Smoothing Neural Network for Constrained Non-Lipschitz Optimization With Applications

Wei Bian and Xiaojun Chen

***Abstract*— In this paper, a smoothing neural network (SNN) is proposed for a class of constrained non-Lipschitz optimization problems, where the objective function is the sum of a nonsmooth, nonconvex function, and a non-Lipschitz function, and the feasible set is a closed convex subset of $\mathbb{R}^n$. Using the smoothing approximate techniques, the proposed neural network is modeled by a differential equation, which can be implemented easily. Under the level bounded condition on the objective function in the feasible set, we prove the global existence and uniform boundedness of the solutions of the SNN with any initial point in the feasible set. The uniqueness of the solution of the SNN is provided under the Lipschitz property of smoothing functions. We show that any accumulation point of the solutions of the SNN is a stationary point of the optimization problem. Numerical results including image restoration, blind source separation, variable selection, and minimizing condition number are presented to illustrate the theoretical results and show the efficiency of the SNN. Comparisons with some existing algorithms show the advantages of the SNN.**

***Index Terms*— Image and signal restoration, non-Lipschitz optimization, smoothing neural network, stationary point, variable selection.**

I. Introduction

NEURAL networks are promising numerical methods for solving a number of optimization problems in engineering, sciences, and economics [1]–[4]. The structure of a neural network can be implemented physically by designated hardware such as specific integrated circuits where the computational procedure is distributed and parallel. Hence, the neural network approach can solve optimization problems in running time at the order of magnitude much faster than the conventional optimization algorithms executed on general digital computers [5], [6]. Moreover, neural networks can solve many optimization problems with time-varying parameters [7], [8] since the dynamical techniques can be applied to the continuous-time neural networks.

Recently, neural networks for solving optimization problems have been extensively studied. In 1986, Tank and Hopfield introduced the Hopfield neural network for solving the linear programming problem [9]. This paper inspired many researchers to develop various neural networks for solving optimization problems. Kennedy and Chua proposed a neural network for solving a class of nonlinear programming problems in [10].

We have noticed that in many important applications, the optimization problems are not differentiable. However, the neural networks for smooth optimization problems cannot solve nonsmooth optimization problems, because the gradients of the objective and constrained functions are required in such neural networks. Hence, the study of neural networks for nonsmooth optimization problems is necessary. In [11], Forti, Nistri, and Quincampoix presented a generalized neural network for solving a class of nonsmooth convex optimization problems. In [12], using the penalty function method and differential inclusion, Xue and Bian defined a neural network for solving a class of nonsmooth convex optimization problems. Other interesting results for nonsmooth convex optimization using neural networks can be found in [13]–[15].

The efficiency of the neural networks for solving convex optimization problems relies on the convexity of functions. A neural network for a nonconvex quadratic optimization is presented in [16]. Xia, Feng, and Wang gave a recurrent neural network for solving a class of differentiable and nonconvex optimization problems in [6]. Another neural network using a differential inclusion for a class of nonsmooth and nonconvex optimization problems was proposed in [17]. Some results on the convergence analysis of neural networks with discontinuous activations are given in [18]–[20], which are potentially useful for using the neural networks to solve nonsmooth optimization problems.

For nonsmooth optimization, the Lipschitz continuity is a necessary condition to define the Clarke stationary point. However, some real-world optimization problems are non-Lipschitz, such as problems in image restoration, signal recovery, variable selection, etc. For example, the l_2-l_p problem

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2+\lambda\|x\|_p^p \tag{1}$$

with $0 < p < 1$. Problem (1) attracts great attention in variable selection and sparse reconstruction [21]–[33]. The $l_2 - l_1$ problem is also known as Lasso [24], which is a

Manuscript received June 16, 2011; accepted December 20, 2011. Date of publication January 5, 2012; date of current version February 29, 2012. The work of W. Bian was supported in part by Hong Kong Polytechnic University Postdoctoral Fellowship Scheme, the National Science Foundation of China, under Grant 11101107 and Grant 10971043, Heilongjiang Province Foundation under Grant A200803, the Education Department of Heilongjiang Province under Grant 1251112, and the Project HIT. NSRIF. 2009048. The work of X. Chen was supported in part by Hong Kong Research Grant Council under Grant PolyU5005/09P.

W. Bian is with the Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China. She is now with the Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: bianweilvse520@163.com).

X. Chen is with the Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: maxjchen@polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2011.2181867

convex optimization problem and can be efficiently solved. In some cases, some solutions of the l_2-l_1 problem are in the solution set of the corresponding l_2-l_0 problem [27]. l_1 regularization provides the best convex approximations to l_0 regularization and it is computationally efficient. Nevertheless, l_1 regularization can not handle the collinearity problem and may yield inconsistent selections when applied to variable selection in some situations. Moreover, l_1 regularization often introduces extra bias in estimation [23] and cannot recovery a signal or image with the least measurements when it is applied to compressed sensing [25]–[31].

In [31], Chartrand and Staneva showed that by replacing the l_1 norm with the l_p $(p < 1)$ norm, exact reconstruction was possible with substantially fewer measurements. Using the basis pursuit method, Saab, Yilmaz, Mckeown, and Abugharbieh [25] illustrated experimentally that the separation performance for underdetermined blind source separation of anechoic speech mixtures was improved when one used l_p basis pursuit with $p < 1$ compared to $p = 1$. Chen, Xu, and Ye [32] established a lower bound theorem to classify zero and nonzero entries in every local solution of l_2-l_p problem (1). Recently, it has been proved that finding a global minimizer of (1) is strongly NP-hard [33].

The following problem is a generalized version of (1):

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{n} \psi(|A_i x - b_i|) + \lambda \sum_{i=1}^{d} \omega_i(|x_i|) \tag{2}$$

where A_i is the ith row of A, ω_i is a non-Lipschitz function, and ψ is the Huber function [34], [35].

Penalty optimization problems are also an important source of non-Lipschitz optimization problems. Penalty techniques are widely used in constrained optimization. However, using smooth convex penalty functions cannot get exact solutions of the original problems. Non-Lipschitz penalty functions have been used to get exact solutions [36].

Basing on the above source problems, in this paper, we consider the following optimization problem:

$$\begin{aligned} \min \quad & f(x) + \sum_{i=1}^{r} \varphi\left(|d_i^T x|^p\right) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned} \tag{3}$$

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is a locally Lipschitz function, $\mathcal{X}$ is a closed convex subset of $\mathbb{R}^n$, $p \in (0, 1)$ is a positive parameter and $d_i \in \mathbb{R}^n$, $i = 1, 2, \ldots, r$, $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is a differentiable and globally Lipschitz function with Lipschitz constant l_φ.

Some popular formats of φ satisfying the above conditions can be given by

$$\varphi(z) = \lambda z, \quad \varphi(z) = \lambda \frac{\alpha z}{1 + \alpha z}, \quad \varphi(z) = \lambda \log(\alpha z + 1) \tag{4}$$

where λ and α are positive parameters.

In our optimization model (3), the function $|\cdot|^p$ for $0 < p < 1$ is neither convex nor Lipschitz continuous, and f is not necessarily convex or smooth. It is worth noting that our analysis can be easily extended to the following problem:

$$\begin{aligned} \min \quad & f(x) + \sum_{i=1}^{r} \varphi_i\left(|d_i^T x|^p\right) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned} \tag{5}$$

that is, these functions φ_i are not necessarily same.

Specially, when d_i is the ith column of the identity matrix with dimension $n \times n$, then (5) is reformatted as

$$\begin{aligned} \min \quad & f(x) + \sum_{i=1}^{n} \varphi_i\left(|x_i|^p\right) \\ \text{s.t.} \quad & x \in \mathcal{X}. \end{aligned} \tag{6}$$

To the best of our knowledge, all methods for solving non-Lipschitz optimization problems are discrete, such as reweighted l_1 methods, orthogonal matching pursuit algorithms, iterative thresholds algorithms, smoothing conjugate gradient methods, etc. In this paper, for the first time, we use the continuous neural network method to solve non-Lipschitz optimization problem (3) and we adopt the smoothing techniques [21], [37]–[39] into the proposed neural network. Almost all neural networks for solving optimization problems are relying on the Clarke generalized gradient, which is defined for the locally Lipschitz function. For the non-Lipschitz items in (3), the Clarke generalized gradient cannot be defined and it is difficult to give a kind of gradient of the non-Lipschitz functions, which are effective for optimization and can be calculated easily. However, the smoothing approximations of the non-Lipschitz items used in this paper are continuously differentiable and their gradients can be calculated easily. Moreover, from the theoretical analysis and numerical experiments in this paper, we can find that they are effective for solving optimization problem (3). This is the main reason that we use the neural network based on the smoothing method instead of using the neural network directly to solve (3). Moreover, there are some difficulties in solving and implementing neural networks modeled by differential inclusion [11], [12], [16], [17]. The neural network based on smoothing techniques for nonsmooth optimization is modeled by a differential equation instead of differential inclusion, which can avoid solving differential inclusion. Since the objective function of (3) is non-Lipschitz, the Clarke stationary point of (3) cannot be defined. We introduce a new definition of stationary points of optimization problem (3).

We will use the projection operator to handle the constraint in (3) since the feasible set $\mathcal{X}$ is a closed convex set. Moreover, in many applications, the projection operator has an explicit form, for instance, $\mathcal{X}$ is defined by box constraints or affine equality constraints.

The remaining part of this paper is organized as follows. In Section II, we introduce some necessary definitions and preliminary results including the definitions of smoothing function and stationary point of (3). In Section III, we propose a smoothing neural network (SNN) modeled by a differential equation to solve (3). We study the global existence and limit behavior of the solutions of the proposed neural network. In theory, we prove that any accumulation point of the solutions of the proposed neural network is a stationary point of (3).

Especially, when the proposed neural network is applied to solve (6), its accumulation points can satisfy a stronger inclusion property which reduces to the definition of the Clarke stationary point for $p \geq 1$. In Section IV, we present numerical results to show the efficiency of the SNN for image restoration, blind source separation, variable selection, and optimizing condition number.

