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

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2_2 + \lambda \|x\|^p_p$$

(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 2009068. 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: bianwell@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

2162–237X/S31.00 © 2012 IEEE
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 $l_p$ norm ($p < 1$) 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):

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

(2)

where $A_i$ is the $i$th row of $A$, $\omega_k$ is a non-Lipschitz function, and $\psi$ 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:

$$\min_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{r} \varphi_i(|d_i^T x|^p)$$

(3)

s.t. $x \in \mathcal{X}$

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 $L_\varphi$.

Some popular formats of $\varphi$ satisfying the above conditions can be given by

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

(4)

where $\lambda$ and $\alpha$ 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:

$$\min_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{r} \varphi_i(|d_i^T x|^p)$$

(5)

s.t. $x \in \mathcal{X}$

that is, these functions $\varphi_i$ are not necessarily same.

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

$$\min_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{n} \varphi_i(|x_i|^p)$$

(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 \( x = (x_1, x_2, \ldots, x_n)^T \) and \( y = (y_1, y_2, \ldots, y_n)^T \), \( (x, y) = x^T y = \sum_{i=1}^{n} x_i y_i \) is the scalar product of \( x \) and \( y \). \( x_i \) denotes the \( i \)th element of \( x \). \( \| x \| \) denotes the Euclidean norm defined by \( \| x \| = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \). For a matrix \( A \in \mathbb{R}^{m \times n} \), 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 \( \Omega \subseteq \mathbb{R}^n \), \( \text{int}(\Omega) \) and \( \text{cl}(\Omega) \) mean the interior and closure of \( \Omega \) in \( \mathbb{R}^n \), respectively. For a subset \( U \subseteq \mathbb{R}^n \), \( \partial U \) indicates the convex hull of \( U \). \( P_{\Omega} : \mathbb{R}^n \rightarrow \Omega \) is the projection operator from \( \mathbb{R}^n \) onto \( \Omega \) and \( \partial P_{\Omega}(x) \) means the normal cone of \( \Omega \) at \( x \). \( AC[0, \infty) \) represents the set of all absolutely continuous functions \( x : [0, \infty) \rightarrow \mathbb{R}^n \).

## 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 \( h : \mathbb{R}^n \rightarrow \mathbb{R} \) 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

\[
\partial h(x) = \text{co} \left\{ \lim_{x_i \rightarrow x : x_i \in D_h} \nabla h(x_i) \right\}.
\]

**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) \( \partial h \) is upper semicontinuous at \( x \).

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

From [40, Proposition 2.4.2], we know that

\[
N_{\Omega}(x) = \text{cl} \left( \bigcup_{v \geq 0} v \partial d_{\Omega}(x) \right)
\]

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

The projection operator to \( \Omega \) 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]):**

\[
\begin{align*}
\| u - P_{\Omega}(v) - P_{\Omega}(v) - u \| & \geq 0 \quad \forall v \in \mathbb{R}^n, \quad u \in \Omega; \\
\| P_{\Omega}(u) - P_{\Omega}(v) \| & \leq \| u - v \| \quad \forall u, v \in \mathbb{R}^n.
\end{align*}
\]

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. As follows.

**Definition 2.1:** Let \( h : \mathbb{R}^n \rightarrow \mathbb{R} \) be a locally Lipschitz function. We call \( \tilde{h} : (0, +\infty) \rightarrow \mathbb{R} \) 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 \( \kappa_h > 0 \) such that

\[
|\nabla \mu \tilde{h}(x, \mu) - \kappa_h| \leq \kappa_h \
\forall \mu \in [0, \infty), \quad x \in \mathbb{R}^n.
\]

4) \( \lim_{\mu \downarrow 0} \nabla \tilde{h}(x, \mu) \leq \nabla \tilde{h}(x) \).

