IEEE Transactions on Neural Networks*, *7*(6), 1544–1548.

Xia, Y. (2003). Global convergence analysis of Lagrangian networks. *IEEE Transactions on Circuits and Systems*, *50*(6), 818–822.

Xia, Y., Feng, G., & Wang, J. (2004). A recurrent neural network with exponential convergence for solving convex quadratic program and related linear piecewise equations. *Neural Networks*, *17*(7), 1003–1005.

Xia, Y., Feng, G., & Wang, J. (2008). A novel neural network for solving nonlinear optimization problems with inequality constraints. *IEEE Transactions on Neural Networks*, *19*, 1340–1353.

Xia, Y., Leung, H., & Wang, J. (2002). A projection neural network and its application to constrained optimization problems. *IEEE Transactions on Circuits and Systems. I*, *49*, 447–458.

Xue, X., & Bian, W. (2008). Subgradient-based neural networks for nonsmooth convex optimization problems. *IEEE Transactions on Circuits and Systems. I*, *55*, 2378–2391.

Yan, Z., & Wang, J. (2012). Model predictive control of nonlinear systems with unmodeled dynamics based on feedforward and recurrent neural networks. *IEEE Transactions on Industrial Informatics*, *8*, 746–756.