Notation: Given column vectors $x = (x_1, x_2, \dots, x_n)^T$ and $y = (y_1, y_2, \dots, y_n)^T$, $\langle x, y\rangle = x^T y = \sum_{i=1}^n x_i y_i$ is the scalar product of x and y. x_i denotes the ith element of x. $\|x\|$ denotes the Euclidean norm defined by $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$. For a matrix $A \in \mathbb{R}^{n\times n}$, the norm $\|A\|$ is defined by $\|A\| = \max_{\|x\|=1} \|Ax\|$. For any $y \in \mathbb{R}$, $\nabla\varphi(y)$ means the value of $\nabla\varphi$ at a point y. For a closed convex subset $\Omega \subseteq \mathbb{R}^n$, $\mathrm{int}(\Omega)$, and $\mathrm{cl}(\Omega)$ mean the interior and closure of Ω in $\mathbb{R}^n$, respectively. For a subset $U \subseteq \mathbb{R}^n$, co U indicates the convex hull of U. $P_\Omega : \mathbb{R}^n \to \Omega$ is the projection operator from $\mathbb{R}^n$ onto Ω and $N_\Omega(x)$ means the normal cone of Ω at x. $AC[0, \infty)$ represents the set of all absolutely continuous functions $x : [0, \infty) \to \mathbb{R}^n$.

II. Preliminary Results

In this section, we state some definitions and properties needed in this paper. We refer the readers to [39]–[42].

Definitions of local Lipschitz and regular real-valued functions, upper semicontinuous and lower semicontinuous set-valued functions can be found in [40].

Suppose that h: $\mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz, hence h is differentiable almost everywhere. Let D_h denote the set of points at which h is differentiable. Then the Clarke generalized gradient is given by

$$\partial h(x) = \mathrm{co}\left\{ \lim_{x_k \to x;\, x_k \in D_h} \nabla h(x_k) \right\}. \tag{7}$$

Proposition 2.1 ([40]): For any fixed $x \in \mathbb{R}^n$, the following statements hold.

1) $\partial h(x)$ is a nonempty, convex and compact set of $\mathbb{R}^n$.
2) ∂h is upper semicontinuous at x.

For a nonempty, closed and convex set Ω in $\mathbb{R}^n$, the definitions of the tangent and normal cones of Ω are as in [40].

From [40, Proposition 2.4.2], we know that

$$N_\Omega(x) = \mathrm{cl}\left(\bigcup_{\nu \geq 0} \nu \partial d_\Omega(x) \right) \tag{8}$$

where $d_\Omega(x) = \min_{u\in\Omega} \|x - u\|$ is a globally Lipschitz function.

The *projection operator* to Ω at x is defined by

$$P_\Omega(x) = \arg\min_{u\in\Omega} \|u - x\|^2$$

and it has the following properties.

Proposition 2.2 ([42]):

$$\langle v - P_\Omega(v), P_\Omega(v) - u\rangle \geq 0 \quad \forall\, v \in \mathbb{R}^n, \quad u \in \Omega;$$
$$\|P_\Omega(u) - P_\Omega(v)\| \leq \|u - v\| \quad \forall\, u, v \in \mathbb{R}^n.$$

Many smoothing approximations for nonsmooth optimization problems have been developed in the past decades. The main feature of smoothing methods is to approximate the nonsmooth functions by the parameterized smooth functions. This paper uses a class of smoothing functions defined as follows.

Definition 2.1: Let $h : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. We call $\tilde{h} : \mathbb{R}^n \times [0, +\infty) \to \mathbb{R}$ a *smoothing function* of h, if $\tilde{h}$ satisfies the following conditions.

1) For any fixed $\mu > 0$, $\tilde{h}(\cdot, \mu)$ is continuously differentiable in $\mathbb{R}^n$, and for any fixed $x \in \mathbb{R}^n$, $\tilde{h}(x, \cdot)$ is differentiable in $[0, +\infty)$.
2) For any fixed $x \in \mathbb{R}^n$, $\lim_{\mu\downarrow 0} \tilde{h}(x, \mu) = h(x)$.
3) There is a positive constant $\kappa_{\tilde{h}} > 0$ such that
$$|\nabla_\mu \tilde{h}(x, \mu)| \leq \kappa_{\tilde{h}} \quad \forall\, \mu \in [0, \infty), \quad x \in \mathbb{R}^n.$$
4) $\{\lim_{z\to x, \mu\downarrow 0} \nabla_z \tilde{h}(z, \mu)\} \subseteq \partial h(x)$.

We should state that (2) and (3) of Definition 2.1 imply that for any fixed $x \in \mathbb{R}^n$

$$\lim_{z\to x, \mu\downarrow 0} \tilde{h}(z, \mu) = h(x) \tag{9}$$

and

$$|\tilde{h}(x, \mu) - h(x)| \leq \kappa_{\tilde{h}}\mu \quad \forall\, \mu \in [0, \infty), \quad x \in \mathbb{R}^n. \tag{10}$$

Many existing results in [37]–[39] give us some theoretical basis for constructing smoothing functions. Throughout this paper, we always denote $\tilde{f}$ a smoothing function of f in (3).

Next, we will focus on the smoothing approximations of non-Lipschitz items $\varphi(|d_i^T x|^p)$, $i = 1, 2, \dots, r$. In this paper, we use the smooth functions $\varphi(\theta^p(d_i^T x, \mu))$ to approximate $\varphi(|d_i^T x|^p)$, $i = 1, 2, \dots, r$, where $\theta(s, \mu)$ is a smoothing function of $|s|$ defined by

$$\theta(s, \mu) = \begin{cases} |s|, & \text{if } |s| > \mu \\ \dfrac{s^2}{2\mu} + \dfrac{\mu}{2}, & \text{if } |s| \leq \mu. \end{cases} \tag{11}$$

Since $\lim_{\mu\downarrow 0} \theta(d_i^T x, \mu) = |d_i^T x|$, we can easily obtain that

$$\lim_{\mu\downarrow 0} \varphi(\theta^p(d_i^T x, \mu)) = \varphi(|d_i^T x|^p) \quad i = 1, 2, \dots, r. \tag{12}$$

For any fixed $\mu > 0$, $i = 1, 2, \dots, r$, we obtain

$$\nabla_x \theta^p(d_i^T x, \mu) = \begin{cases} p|d_i^T x|^{p-1}\mathrm{sign}(d_i^T x) d_i, & \text{if } |d_i^T x| > \mu \\ p\left(\frac{|d_i^T x|^2}{2\mu} + \frac{\mu}{2}\right)^{p-1} \frac{d_i^T x d_i}{\mu}, & \\ & \text{if } |d_i^T x| \leq \mu. \end{cases} \tag{13}$$

Then, we have

$$\nabla_x \varphi(\theta^p(d_i^T x, \mu)) = \nabla\varphi(\theta^p(d_i^T x, \mu)) \cdot \nabla_x \theta^p(d_i^T x, \mu) \tag{14}$$

which means that $\varphi(\theta^p(d_i^T x, \mu))$ is continuously differentiable in x for any fixed $\mu > 0$, $i = 1, 2, \dots, r$.

On the other hand, for any fixed $x \in \mathbb{R}^n$, we have

$$\nabla_\mu \theta^p(d_i^T x, \mu) = p\theta^{p-1}(d_i^T x, \mu)\nabla_\mu \theta(d_i^T x, \mu) \tag{15}$$

where

$$\nabla_\mu \theta(d_i^T x, \mu) = \begin{cases} 0, & \text{if } |d_i^T x| > \mu \\ -\dfrac{|d_i^T x|^2}{2\mu^2} + \dfrac{1}{2}, & \text{if } |d_i^T x| \leq \mu. \end{cases} \tag{16}$$

Thus

$$\nabla_\mu \varphi(\theta^p(d_i^T x, \mu)) = \nabla \varphi(\theta^p(d_i^T x, \mu)) \cdot \nabla_\mu \theta^p(d_i^T x, \mu) \quad (17)$$

which means that $\varphi(\theta^p(d_i^T x, \mu))$ is continuously differentiable in μ for any fixed $x \in \mathbb{R}^n$, $i = 1, 2, \ldots, r$.

In [32], Chen, Xu, and Ye studied unconstrained l_2-l_p optimization problem (1) and gave a definition for the first order necessary condition of problem (1), where the objective function is not Lipschitz continuous.

Combining the first order necessary condition of [32, 1] and the definition of the Clarke stationary point, we define the stationary point of (3) as follows.

Definition 2.2: For $x^* \in \mathcal{X}$, we call x^* a *stationary point* of (3) if

$$0 \in (x^*)^T \partial f(x^*) + p \sum_{i=1}^{r} |d_i^T x^*|^p \nabla \varphi \left(|d_i^T x^*|^p\right) + (x^*)^T N_{\mathcal{X}}(x^*).$$

III. Main Results

In this paper, we propose a SNN modeled by the following differential equation:

$$\begin{cases} \dot{x}(t) = -x(t) + P_{\mathcal{X}}[x(t) - \nabla_x \tilde{f}(x(t), \mu(t)) \\ \qquad - \sum_{i=1}^{r} \nabla_x \varphi(\theta^p(d_i^T x(t), \mu(t)))] \\ x(0) = x_0 \end{cases}$$

where $\tilde{f} : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}$ is a smoothing function of f, $\mu : [0, \infty) \to [0, \infty)$ is a continuously differentiable and decreasing function satisfying $\lim_{t\to\infty} \mu(t) = 0$.

Remark 3.1: We can choose

$$\mu(t) = \frac{\mu_0}{(t+1)^\alpha} \quad \text{or} \quad \mu(t) = \mu_0 e^{-\alpha t}$$

where μ_0 and α are positive numbers.

The proposed neural network SNN is modeled by a nonautonomous differential equation. There are many kinds of nonautonomous neural network models, such as the Cohen-Grossberg-Type neural network in [43] and the uncertain stochastic Hopfield neural network in [44], which have many implementation methods. The proposed network SNN can be seen as a neural network with one continuous variable input $\mu(t)$. The dynamic differential equation in the SNN can be realized in a network with the architecture block diagram structure in Fig. 1.