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

\[
\lim_{z \rightarrow x, \mu \downarrow 0} \tilde{h}(z, \mu) = h(x)
\]

and

\[
\tilde{h}(x, \mu) - h(x) \leq \kappa_h \mu \quad \forall \mu \in [0, \infty), \quad x \in \mathbb{R}^n.
\]

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 \( \varphi(|d_i^T x|^p) \), \( i = 1, 2, \ldots, r \). In this paper, we use the smooth functions \( \varphi(\theta^p(d_i^T x, \mu)) \) to approximate \( \varphi(|d_i^T x|^p) \), \( i = 1, 2, \ldots, r \), where \( \theta(s, \mu) \) is a smoothing function of \( |s| \) defined by

\[
\theta(s, \mu) = \begin{cases} 
|s|, & \text{if } |s| > \mu \\
\frac{s^2}{2\mu} + \mu, & \text{if } |s| \leq \mu.
\end{cases}
\]

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

\[
\lim_{\mu \downarrow 0} \varphi(\theta^p(d_i^T x, \mu)) = \varphi(|d_i^T x|^p), \quad i = 1, 2, \ldots, r.
\]

For any fixed \( \mu > 0 \), \( i = 1, 2, \ldots, r \), we obtain

\[
\nabla_s \varphi(\theta^p(d_i^T x, \mu)) = \varphi(\theta^p(d_i^T x, \mu)) \cdot \nabla_x \theta^p(d_i^T x, \mu)
\]

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

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

\[
\nabla_\mu \theta^p(d_i^T x, \mu) = p \theta^p(d_i^T x, \mu) \nabla_\mu \theta(d_i^T x, \mu)
\]

where

\[
\nabla_\mu \theta(d_i^T x, \mu) = \begin{cases} 
0, & \text{if } |d_i^T x| > \mu \\
-\frac{|d_i^T x|^2}{2\mu^2} + \frac{1}{2}, & \text{if } |d_i^T x| \leq \mu.
\end{cases}
\]
Thus
\[ \nabla_{\mu} \varphi(\theta^p(d_1^T x, \mu)) = \nabla \varphi(\theta^p(d_1^T x, \mu)) \cdot \nabla_{\mu} \theta^p(d_1^T x, \mu) \quad (17) \]
which means that \( \varphi(\theta^p(d_1^T x, \mu)) \) is continuously differentiable in \( \mu \) for any fixed \( x \in \mathbb{R}^n \), \( i \in 1, 2, \ldots, r \).

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

**III. MAIN RESULTS**

In this paper, we propose a SNN modeled by the following differential equation:
\[
\begin{cases}
\dot{x}(t) = -x(t) + P_{\mathcal{X}} [x(t) - \nabla_x \tilde{f}(x(t), \mu(t))] \\
- \sum_{i=1}^r \nabla_x \varphi(\theta^p(d_i^T x(t), \mu(t))) \\
x(0) = x_0
\end{cases}
\]
where \( \tilde{f} : \mathbb{R}^n \times [0, \infty) \rightarrow \mathbb{R} \) is a smoothing function of \( f \), \( \mu : [0, \infty) \rightarrow [0, \infty) \) is a continuously differentiable and decreasing function satisfying \( \lim_{t \to \infty} \mu(t) = 0 \).

**Remark 3.2:** The gradients of the three functions \( \varphi \) 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 \( \mu(t) \) 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
\[
\{ x \in \mathcal{X} : f(x) + \sum_{i=1}^r \varphi(|d_i^T x|^p) \leq \gamma \}
\]
is bounded for any \( \gamma > 0 \) then, for any initial point \( x_0 \in \mathcal{X} \), the SNN has a solution \( x : [0, \infty) \rightarrow \mathbb{R}^n \) 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 \( \sup_{t \in [0, \infty)} \| x(t) \| \leq \rho \).

**Proof:** Since the right-hand function of the SNN is continuous in \( x \) and \( t \), there are \( T > 0 \) and an absolutely continuous function \( x : [0, T) \rightarrow \mathbb{R}^n \) 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 \( (e^s/e^t - 1) > 0, h(s) \in \mathcal{X}, \forall 0 \leq s \leq t, \int_0^t (e^s/e^t - 1) ds = 1 \) 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 
\[
\frac{d}{dt} \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_t, \mu_t) \right) \nonumber \]
\[ = \left\langle \nabla \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \nabla \varphi(\theta^p(d_i^T x_t, \mu_t)), x_t \rightangle 
+ \left\langle \nabla_{\mu} \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \nabla_{\mu} \varphi(\theta^p(d_i^T x_t, \mu_t)) \right\rangle \dot{\mu}_t. \quad (19) \]