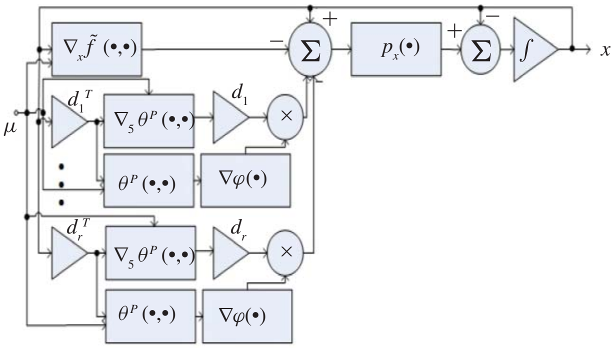


Fig. 1. Architecture block diagram structure of the SNN.

To present the main results of this paper, we need some properties of $\varphi(\theta^p(d_i^T x, \mu))$, $i = 1, 2, \ldots, r$ as follows.

Proposition 3.1: For any $i = 1, 2, \ldots, r$, the following statements hold.

1) $|\nabla_\mu \varphi(\theta^p(d_i^T x, \mu))| \le p l_\varphi \mu^{p-1}$ $\forall x \in \mathbb{R}^n$.
2) $|\varphi(\theta^p(d_i^T x, \mu)) - \varphi(|d_i^T x|^p)| \le l_\varphi \mu^p$ $\forall x \in \mathbb{R}^n$.
3) $\nabla_x \theta^p(d_i^T x, \mu)$ is globally Lipschitz in x for any fixed $\mu > 0$.
4) If $\nabla \varphi$ is locally (globally) Lipschitz on $[0, +\infty)$, then $\nabla_x \varphi(\theta^p(d_i^T x, \mu))$ is locally (globally) Lipschitz in x for any fixed $\mu > 0$.

Proof: See Appendix A. ■

Remark 3.2: The gradients of the three functions φ given in (4) are globally Lipschitz.

Now, we study the global existence, boundedness and feasibility of the solutions of the SNN. For simplicity, we use x_t and μ_t to denote $x(t)$ and $\mu(t)$ in the proofs of the main results of this paper, respectively.

Theorem 3.1: Suppose the level set

$$\left\{ x \in \mathcal{X} : \quad f(x) + \sum_{i=1}^{r} \varphi(|d_i^T x|^p) \le \gamma \right\}$$

is bounded for any $\gamma > 0$. Then, for any initial point $x_0 \in \mathcal{X}$, the SNN has a solution $x : [0, \infty) \to \mathbb{R}^n$ in $AC[0, \infty)$. Moreover, any solution $x \in AC[0, \infty)$ of the SNN satisfies:

1) $x(t) \in \mathcal{X}$, $\forall t \in [0, \infty)$;
2) there is a $\rho > 0$ such that $\sup_{t\in[0,\infty)} \|x(t)\| \le \rho$.

Proof: Since the right-hand function of the SNN is continuous in x and t, there are a $T > 0$ and an absolutely continuous function $x : [0, T) \to \mathbb{R}^n$ such that it is a local solution of the SNN [45]. Suppose that $[0, T)$ is the maximal existence interval of this solution.

First, we prove that $x_t \in \mathcal{X}$, $\forall t \in [0, T)$. The SNN can be rewritten as

$$\dot{x}_t + x_t = h(t)$$

where

$$h(t) = P_{\mathcal{X}} \left[x_t - \nabla_x \tilde{f}(x_t, \mu_t) - \sum_{i=1}^{r} \nabla_x \varphi(\theta^p(d_i^T x_t, \mu_t)) \right]$$

is a continuous function taking its values in $\mathcal{X}$. A simple integration procedure gives

$$x_t = e^{-t} x_0 + e^{-t} \int_0^t h(s) e^s ds$$

which can be rewritten as

$$x_t = e^{-t} x_0 + (1 - e^{-t}) \int_0^t h(s) \frac{e^s}{e^t - 1} ds. \quad (18)$$

Since $(e^s/e^t - 1) > 0$, $h(s) \in \mathcal{X}$, $\forall\, 0 \le s \le t$, $\int_0^t (e^s/e^t - 1)\, ds = 1$ and $\mathcal{X}$ is convex, we have

$$\int_0^t h(s) \frac{e^s}{e^t - 1} ds \in \mathcal{X} \quad \forall t \in [0, T).$$

Combining the above result and $x_0 \in \mathcal{X}$ in (18), we obtain that

$$x_t \in \mathcal{X} \quad \forall t \in [0, T).$$

Differentiating $\tilde{f}(x_t, \mu_t)+\sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t))$ along this solution of the SNN, we obtain

$$\begin{aligned} &\frac{d}{dt}\left[\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right)\right] \\ &= \left\langle \nabla_x \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \nabla_x \varphi(\theta^p(d_i^T x_t, \mu_t)), \dot{x}_t \right\rangle \\ &\quad + \left(\nabla_\mu \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \nabla_\mu \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right)\right)\dot{\mu}_t. \end{aligned} \tag{19}$$

Let $\upsilon = x_t - \nabla_x \tilde{f}(x_t, \mu_t) - \sum_{i=1}^{r} \nabla_x \varphi(\theta^p(d_i^T x_t, \mu_t))$ and $u = x_t$ in Proposition 2.2, from the SNN, we have

$$\left\langle \nabla_x \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \nabla_x \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right), \dot{x}_t \right\rangle \le -\|\dot{x}_t\|^2. \tag{20}$$

From Definition 2.1 and $\dot{\mu}_t \le 0$, $\forall t \in [0, T)$, there is a $\kappa_{\tilde{f}} > 0$ such that

$$\nabla_\mu \tilde{f}(x_t, \mu_t)\dot{\mu}_t \le -\kappa_{\tilde{f}}\dot{\mu}_t. \tag{21}$$

By Proposition 3.1, we obtain

$$\sum_{i=1}^{r} \nabla_\mu \varphi(\theta^p(d_i^T x_t, \mu_t))\dot{\mu}_t \le -rpl_\varphi \mu_t^{p-1}\dot{\mu}_t. \tag{22}$$

Substituting (20)–(22) into (19) we have

$$\begin{aligned} &\frac{d}{dt}\Bigg[\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right) \\ &\qquad + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p\Bigg] \le -\|\dot{x}_t\|^2. \end{aligned} \tag{23}$$

This implies that $\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t)) + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p$ is a non-increasing function in t on $[0, T)$, which follows that for any $t \in [0, T)$

$$\begin{aligned} &\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right) + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p \\ &\le \tilde{f}(x_0, \mu_0) + \sum_{i=1}^{r} \varphi\left(\theta^p(d_i^T x_0, \mu_0)\right) + \kappa_{\tilde{f}}\mu_0 + rl_\varphi \mu_0^p. \end{aligned}$$

On the other hand, from (10) and Proposition 3.1 in (2), we obtain

$$\begin{aligned} &\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right) + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p \\ &\ge f(x_t) + \sum_{i=1}^{r} \varphi(|d_i^T x_t|^p). \end{aligned} \tag{24}$$

Thus, for any $t \in [0, T)$

$$\begin{aligned} f(x_t) + \sum_{i=1}^{r} \varphi\left(|d_i^T x_t|^p\right) &\le \tilde{f}(x_0, \mu_0) \\ &+ \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_0, \mu_0)) + \kappa_{\tilde{f}}\mu_0 + rl_\varphi \mu_0^p. \end{aligned} \tag{25}$$

Owning to the level boundedness of the objective function in $\mathcal{X}$, we obtain $x : [0, T) \to \mathbb{R}^n$ is bounded. Using the theorem about the extension of a solution [45], we get that this solution of the SNN can be extended, which leads a contradiction to the supposition that $[0, T)$ is the maximal existence interval of x_t. Therefore, this solution of the SNN exists globally.

Moreover, since (25) holds for all $t \in [0, \infty)$, from the level bounded assumption in this theorem, this solution x_t of the SNN is uniformly bounded, which means there is a $\rho > 0$ such that

$$\|x_t\| \le \rho \qquad \forall t \in [0, \infty).$$ ■

The next proposition shows that the locally Lipschitz continuity of $\nabla\varphi(\cdot)$ and $\nabla_x \tilde{f}(\cdot, \mu)$ can guarantee the uniqueness of the solution of the SNN.

Proposition 3.2: With the conditions of Theorem 3.1, if $\nabla\varphi(\cdot)$ and $\nabla_x \tilde{f}(\cdot, \mu)$ is locally Lipschitz for any fixed $\mu > 0$, then the solution of the SNN with an initial point $x_0 \in \mathcal{X}$ is unique.

Proof: See Appendix B. ■

The next theorem gives the main results of this paper, which presents some convergent properties of the SNN.

Theorem 3.2: With the conditions of Theorem 3.1, any solution $x \in AC[0, \infty)$ of the SNN with an initial point $x_0 \in \mathcal{X}$ satisfies the following properties:

1) $\lim_{t\to\infty} f(x(t)) + \sum_{i=1}^{r} \varphi(|d_i^T x(t)|^p)$ exists;
2) $\dot{x}(t) \in L^2[0, \infty)$ and $\lim_{t\to\infty} \|\dot{x}(t)\| = 0$;
3) any accumulation point of $x(t)$ is a stationary point of (3);
4) if all stationary points of (3) are isolated, then $x(t)$ converges to one stationary point of (3) as $t \to \infty$.

Proof: From Theorem 3.1, there is a $\rho > 0$ such that $\|x_t\| \le \rho$, $\forall t \ge 0$. Thus, from (24), we obtain

$$\begin{aligned} &\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi\left(\theta^p(d_i^T x_t, \mu_t)\right) + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p \\ &\ge \inf_{x\in\{x:\|x\|\le\rho\}} \left[f(x) + \sum_{i=1}^{r} \varphi(|d_i^T x|^p)\right]. \end{aligned} \tag{26}$$

Equations (23) and (26) imply that

$$\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t)) + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p$$

is non-increasing and bounded from below on $[0, \infty)$, then

$$\lim_{t\to\infty}\left[\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t)) + \kappa_{\tilde{f}}\mu_t + rl_\varphi \mu_t^p\right] \tag{27}$$

exists.

Thus, from (9), Proposition 3.1 and $\lim_{t\to\infty}\mu_t = 0$, we obtain

$$\lim_{t\to\infty}\left[\tilde{f}(x_t,\mu_t)+\sum_{i=1}^{r}\varphi(\theta^p(d_i^T x_t,\mu_t))+\kappa_{\tilde{f}}\mu_t+rl_\varphi\mu_t^p\right]$$
$$=\lim_{t\to\infty}\left[f(x_t)+\sum_{i=1}^{r}\varphi(|d_i^T x_t|^p)\right].$$

On the other hand, (23) and (27) imply that

$$\lim_{t\to\infty}\frac{d}{dt}\left[\tilde{f}(x_t,\mu_t)+\sum_{i=1}^{r}\varphi(\theta^p(d_i^T x_t,\mu_t))\right.$$
$$\left.+\kappa_{\tilde{f}}\mu_t+rl_\varphi\mu_t^p\right]=0. \tag{28}$$

Owning to (23), we obtain

$$\lim_{t\to\infty}\|\dot{x}_t\|=0.$$

From (26), integrating (23) from 0 to t gives

$$\int_0^\infty\|\dot{x}_t\|^2dt<\infty.$$

Since $x:[0,\infty)\to\mathbb{R}^n$ is uniformly bounded, x_t has at least one accumulation point. If x^* is an accumulation point of x_t, there exists a sequence $\{t_k\}$ such that $\lim_{k\to\infty}x_{t_k}=x^*$ and $\lim_{k\to\infty}\mu_{t_k}=0$ as $\lim_{k\to\infty}t_k=\infty$.

From Theorem 3.1, we know that $x^*\in\mathcal{X}$.

Let

$$g_{t_k}=x_{t_k}-P_{\mathcal{X}}\left[x_{t_k}-\nabla_x\tilde{f}(x_{t_k},\mu_{t_k})\right.$$
$$\left.-\sum_{i=1}^{r}\nabla_x\varphi(\theta^p(d_i^T x_{t_k},\mu_{t_k}))\right]. \tag{29}$$

Owning to $\lim_{k\to\infty}\dot{x}_{t_k}=0$ and the SNN, we obtain

$$\lim_{k\to\infty}g_{t_k}=0. \tag{30}$$

Using Proposition 2.2 into (29), we have

$$\Big\langle\nabla_x\tilde{f}(x_{t_k},\mu_{t_k})+\textstyle\sum_{i=1}^{r}\nabla_x\varphi(\theta^p(d_i^T x_{t_k},\mu_{t_k}))-g_{t_k},$$
$$u-(x_{t_k}-g_{t_k})\Big\rangle\ge 0\quad\forall u\in\mathcal{X}. \tag{31}$$

From the definition of normal cone and the above inequality, we have

$$0\in\nabla_x\tilde{f}(x_{t_k},\mu_{t_k})+\sum_{i=1}^{r}\nabla_x\varphi\left(\theta^p(d_i^T x_{t_k},\mu_{t_k})\right)-g_{t_k}$$
$$+N_{\mathcal{X}}(x_{t_k}-g_{t_k}). \tag{32}$$

From (32) and

$$N_{\mathcal{X}}(x_{t_k}-g_{t_k})=\bigcup_{\nu\ge0}\nu\partial d_{\mathcal{X}}(x_{t_k}-g_{t_k})$$

there exist $\nu_{t_k}\ge0$ and $\xi_{t_k}\in\partial d_{\mathcal{X}}(x_{t_k}-g_{t_k})$ such that

$$0=\nabla_x\tilde{f}(x_{t_k},\mu_{t_k})+\sum_{i=1}^{r}\nabla_x\varphi(\theta^p(d_i^T x_{t_k},\mu_{t_k}))-g_{t_k}+\nu_{t_k}\xi_{t_k}.$$

From Definition 2.1, we have

$$\left\{\lim_{k\to\infty}\nabla_x\tilde{f}(x_{t_k},\mu_{t_k})\right\}\subseteq\partial f(x^*). \tag{33}$$

Since $x\multimap\partial d_{\mathcal{X}}(x)$ is upper semicontinuous and $\lim_{k\to\infty}x_{t_k}-g_{t_k}=x^*$, then

$$\left\{\lim_{k\to\infty}\xi_{t_k}\right\}\subseteq\partial d_{\mathcal{X}}(x^*).$$

Combining the above relationship with $\lim_{k\to\infty}x_{t_k}=x^*$, we have

$$\left\{\lim_{k\to\infty}\nu_{t_k}\xi_{t_k}\right\}\subseteq\bigcup_{\nu\ge0}\nu\partial d_{\mathcal{X}}(x^*)=N_{\mathcal{X}}(x^*). \tag{34}$$

Thus

$$0\in\partial f(x^*)+\sum_{i=1}^{r}\lim_{k\to\infty}\nabla_x\varphi(\theta^p(d_i^T x_{t_k},\mu_{t_k}))+N_{\mathcal{X}}(x^*). \tag{35}$$

From (13), we obtain

$$x_{t_k}^T\nabla_x\theta^p(d_i^T x_{t_k},\mu_{t_k})$$
$$=\begin{cases}p|d_i^T x_{t_k}|^p, & \text{if } |d_i^T x_{t_k}|>\mu_{t_k}\\ p\left(\frac{|d_i^T x_{t_k}|^2}{2\mu_{t_k}}+\frac{\mu_{t_k}}{2}\right)^{p-1}\frac{|d_i^T x_{t_k}|^2}{\mu_{t_k}}, & \text{if } |d_i^T x_{t_k}|\le\mu_{t_k}.\end{cases}$$

Thus

$$\lim_{k\to\infty}x_{t_k}^T\nabla_x\theta^p\left(d_i^T x_{t_k},\mu_{t_k}\right)=p|d_i^T x^*|^p$$

which implies that

$$\lim_{k\to\infty}x(t_k)^T\sum_{i=1}^{r}\nabla_x\varphi\left(\theta^p(d_i^T x_{t_k},\mu_{t_k})\right)$$
$$=p\sum_{i=1}^{r}\nabla\varphi\left(|d_i^T x^*|^p\right)|d_i^T x^*|^p. \tag{36}$$

Therefore, we obtain

$$0\in(x^*)^T\partial f(x^*)+p\sum_{i=1}^{r}|d_i^T x^*|^p\nabla\varphi(|d_i^T x^*|^p)+(x^*)^T N_{\mathcal{X}}(x^*)$$

which means x^* is a stationary point of (3).

If x_t has two accumulation points x^* and y^*, there are two sequences t_k and s_k such that

$$\lim_{k\to\infty}t_k=\lim_{k\to\infty}s_k=\infty,\ \lim_{k\to\infty}x_{t_k}=x^*\ \text{ and }\lim_{k\to\infty}x_{s_k}=y^*.$$

Since $x(t)$ is continuous on $[0,\infty)$ and uniformly bounded, there exists a path from x^* to y^*, denoted $w(x^*,y^*)$, such that any point on $w(x^*,y^*)$ is a stationary point of (3), which leads a contradiction to that x^* and y^* are isolated. ■

Remark 3.3: When $p\ge1$, the critical point set of (3) is

$$\mathcal{C}=\left\{x\in\mathcal{X}:0\in\partial f(x)+\sum_{i=1}^{r}\partial\varphi(|d_i^T x|^p)+N_{\mathcal{X}}(x)\right\}$$

which coincides with the optimal solution set of (3) when f is convex.

From (35), we can confirm that the solution of the SNN is convergent to the critical point set of (3) when $p\ge1$, and to the optimal solution set of (3) when f is convex and $p\ge1$.

Moreover, if x^* is a limit point of the solution of the SNN such that $d_i^T x^* \neq 0, i = 1, 2, \ldots, r$, then from (35), we obtain that

$$0 \in \partial f(x^*) + \sum_{i=1}^{r} \nabla\varphi(|d_i^T x^*|^p) + N_{\mathcal{X}}(x^*).$$

Many application models can be formatted as (6), where the components of variable x are separated in the non-Lipschitz items of the objective function. When the SNN is used to solve (6), similar to the analysis in Theorem 3.2, we have the following corollary.

Corollary 3.1: If the SNN is used to solve (6), for any initial point $x_0 \in \mathcal{X}$, any accumulation point x^* of the solutions of the SNN satisfies

$$\begin{cases} x^* \in \mathcal{X} \\ 0 \in X^*\partial f(x^*) + p\nabla\varphi(|x^*|^p)|x^*|^p + X^* N_{\mathcal{X}}(x^*) \end{cases} \tag{37}$$

where $X^* = \text{diag}(x_1^*, \ldots, x_n^*)$ and

$$\nabla\varphi(|x^*|^p)|x^*|^p = (\nabla\varphi_1(|x_1^*|^p)|x_1^*|^p, \ldots, \nabla\varphi_n(|x_n^*|^p)|x_n^*|^p)^T.$$

We should notice that the inclusion property (37) in Corollary 3.1 is stronger than the conditions in Definition 2.2 for the SNN to solve (6). Moreover, the property (37) is a generalization of the first order necessary condition and reduces to the Clarke stationary point when $p \geq 1$ [32].

IV. Numerical Experiments

In this section, we use the following notations to report our numerical experiments.

1) SNN: use codes for ODE in MATLAB to implement the proposed neural network SNN in this paper.
2) x^*: numerical solution obtained by the corresponding algorithms.
3) N^*: number of nonzero elements of x^*.
4) PSNR(dB) and SNR(dB): signal-to-noise ratios defined by

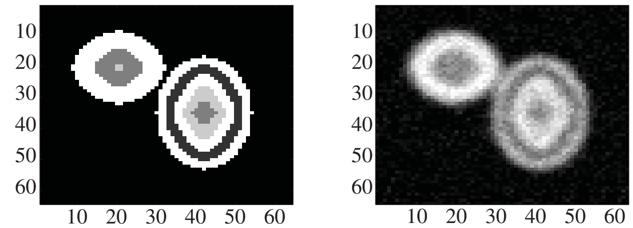
$$\text{PSNR} = -10\lg\frac{\|x^* - s\|^2}{n}, \quad \text{SNR} = -20\lg\frac{\|x^* - s\|}{\|s\|}$$

where s is the original signal and n is the dimension of the original signal.
5) $RA(\%)$: recovery rate, where we say the recovery of signal s is effective if its SNR is larger than 40.
6) Oracle and Radio: the oracle estimator is defined by Oracle $= \sigma^2\text{tr}(A_\Lambda^T A_\Lambda)^{-1}$, where $\Lambda = \text{support}(s)$. And we define Radio $= \|x^* - s\|/\text{Oracle}$.