Let $v = x_t - \nabla \tilde{f}(x_t, \mu_t) - \sum_{i=1}^{r} \nabla \varphi(\theta^p(d_i^T x_t, \mu_t))$ and $u = x_t$ in Proposition 2.2, from the SNN, we have
\[ \left\langle \nabla \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \nabla \varphi \left( \theta^p(d_i^T x_t, \mu_t) \right), x_t \right\rangle \leq -\|x_t\|^2. \quad (20) \]

From Definition 2.1 and $\dot{\mu}_t \leq 0$, $\forall t \in [0, T)$, there is a $\kappa_{\tilde{f}} > 0$ such that
\[ \nabla_{\mu} \tilde{f}(x_t, \mu_t) \dot{\mu}_t \leq -\kappa_{\tilde{f}} \dot{\mu}_t. \quad (21) \]

By Proposition 3.1, we obtain
\[ \sum_{i=1}^{r} \nabla_{\mu} \varphi(\theta^p(d_i^T x_t, \mu_t)) \dot{\mu}_t \leq -r \rho \mu_t^{-1} \dot{\mu}_t. \quad (22) \]

Substituting (20)–(22) into (19) we have
\[
\frac{d}{dt} \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_t, \mu_t) \right) 
+ \kappa_{\tilde{f}} \dot{\mu}_t + r \rho \mu_t^{-1} \dot{\mu}_t \leq -\|x_t\|^2. \quad (23) \]

This implies that $\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t)) + \kappa_{\tilde{f}} \dot{\mu}_t + r \rho \mu_t^{-1} \dot{\mu}_t$ is a non-increasing function in $t$ on $[0, T)$, which follows that for any $t \in [0, T)$
\[
\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_t, \mu_t) \right) + \kappa_{\tilde{f}} \dot{\mu}_t + r \rho \mu_t^{-1} \dot{\mu}_t 
\leq \tilde{f}(x_0, \mu_0) + \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_0, \mu_0) \right) + \kappa_{\tilde{f}} \mu_0 + r \rho \mu_0^{-1} \mu_0. \]

On the other hand, from (10) and Proposition 3.1 in (2), we obtain
\[
\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_t, \mu_t) \right) + \kappa_{\tilde{f}} \mu_t + r \rho \mu_t^{-1} \mu_t \nonumber \]
\[ \geq \tilde{f}(x_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t) P^p). \quad (24) \]

Thus, for any $t \in [0, T)$
\[
f(x_t) + \sum_{i=1}^{r} \varphi \left( |d_i^T x_t|^p \right) \leq \tilde{f}(x_0, \mu_0) 
+ \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_0, \mu_0) \right) + \kappa_{\tilde{f}} \mu_0 + r \rho \mu_0^{-1} \mu_0. \quad (25) \]

Owing to the level boundedness of the objective function in $\mathcal{X}$, we obtain $x : [0, T) \rightarrow \mathbb{R}^n$ 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
\[ \|x_t\| \leq \rho \quad \forall t \in [0, \infty). \]

The next proposition shows that the locally Lipschitz continuity of $\nabla \varphi(\cdot)$ and $\nabla_{\mu} \tilde{f}(\cdot, \mu)$ 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 $\nabla_{\mu} \tilde{f}(\cdot, \mu)$ 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. $\lim_{t \to \infty} f(x(t)) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x(t)))$ exists;
2. $x(t) \in L^2(0, \infty)$ and $\lim_{t \to \infty} \|x(t)\| = 0$;
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 $\|x_t\| \leq \rho$, $\forall t \geq 0$. Thus, from (24), we obtain
\[
\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi \left( \theta^p(d_i^T x_t, \mu_t) \right) + \kappa_{\tilde{f}} \mu_t + r \rho \mu^{-1} \mu_t 
\geq \inf_{x \in \{x : \|x\| \leq \rho\}} \left[ f(x) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x)^p) \right]. \quad (26) \]