A. Image Restoration

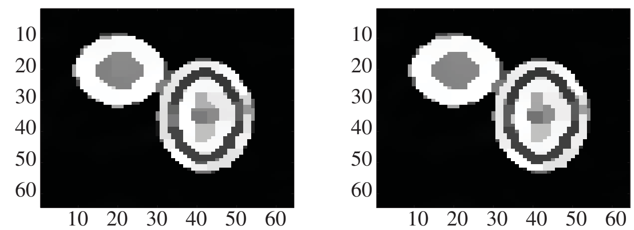
Image restoration is to resconstruct an image of an unknown scene from an observed image, which has wide applications in engineering and sciences. In this experiment, we perform the SNN on the restoration of 64×64 circle image. The observed image is distorted from the unknown true image mainly by two factors: the blurring and the random noise. The blurring is a 2-D Gaussian function

$$h(i, j) = e^{-2\left(\frac{i}{3}\right)^2 - 2\left(\frac{j}{3}\right)^2}$$



(a) (b)

Fig. 2. (a) Original images. (b) Observed images.



(a) (b)

Fig. 3. Restored images with (a) $p = 0.5$ and (b) $p = 0.3$.

which is truncated such that the function has a support of 7×7. A Gaussian noise with zero mean and standard derivation of 0.05 is added to the blurred image. Fig. 2 presents the original and the observed images that is used in [21] and [22]. The peak signal-to-noise ratio (PSNR) of the observed image is 15.50 dB.

Then, we consider the following minimization problem:

$$\begin{aligned} \min \quad & \|Ax - b\|^2 + 0.02\sum_{i=1}^{r}\psi(|d_i^T x|^p) \\ \text{s.t.} \quad & x \in \mathcal{X} = \{x : x_i \in [0, 1], \ i = 1, 2, \cdots, n\} \end{aligned} \tag{38}$$

where $n = 64 \times 64$, $A \in \mathbb{R}^{n \times n}$ is the blurring matrix, $b \in \mathbb{R}^n$ is the observed image, $p \in [0, 1)$ is a positive parameter, $\psi : [0, +\infty) \to [0, +\infty)$ is the potential function defined by

$$\psi(z) = \frac{0.5z}{1 + 0.5z}.$$

$d_i^T \in \mathbb{R}^n$ is the ith row of the first-order difference matrix D, which is used to define the difference between each pixel and its four adjacent neighbors [21], [22].

Let $\mu(t) = e^{-0.1t}$ and $x_0 = P_{\mathcal{X}}(b)$. In this experiment, we use ode15s in MATLAB to implement the SNN and we stop when the computation iteration exceeds 1500. The restored images with $p = 0.5$ and $p = 0.3$ are presented in Fig. 3. The ultimate values of $\mu(t)$ for the two cases are about 0.01 and 0.007. Fig. 4 shows the PSNRs along the solutions of the SNN when $p = 0.5$ and $p = 0.3$. From this figure, we see that the PSNRs are strictly increasing along the solutions of the SNN. Table I presents the objective values and PSNRs for the original image, observed image, and restored images with $p = 0.5$ and $p = 0.3$. The PSNRs of the graduated nonconvexity GNC algorithm in [22] and the smoothing projected gradient SPG algorithm in [21] are 19.42 dB and 19.97 dB, respectively. From Table I, we see that PSNRs obtained by the SNN are

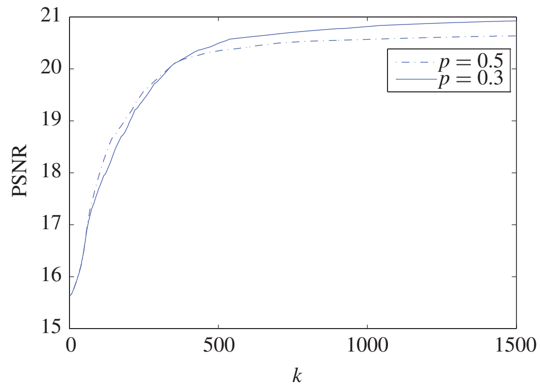


Fig. 4. PSNRs with $p = 0.5$ and $p = 0.3$.

TABLE I
OBJECTIVE VALUES AND PSNRS IN IMAGE RESTORATION

	Original	Observed	$p = 0.5$	$p = 0.3$
Objective value	14.486	39.461	16.784	16.784
PSNR(dB)		15.503	20.686	20.924

higher than those obtained in [21] and [22]. Moreover, using $p = 0.3$ is better than using $p = 0.5$ for this problem.

B. Underdetermined Blind Source Separation (BSS)

BSS describes the process of extracting some underlying original source signals from a number of observed mixture signals, where the mixing model is either unknown or the knowledge about the mixing process is limited. The problem of BSS has been extensively studied over the past decades due to its potential applications in a wide range of areas such as digital communications, speech enhancement, medical data analysis, and remote sensing [25], [29], [46].

In the anechoic instantaneous mixing model, at each time l, one has N sources $x(l)$, K mixtures $b(l)$, and K noises $n(l)$ such that

$$b(l) = Ax(l) + n(l)l = 1, \dots, L$$

where $A = (a_{ij})_{K \times N}$ with $\sum_{k=1}^{K} |a_{kj}|^2 = 1$ for $j = 1, 2, \dots, N$.

In this experiment, we focus on the reformulation of sparse source signal s when the mixing matrix A has been fixed and the sparsity of s is unknown. To evaluate the efficiency of algorithms in recovering the original signals, we use the signal-to-noise ratio. The larger the SNR the algorithm has, the better.

We use the following MATLAB code to generate real-valued original signals $S \in \mathbb{R}^{N \times L}$ with length L. The lth column of S is the signal at time l, which is with dimension N and sparsity T.

```
S = zeros(N, L);
for m = 1:L
q = randperm(N);
S[q(1:T),m] = [sqrt(4)*randn(T,1)];
end.
```

And we use the following MATLAB code to generate mixing matrix A and mixture signals $b \in \mathbb{R}^{K \times L}$.

TABLE II
SIGNAL RESTORATION WITHOUT NOISE WITH
$N = 8$, $T = 2$, $K = 4, 5, 6, 7$, AND $L = 100$

p	K	4	5	6	7
0.1	$RA(\%)$	65	86	97	100
	Mean-SNR(dB)	13.52	19.57	32.43	83.22
0.3	$RA(\%)$	68	87	95	100
	Mean-SNR(dB)	12.92	20.24	31.19	77.46
0.5	$RA(\%)$	62	86	93	100
	Mean-SNR(dB)	12.03	19.54	31.06	71.06
0.7	$RA(\%)$	60	82	92	100
	Mean-SNR(dB)	12.02	18.97	30.78	63.39
1	$RA(\%)$	48	65	84	92
	Mean-SNR(dB)	11.47	17.86	28.73	50.17

```
A = randn(K,N);
for j = 1:N
A(:, j) = A(:, j)/norm[A(:, j)];
end
noise = σ*randn(K, L);
b = A*S-noise;
```

1) Signal Without Noise: In this part, we consider the signal without noise and wish to restore the sparsest signal. In many papers [25], [29], [30], this class of BSS problems is formally stated as

$$\min \sum_{i=1}^{n} |x_i(l)|^p, \quad \text{s.t. } Ax(l) = b(l) \tag{39}$$

$l = 1, 2, \dots, L$, where $0 < p \leq 1$. For each $l \in \{1, 2, \dots, L\}$

$$\mathcal{X} = \{x \in \mathbb{R}^n : Ax = b(l)\}$$
$$P_{\mathcal{X}}(x) = x - A^T(AA^T)^{-1}(Ax - b(l)).$$

We define Mean-SNR as the mean SNR of the L tests. Choose initial point $x_0 = A^T(AA^T)^{-1}b$ and $\mu(t) = 4e^{-0.1t}$. Use ode15s in MATLAB to implement the SNN and stop when $\mu(t) \leq 0.005$. Table II presents the numerical results of using the SNN to solve (39) with $N = 8$, $T = 2$, $L = 100$, and $K = 4, 5, 6, 7$. Saab, Yilmaz, Mckeown, and Abugharbieh [25] pointed out that the choice of $0.1 \leq p \leq 0.4$ is appropriate in the case of speech. From Table II, we see that the numerical results using $p \leq 0.5$ is better than that using $p > 0.5$ in this experiment.

Next, we test the SNN by a problem with $N = 512$, $T = 130$ and $p = 0.5$. The SNRs of the SNN for solving (39) with $K = 225$, 250, 280, 300, and 330 are 10.20 dB, 55.39 dB, 59.17 dB, 60.00 dB, and 63.10 dB, respectively.

2) Signal With Noise: In this part, we consider the signal restoration with noise, where noise is the independent identically distributed Gaussian noise with zero mean and variance σ^2. We evaluate the recovery of signals with noise by the SNR and "Radio." The "Radio" is closer to 1, the algorithm is more robust.

In this experiment, we let $N = 512$, $T = 130$, $L = 10$, $K = 270$, and $\sigma = 0.1$ in the MATLAB code to generate the mixing signals with noises. To solve the signal recovery with

TABLE III
SNRs AND RADIOS OF THE SNN WITH
$N = 512$, $L = 10$, AND $p = 0.5$

K	270	280	290	300	310	320
Max-SNR(dB)	18.29	18.64	19.06	19.51	20.33	20.71
Min-SNR(dB)	11.57	15.42	16.23	16.74	17.42	19.07
Mean-SNR(dB)	15.43	17.77	17.80	17.96	18.68	19.61
Oracle	2.49	2.40	2.32	2.73	2.25	2.22
Max-Radio	2.39	1.46	1.33	1.32	1.30	1.18
Min-Radon	1.11	1.16	1.07	1.05	1.05	1.00
Mean-Radio	1.61	1.27	1.26	1.25	1.19	1.07

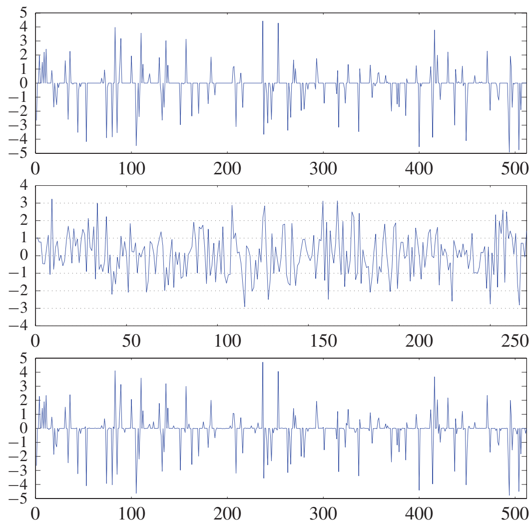


Fig. 5. Original (first), mixture (second), and recovered (third) signal with $N = 512$, $T = 130$, $K = 270$, and $p = 0.5$.

noise, we use the unconstrained l_2-l_p model as follows:

$$\min \|Ax(l) - b(l)\|^2 + \lambda \sum_{i=1}^{n} |x_i(l)|^p, \; l = 1, \ldots, L. \quad (40)$$

Let $N = 512$, $T = 130$, and $L = 10$. For different choices of K, Table III presents the numerical results of the SNN with $p = 0.5$, $\lambda = 0.05$ and random initial points to solve (40), where Max-SNR(Max-Radio), Min-SNR(Min-Radio), Mean-SNR(Mean-Radio) are the maximal, minimal, and mean SNR(Radio) of randomly generated ten tests. From Table III, we can see that the "Radio" is always close to 1 using the SNN to solve the underdetermined blind source separation with noise. Fig. 5 shows the original signal, mixing signal, and recovered signal of a random test with $N = 512$, $K = 270$, and $p = 0.5$.

C. Variable Selection

In this section, we report numerical results for testing a prostate cancer problem. The data set is downloaded from the UCI Standard database [47]. The data set consists of the medical records of 97 patients who were about to receive a radical prostatectomy, which is divided into two parts: a training set with 67 observations and a test set with 30 observations. The predictors are eight clinical measures: lcavol, lweight, age,

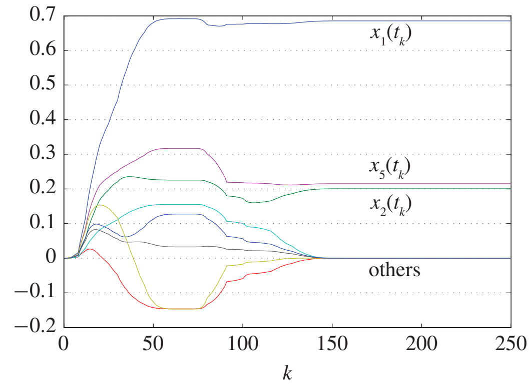


Fig. 6. Convergence of solution $x(t_k)$ of the SNN with ψ_1 and $\lambda = 4.5$.

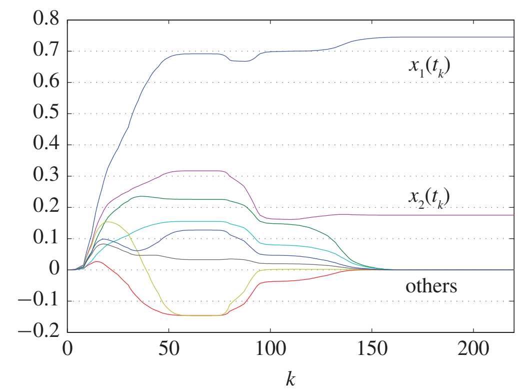


Fig. 7. Convergence of solution $x(t_k)$ of the SNN with ψ_1 and $\lambda = 5.5$.

lbph, svi, lcp, pleason, and pgg45. In this experiment, we want to find fewer main factors with smaller prediction error. Thus, we solve this problem by the following unconstrained minimization problem:

$$\min \quad \lg(\|Ax - b\|^2 + 1) + \lambda \sum_{i=1}^{8} \psi(|x_i|^p) \quad (41)$$

where $A \in \mathbb{R}^{97\times 8}$, $p = 0.5$, ψ is defined by one of the following three formats:

$$\psi_1(z) = z, \quad \psi_2(z) = \frac{3z}{1+3z}, \quad \psi_3(z) = \lg(3z+1).$$

(41) is an unconstrained non-Lipschitz optimization, where the first item in the objective function of (41) is smooth, but nonconvex. Every element of x is a clinical measure of the predictors. We call a clinical measure is a main factor if the relevant element of obtain optimal solution x^* is nonzero. The prediction error is defined by the mean square error over the 30 observations in the test set. In [32], the authors use l_2-l_p model and OMP-SCG method to solve this problem with $A \in \mathbb{R}^{67\times 8}$. In their numerical results, the authors can obtain three main factors with prediction error 0.443.

Choose $x_0 = (0, \ldots, 0)^T$ and $\mu(t) = e^{-0.1t}$. In this experiment, we use ode15s in MATLAB to implement the SNN and stop when $\mu(t) \leq 10^{-6}$. The numerical results using the SNN to solve (41) are listed in Table IV, "Error" is the prediction error of x^*. Table IV shows that we can find

TABLE IV
VARIABLE SELECTION WITH DIFFERENT VALUES OF λ

	ψ_1			ψ_2			ψ_3		
λ	4.5	5.5	6	5	10.3	10.5	2.5	4.1	4.2
x_1^*(lcavol)	0.655	0.769	0.794	0.671	0.799	0.829	0.685	0.745	0.840
x_2^*(lweight)	0.208	0.130	1.58e-10	0.209	0.159	4.15e-9	0.201	3.00e-7	1.49e-8
x_3^*(age)	1.24e-7	3.81e-10	1.28e-9	–3.28e-8	5.12e-9	1.96e-10	–4.14e-12	5.53e-8	1.71e-8
x_4^*(lbph)	6.89e-7	2.84e-9	1.14e-9	2.30e-8	1.85e-8	3.42e-9	–6.58e-12	1.73e-7	5.87e-10
x_5^*(svi)	0.232	4.93e-9	1.62e-10	0.228	3.42e-8	3.71e-9	0.215	0.175	1.98e-8
x_6^*(lcp)	7.42e-13	2.20e-9	8.04e-8	6.45e-8	1.20e-8	1.45e-9	–3.18e-13	–3.31e-8	2.06e-8
x_7^*(pleason)	3.67e-12	2.04e-9	6.87e-8	4.81e-9	1.24e-8	1.34e-9	–3.37e-12	2.62e-8	1.26e-8
x_8^*(pgg45)	1.71e-8	3.22e-9	–1.39e-10	1.46e-7	2.08e-8	2.48e-9	–4.46e-12	6.17e-8	1.95e-8
N^*	3	2	1	3	2	1	3	2	1
Error	0.394	0.399	0.497	0.396	0.479	0.483	0.394	0.478	0.480

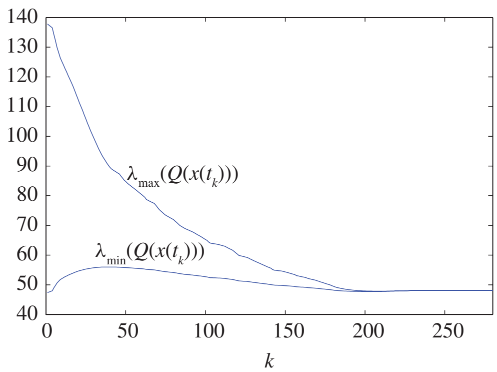


Fig. 8. Convergence of $\lambda_{\max}(Q(x(t_k)))$ and $\lambda_{\min}(Q(x(t_k)))$ along the solution of the SNN with a random initial point in $\mathcal{X}$.

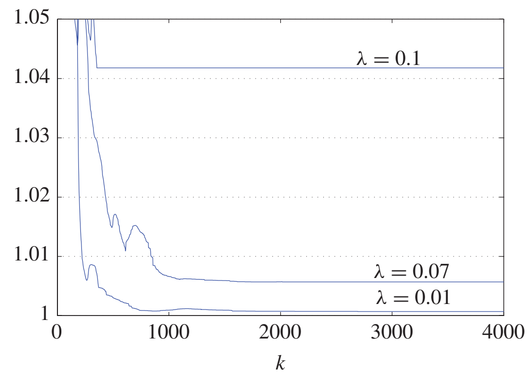


Fig. 9. Optimal values with three different values of λ.

fewer main factors with smaller prediction errors by using the SNN to solve optimization model (41). Figs. 6 and 7 show the convergence of the solutions of the SNN using ψ_1 with $\lambda = 4.5$ and $\lambda = 5.5$.

D. Optimizing Condition Number

Optimizing eigenvalue functions has been studied for decades. The condition number of a positive definite matrix $Q \in \mathbb{R}^{m\times m}$ is defined by

$$\kappa(Q) = \frac{\lambda_{\max}(Q)}{\lambda_{\min}(Q)}$$

where $\lambda_{\max}(Q)$ and $\lambda_{\min}(Q)$ are the maximal and minimal eigenvalues of Q. The quantity $\kappa(Q)$ has been widely used in the sensitivity analysis of interpolation and approximation. The problem of minimizing condition number is an important class of nonsmooth nonconvex optimization problems [48], [49]. In this example, we consider the test problem

$$\begin{aligned} \min \quad & \kappa(Q(x)) = \kappa\left(I + \sum_{i=1}^{n} x_i Q_i\right) \\ \text{s.t.} \quad & x \in \mathcal{X} = \{x : 0 \le x_i \le 1,\ i = 1, 2, \ldots, n\} \end{aligned} \tag{42}$$

where $n = 100$, I is the 5×5 identity matrix, $Q_i \in \mathbb{R}^{5\times 5}$ is a symmetric positive definite matrix generated by MATLAB randomly, $i = 1, 2, \ldots, n$.