Equations (23) and (26) imply that
\[
\tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t)) + \kappa_{\tilde{f}} \mu_t + r \rho \mu^{-1} \mu_t 
\text{is non-increasing and bounded from below on } [0, \infty), \text{ then}
\lim_{t \to \infty} \left[ \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \varphi(\theta^p(d_i^T x_t, \mu_t)) + \kappa_{\tilde{f}} \mu_t + r \rho \mu^{-1} \mu_t \right] 
\text{exists.} \quad (27) \]
Thus, from (9), Proposition 3.1 and \( \lim_{t \to \infty} \mu_t = 0 \), we obtain
\[
\lim_{t \to \infty} \left[ \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \phi(\theta^p(d_t^T x_t, \mu_t)) + \kappa \mu_t + r l \phi \mu_t^p \right] = \lim_{t \to \infty} \left[ f(x_t) + \sum_{i=1}^{r} \phi([d_t^T x_t] r^p) \right].
\]
On the other hand, (23) and (27) imply that
\[
\lim_{t \to \infty} \frac{d}{dt} \left[ \tilde{f}(x_t, \mu_t) + \sum_{i=1}^{r} \phi(\theta^p(d_t^T x_t, \mu_t)) + \kappa \mu_t + r l \phi \mu_t^p \right] = 0. \tag{28}
\]
Owing to (23), we obtain
\[
\lim_{t \to \infty} \| \dot{x}_t \| = 0.
\]
From (26), integrating (23) from 0 to \( t \) gives
\[
\int_0^\infty \| \dot{x}_t \|^2 dt < \infty.
\]
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 \( \lim_{k \to \infty} l_k = \infty \).

From Theorem 3.1, we know that \( x^* \in \mathcal{X} \). Let
\[
g_{t_k} = x_{t_k} - P_{\mathcal{X}} \left[ x_{t_k} - \nabla_x \tilde{f}(x_{t_k}, \mu_{t_k}) - \sum_{i=1}^{r} \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k})) \right]. \tag{29}
\]
Owing 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
\[
\left[ \nabla_x \tilde{f}(x_{t_k}, \mu_{t_k}) + \sum_{i=1}^{r} \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k})) - g_{t_k}, u - (x_{t_k} - g_{t_k}) \right] \geq 0 \quad \forall u \in \mathcal{X}. \tag{31}
\]
From the definition of normal cone and the above inequality, we have
\[
0 \in \nabla_x \tilde{f}(x_{t_k}, \mu_{t_k}) + \sum_{i=1}^{r} \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k})) - g_{t_k}
+ N_{\mathcal{X}}(x_{t_k} - g_{t_k}). \tag{32}
\]
From (32) and
\[
N_{\mathcal{X}}(x_{t_k} - g_{t_k}) = \bigcup_{v \geq 0} v \partial d_{\mathcal{X}}(x_{t_k} - g_{t_k})
\]
there exist \( v_{t_k} \geq 0 \) and \( \xi_{t_k} \in \partial d_{\mathcal{X}}(x_{t_k} - g_{t_k}) \) such that
\[
0 = \nabla_x \tilde{f}(x_{t_k}, \mu_{t_k}) + \sum_{i=1}^{r} \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k})) - g_{t_k} + v_{t_k} \xi_{t_k}. \tag{33}
\]
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 \( x \to \infty \) \( \partial d_{\mathcal{X}}(x) \) is upper semicontinuous and \( \lim_{k \to \infty} x_{t_k} - g_{t_k} = x^* \), then
\[
\left\{ \lim_{k \to \infty} \xi_{t_k} \right\} \subseteq \partial d_{\mathcal{X}}(x^*). \tag{34}
\]
Combining the above relationship with \( \lim_{k \to \infty} x_{t_k} = x^* \), we have
\[
\left\{ \lim_{k \to \infty} v_{t_k} \xi_{t_k} \right\} \subseteq \bigcup_{v \geq 0} v \partial d_{\mathcal{X}}(x^*) = N_{\mathcal{X}}(x^*). \tag{35}
\]
Thus
\[
0 \in \partial f(x^*) + \sum_{i=1}^{r} \lim_{k \to \infty} \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k})) + N_{\mathcal{X}}(x^*). \tag{35}
\]
From (13), we obtain
\[
x_{t_k}^T \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k}))
= \begin{cases} 
p|d_{t_k}^T x_{t_k}|^p, & \text{if } |d_{t_k}^T x_{t_k}| > \mu_{t_k} \\
p \left( \frac{|d_{t_k}^T x_{t_k}|^2 + \mu_{t_k}^2}{2\mu_{t_k}} \right)^{p-1} \frac{|d_{t_k}^T x_{t_k}|^2}{\mu_{t_k}}, & \text{if } |d_{t_k}^T x_{t_k}| \leq \mu_{t_k}. \end{cases} \tag{36}
\]
which implies that
\[
\lim_{k \to \infty} x(t_k)^T \sum_{i=1}^{r} \nabla_x \phi(\theta^p(d_{t_k}^T x_{t_k}, \mu_{t_k})) = p|d_t^T x^*|^p.
\]
Therefore, we obtain
\[
0 \in (x^*)^T \partial f(x^*) + p\sum_{i=1}^{r} |d_t^T x^*|^p \nabla \phi(|d_t^T x^*|^p) + (x^*)^T N_{\mathcal{X}}(x^*)
\]
which means \( x^* \) is a stationary point of (3).