We know that the optimal value of this problem is 1.

Denote $\lambda_1(Q), \ldots, \lambda_5(Q)$ the non-decreasing ordered eigenvalues of Q. We use the smoothing function of the objective function given in [49], specially

$$\tilde{f}(x, \mu) = -\frac{\ln\left(\sum_{i=1}^{5} e^{\lambda_i(Q(x))/\mu}\right)}{\ln\left(\sum_{i=1}^{5} e^{-\lambda_i(Q(x))/\mu}\right)}.$$

Then, the SNN can be rewritten as

$$\dot{x}(t) = -x(t) + P_{\mathcal{X}}[x(t) - \nabla_x \tilde{f}(x(t), \mu(t))]. \tag{43}$$

With a random initial point in $\mathcal{X}$, Fig. 8 presents the convergence of $\lambda_{\max}(Q(x(t_k)))$ and $\lambda_{\min}(Q(x(t_k)))$ along the solution of (43).

We know that adding a non-Lipschitz item into the objective function can often bring some influences to the original optimization problems, such as the sparsity. In the following, we consider what influences may occur in this kind of problems. Then, we consider the revised optimization model

$$\begin{aligned} \min \quad & \kappa(Q(x)) = \kappa\left(I + \sum_{i=1}^{n} x_i Q_i\right) + \lambda \sum_{i=1}^{n} |x_i|^p \\ \text{s.t.} \quad & x \in \mathcal{X} = \{x : 0 \le x_i \le 1,\ i = 1, 2, \ldots, n\}. \end{aligned} \tag{44}$$

The objective function of (44) is the combining of a nonsmooth, nonconvex function, and a non-Lipschitz function.

TABLE V
CONDITION NUMBERS FOR DIFFERENT VALUES OF λ

λ	$\lambda_{\max}(Q(x^*))$	$\lambda_{\min}(Q(x^*))$	$\kappa(Q(x^*))$	N^*
0	51.6629	51.6299	1.0004	72
0.01	40.8222	40.7932	1.0007	48
0.03	21.021	21.0411	1.0010	31
0.05	11.8538	11.8712	1.0015	21
0.07	8.6105	8.6596	1.0057	17
0.1	7.0827	7.3791	1.0418	10

We use the SNN modeled by

$$\dot{x}(t) = -x(t) + P_{\mathcal{X}}\Big[x(t) - \nabla_x \tilde{f}(x(t), \mu(t)) \\ -\lambda \nabla_x \theta^p(x(t), \mu(t))\Big] \quad (45)$$

where θ is defined as in (11), λ is a positive parameter.

Table V shows the numerical results for different values of λ in (45). From the results in Table V, we see that as λ is increasing, the sparsity of the optimal solution is increasing and the eigenvalues are decreasing. Thus, if we want to get a more sparse solution, we can add a non-Lipschitz item to the original problem. Fig. 9 presents the convergence of the optimal values with three different values of λ in Table V.

V. CONCLUSION

In this paper, we studied a class of constrained non-Lipschitz optimization problems, which has wide applications in engineering, sciences, and economics. To avoid solving differential inclusions, by applying the smoothing techniques, we proposed the SNN modeled by a differential equation to solve this kind of problem. The definition of a stationary point of the constrained non-Lipschitz optimization was introduced. Under a mild condition, we proved the global existence, boundedness, and limit convergence of the proposed SNN, which shows that any accumulation point of the solutions of the SNN is a stationary point of (3). Specifically, when the SNN was used to solve the special case (6), any accumulation point of the solutions of SNN satisfies a stronger inclusion property which reduces to the definition of the Clarke stationary points for Lipschitz problems with $p \geq 1$. Some numerical experiments and comparisons on image restoration, signal recovery, variable selection, and optimizing condition number were illustrated to show the efficiency and advantages of the SNN.

APPENDIX A
PROOF OF PROPOSITION 3.1

Fix $i \in \{1, 2, \ldots, r\}$.

1) From (16), we can get

$$|\nabla_\mu \theta(d_i^T x, \mu)| \leq \begin{cases} 0, & \text{if } |d_i^T x| > \mu \\ \frac{1}{2}, & \text{if } |d_i^T x| \leq \mu. \end{cases}$$

Thus, from (11), (15), and (17), we obtain

$$|\nabla_\mu \varphi(\theta^p(d_i^T x, \mu))| \leq p l_\varphi \mu^{p-1}. \quad (46)$$

2) When $|d_i^T x| > \mu$, $\theta^p(d_i^T x, \mu) - |d_i^T x|^p = 0$. Then, we only need to consider the case that $|d_i^T x| \leq \mu$.
In [32], Chen, Xu, and Ye have proved that

$$\nabla_s^2 \theta^p(s, \mu) > 0 \quad \forall s \in (-\mu, \mu) \quad (47)$$

which means that $\theta^p(s, \mu)$ is strongly convex in s on $[-\mu, \mu]$.
For any fixed μ, since $\theta^p(s, \mu)$ and $|s|^p$ are symmetrical on $[-\mu, \mu]$ and $\theta^p(s, \mu)$ is strongly convex in s on $[-\mu, \mu]$, then

$$|\theta^p(d_i^T x, \mu) - |d_i^T x|^p| \leq \theta^p(0, \mu) \leq \left(\frac{\mu}{2}\right)^p \\ \forall x \in \mathbb{R}^n \; \mu > 0.$$

Since φ is globally Lipschitz, we obtain 2).

3) First, we prove the Lipschitz property of $\nabla_s \theta^p(s, \mu)$ in s for any fixed $\mu > 0$, which can be derived by the global boundedness of the Clarke generalized gradient of $\nabla_s \theta^p(s, \mu)$. $\nabla_s \theta^p(s, \mu)$ is not differentiable only when $|s| = \mu$. From (7), we only need to prove the global boundedness of the gradient of $\nabla_s \theta^p(s, \mu)$ when $|s| \neq \mu$.
Fix $\mu > 0$. When $|s| > \mu$, we obtain

$$|\nabla_s^2 \theta^p(s, \mu)| = |p(p-1)|s|^{p-2}| \leq p\mu^{p-2}.$$

When $|s| < \mu$, we have

$$\nabla_s^2 \theta^p(s, \mu) \leq p\left(\frac{|s|^2}{2\mu} + \frac{\mu}{2}\right)^{p-1}\left(\frac{1}{\mu}\right) \leq 2p\mu^{p-2}. \quad (48)$$

From (47), then

$$|\nabla_s^2 \theta^p(s, \mu)| \leq 2p\mu^{p-2} \quad \forall s \in (-\mu, \mu). \quad (49)$$

Combining (7), (48), and (49), we have

$$|\xi| \leq 2p\mu^{p-2} \quad \forall s \in \mathbb{R}, \quad \xi \in \partial_s(\nabla_s \theta^p(s, \mu))$$

which implies that

$$|\nabla_s \theta^p(s, \mu) - \nabla_r \theta^p(r, \mu)| \leq 2p\mu^{p-2}|s - r|, \; \forall s, r \in \mathbb{R}.$$

Therefore, for any $x, z \in \mathbb{R}^n$, we get

$$\|\nabla_x \theta^p(d_i^T x, \mu) - \nabla_z \theta^p(d_i^T z, \mu)\| \\ \leq 2p\mu^{p-2}\|d_i d_i^T\| \cdot \|x - z\|. \quad (50)$$

4) From (13), we have

$$\|\nabla_x \theta^p(d_i^T x, \mu)\| \leq 2p\|d_i\|\mu^{p-1} \quad (51)$$

which follows that $\theta^p(d_i^T x, \mu)$ is globally Lipschitz in x for any fixed $\mu > 0$. If $\nabla\varphi$ is locally (globally) Lipschitz, then $\nabla\varphi(\theta^p(d_i^T x, \mu))$ is locally (globally) Lipschitz.

Combining the Lipschitz property in x and the uniform boundedness of $\nabla\varphi(\theta^p(d_i^T x, \mu))$ and $\nabla_x \theta^p(d_i^T x, \mu)$, from (14), we get the local (global) Lipschitz property of $\nabla_x \varphi(\theta^p(d_i^T x, \mu))$ in x for any fixed μ.

APPENDIX B
PROOF OF PROPOSITION 3.2

Denote $x, \hat{x} \in AC[0, \infty)$ two solutions of the SNN with initial point x_0. From Theorem 3.1, there is a $\rho > 0$ such that $\|x_t\| \leq \rho$ and $\|\hat{x}_t\| \leq \rho$, $\forall t \geq 0$. Suppose there exists a $\hat{t} > 0$ such that $x_{\hat{t}} \neq \hat{x}_{\hat{t}}$. Then there exists a $\delta > 0$ such that $x_t \neq \hat{x}_t$, $\forall t \in [\hat{t}, \hat{t} + \delta]$.

Denote

$$k(x, \mu) = -x + P_{\mathcal{X}} \left[x - \nabla_x \tilde{f}(x, \mu) - \sum_{i=1}^{r} \nabla_x \varphi(\theta^p(d_i^T x, \mu)) \right].$$

Form Propositions 2.2 and 3.1, when $\nabla\varphi(\cdot)$ and $\nabla_x \tilde{f}(\cdot, \mu)$ is locally Lipschitz, $k(\cdot, \mu)$ are locally Lipschitz for any fixed $\mu > 0$, which means that there exists an $L_\mu > 0$ such that

$$\|k(y, \mu) - k(\hat{y}, \mu)\| \leq L_\mu \|y - \hat{y}\| \quad \forall y, \hat{y} \in \{u \in \mathbb{R}^n : \|u\| \leq \rho\}. \tag{52}$$

Since x_t, $\hat{x}_t$, and μ_t are continuous and uniformly bounded on $[0, \hat{t} + \delta]$, from (52), there is an $L > 0$ such that

$$\|k(x_t, \mu_t) - k(\hat{x}_t, \mu_t)\| \leq L\|x_t - \hat{x}_t\| \quad \forall t \in [0, \hat{t} + \delta].$$

Differentiating $(1/2)\|x_t - \hat{x}_t\|^2$ along the two solutions of the SNN, we obtain

$$\frac{d}{dt}\frac{1}{2}\|x_t - \hat{x}_t\|^2 \leq L\|x_t - \hat{x}_t\|^2 \quad \forall t \in [0, \hat{t} + \delta].$$

Integrating the above inequality from 0 to $t(\leq \hat{t} + \delta)$ and applying Gronwall's inequality [41], it follows that $x_t = \hat{x}_t$, $\forall t \in [0, \hat{t} + \delta]$, which leads a contradiction.

ACKNOWLEDGMENT

The authors would like to thank the Editor-in-Chief, Associate Editor, and the reviewers for their insightful and constructive comments, which enriched the content and improved the presentation of the results in this paper.

REFERENCES

[1] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. New York: Wiley, 1993.

[2] N. Kalouptsidis, *Signal Processing Systems, Theory and Design*. New York: Wiley, 1997.

[3] Z. G. Hou, M. M. Gupta, P. N. Nikiforuk, M. Tan, and L. Cheng, "A recurrent neural network for hierarchical control of interconnected dynamic systems," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 466–481, Mar. 2007.

[4] Y. S. Xia, J. Wang, and L. M. Fok, "Grasping-force optimization for multifingered robotic hands using a recurrent neural network," *IEEE Trans. Robot. Autom.*, vol. 20, no. 3, pp. 549–554, Jun. 2004.

[5] X. B. Gao and L. Z. Liao, "A new projection-based neural network for constrained variational inequalities," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 373–388, Mar. 2009.

[6] Y. S. Xia, G. Feng, and J. Wang, "A novel recurrent neural network for solving nonlinear optimization problems with inequality constraints," *IEEE Trans. Neural Netw.*, vol. 19, no. 8, pp. 1340–1353, Aug. 2008.

[7] J. L. Liang, Z. D. Wang, and X. H. Liu, "State estimation for coupled uncertain stochastic networks with missing measurements and time-varying delays: The discrete-time case," *IEEE Trans. Neural Netw.*, vol. 20, no. 5, pp. 781–793, May 2009.

[8] B. Y. Zhang, S. Y. Xu, G. D. Zong, and Y. Zou, "Delay-dependent exponential stability for uncertain stochastic Hopfield neural networks with time-varying delays," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 6, pp. 1241–1247, Jun. 2009.

[9] D. W. Tank and J. J. Hopfield, "Simple 'neural' optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. Circuits Syst.*, vol. 33, no. 5, pp. 533–541, May 1986.

[10] M. P. Kennedy and L. O. Chua, "Neural networks for nonlinear programming," *IEEE Trans. Circuit Syst.*, vol. 35, no. 5, pp. 554–562, May 1988.

[11] M. Forti, P. Nistri, and M. Quincampoix, "Generalized neural network for nonsmooth nonlinear programming problems," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 9, pp. 1741–1754, Sep. 2004.

[12] X. P. Xue and W. Bian, "Subgradient-based neural networks for nonsmooth convex optimization problems," *IEEE Trans. Circuits Syst. I*, vol. 55, no. 8, pp. 2378–2391, Sep. 2008.

[13] G. Li, S. Song, and C. Wu, "Generalized gradient projection neural networks for nonsmooth optimization problems," *Sci. Chin. Inf. Sci.*, vol. 53, no. 5, pp. 990–1005, 2010.

[14] Q. S. Liu and J. Wang, "Finite-time convergent recurrent neural network with a hard-limiting activation function for constrained optimization with piecewise-linear objective functions," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 601–613, Apr. 2011.

[15] L. Cheng, Z. G. Hou, Y. Z. Lin, M. Tan, W. C. Zhang, and F. X. Wu, "Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 714–726, May 2011.

[16] M. Forti, P. Nistri, and M. Quincampoix, "Convergence of neural networks for programming problems via a nonsmooth Łojasiewicz inequality," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1471–1486, Nov. 2006.

[17] W. Bian and X. P. Xue, "Subgradient-based neural networks for nonsmooth nonconvex optimization problems," *IEEE Trans. Neural Netw.*, vol. 20, no. 6, pp. 1024–1038, Jun. 2009.

[18] W. L. Lu and J. Wang, "Convergence analysis of a class of nonsmooth gradient systems," *IEEE Trans. Circuits Syst. I*, vol. 55, no. 11, pp. 3514–3527, Dec. 2008.

[19] M. Forti, M. Grazzini, P. Nistri, and L. Pancioni, "Generalized Lyapunov approach for convergence of neural networks with discontinuous or non-Lipschitz activations," *Phys. D*, vol. 214, no. 1, pp. 88–99, Feb. 2006.

[20] M. Forti, "M-matrices and global convergence of discontinuous neural networks," *Int. J. Circuit Theory Appl.*, vol. 35, no. 2, pp. 105–130, Mar.–Apr. 2007.

[21] X. Chen, M. K. Ng, and C. Zhang. (2010). *Nonconvex ℓ_p-Regularization and Box Constrained Model for Image Restoration* [Online]. Available: http://www.polyu.edu.hk/ama/staff/xjchen/nonconvex10Sept2011.pdf

[22] M. Nikolova, M. K. Ng, S. Q. Zhang, and W. K. Ching, "Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 2–25, Mar. 2008.

[23] N. Meinshausen and B. Yu, "Lasso-type recovery of sparse representations for high-dimensional data," *Ann. Stat.*, vol. 37, no. 1, pp. 246–270, 2009.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[25] R. Saab, O. Yilmaz, M. J. Mckeown, and R. Abugharbieh, "Underdetermined anechoic blind source separation via ℓ^q-basis-pursuit with $q \ll 1$," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4004–4017, Aug. 2007.

[26] V. Saligrama and M. Q. Zhao, "Thresholded basis pursuit: LP algorithm for order-wise optimal support recovery for sparse and approximately sparse signals from noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1567–1586, Mar. 2011.

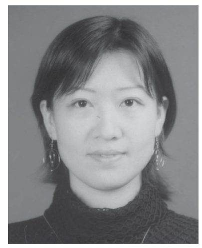
[27] D. L. Donoho, "Neighborly polytopes and sparse solutions of underdetermined linear equations," Dept. Stat., Standford University, Standford, CA, Tech. Rep., 2005.

[28] L. B. Montefusco, D. Lazzaro, and S. Papi, "Nonlinear filtering for sparse signal recovery from incomplete measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2494–2502, Jul. 2009.

[29] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[30] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[31] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Probl.*, vol. 24, no. 3, pp. 1–14, 2008.
[32] X. Chen, F. Xu, and Y. Ye, "Lower bound theory of nonzero entries in solutions of l_2-l_p minimization," *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2832–2852, 2010.
[33] X. Chen, D. Ge, Z. Wang, and Y. Ye. (2011). *Complexity of Unconstrained L_2-L_p Minimization* [Online]. Available: http://www.tju.edu.cn/science/xydt/tztg/201109/P020110908302116195306.pdf
[34] P. Huber, *Robust Estimation*. New York: Wiley, 1981.
[35] J. Q. Fan and H. Peng, "Nonconcave penalized likelihood with a diverging number of parameters," *Ann. Stat.*, vol. 32, no. 3, pp. 928–961, 2004.
[36] A. Auslender, "How to deal with the unbounded in optimization: Theory and algorithms," *Math. Program.*, vol. 79, nos. 1–3, pp. 3–18, 1997.
[37] W. Bian and X. Chen. (2011). *Smoothing Dynamical Approach to Nonsmooth, Nonconvex Constrained Minimization* [Online]. Available: http://www.polyu.edu.hk/ama/staff/xjchen/MPB2012.pdf
[38] J. Kreimer and R. Y. Rubinstein, "Nondifferentiable optimization via smooth approximation: General analytical approach," *Ann. Oper. Res.*, vol. 39, no. 1, pp. 97–119, 1992.
[39] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Berlin, Germany: Springer-Verlag, 1998.
[40] F. H. Clarke, *Optimization and Nonsmooth Analysis*. New York: Wiley, 1983.
[41] J.-P. Aubin and A. Cellina, *Differential Inclusion: Set-Valued Maps and Viability Theory*. New York: Springer-Verlag, 1984.
[42] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*. New York: SIAM, 1987.
[43] Z. H. Yuan, L. H. Huang, D. W. Hu, and B. W. Liu, "Convergence of nonautonomous Cohen–Grossberg-type neural networks with variable delays," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 140–147, Jan. 2008.
[44] B. Y. Zhang, S. Y. Xu, G. D. Zong, and Y. Zou, "Delay-dependent exponential stability for uncertain stochastic Hopfield neural networks with time-varying delays," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 6, pp. 1241–1247, Jun. 2009.
[45] D. Betounes, *Differential Equations: Theory and Applications*. New York: Springer-Verlag, 2009.
[46] Y. Xiang, S. K. Ng, and V. K. Nguyen, "Blind separation of mutually correlated sources using precoders," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 82–90, Jan. 2010.
[47] C. Blake and C. Merz. (1998). *Repository of Machine Learning Databases [DB/OL]*. Dept. Inf. Comput. Sci., Univ. California, Irvine, CA [Online]. Available: http://archive.ics.uci.edu/ml/
[48] M. L. Overton and R. S. Womersley, "Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices," *Math. Program.*, vol. 62, nos. 1–3, pp. 321–357, 1993.
[49] X. Chen, R. S. Womersley, and J. Ye, "Minimizing the condition number of a Gram matrix," *SIAM J. Optim.*, vol. 21, no. 1, pp. 127–148, 2011.



Wei Bian received the B.S. and Ph.D. degrees in mathematics from the Harbin Institute of Technology, Harbin, China, in 2004 and 2009, respectively.

She has been a Lecturer with the Department of Mathematics, Harbin Institute of Technology since 2009. Currently, she is a Post-Doctoral Fellow with the Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong. Her current research interests include neural network theory and optimization methods.

Xiaojun Chen was a Professor with the Department of Mathematical Sciences, Hirosaki University, Hirosaki, Japan. She is a Professor with the Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong. Her current research interests include nonsmooth, nonconvex optimization, stochastic variational inequalities, and approximations on the sphere.