If \( x_t \) has two accumulation points \( x^* \) and \( y^* \), there are two sequences \( s_k \) and \( s_k \) such that
\[
\lim_{k \to \infty} s_k = \infty, \lim_{k \to \infty} x_{s_k} = x^* \text{ and } \lim_{k \to \infty} x_{s_k} = y^*.
\]
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 \( p \geq 1 \), the critical point set of (3) is
\[
C = \left\{ x \in \mathcal{X} : 0 \in \partial f(x) + \sum_{i=1}^{r} \partial \phi(|d_t^T x|^p) + N_{\mathcal{X}}(x) \right\}
\]
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 \( p \geq 1 \), and to the optimal solution set of (3) when \( f \) is convex and \( p \geq 1 \).
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
\[
0 \in \partial f(x^*) + \sum_{i=1}^r \nabla \varphi(|d_i^T x^*|^p) + N_X(x^*).
\]

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 \( x_0 \in X \), any accumulation point \( x^* \) of the solutions of the SNN satisfies
\[
\begin{cases}
x^* \in \mathcal{X} \\
0 \in X^* \partial f(x^*) + p \nabla \varphi(|x^*|^p)|x^*|^p + X^* N_X(x^*)
\end{cases}
\]  
(37)

where \( X^* = \text{diag}(x_1^*, \ldots, x_n^*) \) and
\[
\nabla \varphi(|x_i^*|^p)|x_i^*|^p = (\nabla \varphi_1(|x_1^*|^p)|x_1^*|^p, \ldots, \nabla \varphi_n(|x_n^*|^p)|x_n^*|^p)^T.
\]

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. \( N^k \): number of nonzero elements of \( x^* \).
4. PSNR(dB) and SNR(dB): signal-to-noise ratios defined by
\[
\text{PSNR} = -10 \log \frac{\|x^* - s\|^2}{n}, \text{ SNR} = -20 \log \frac{\|x^* - s\|}{\|s\|}
\]
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 \( \text{Oracle} = \sigma^2 \text{tr}(A^T A)^{-1} \), where \( A = \text{support}(s) \). And we define Radio = \(\|x^* - s\|/\text{Oracle} \).

#### A. Image Restoration

Image restoration is to reconstruct 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 \( 64 \times 64 \) 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
\[
h(i, j) = e^{-z(i)^2 - z(j)^2}
\]
which is truncated such that the function has a support of \( 7 \times 7 \). 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:
\[
\begin{align*}
\min & \quad \|Ax - b\|^2 + 0.02 \sum_{i=1}^r \psi(|d_i^T x|^p) \\
\text{s.t.} & \quad x \in \mathcal{X} = \{x : x_i \in [0, 1], i = 1, 2, \ldots, n\},
\end{align*}
\]  
(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, \( \psi : [0, +\infty) \rightarrow [0, +\infty) \) is the potential function defined by
\[
\psi(z) = \frac{0.5z}{1 + 0.5z}.
\]

\( d_i^T \in \mathbb{R}^n \) is the \( i \)th 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 \( x_0 = P_X(b) \). 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
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

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

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

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 $l$th column of $S$ is the signal at time $l$, which is with dimension $N$ and sparsity $T$.

```matlab
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 $b \in \mathbb{R}^{K \times L}$.

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

**TABLE I**

<table>
<thead>
<tr>
<th>Objective value</th>
<th>Original</th>
<th>Observed</th>
<th>$p = 0.5$</th>
<th>$p = 0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR(dB)</td>
<td>14.486</td>
<td>39.461</td>
<td>16.784</td>
<td>16.784</td>
</tr>
<tr>
<td></td>
<td>15.503</td>
<td>20.686</td>
<td>20.924</td>
<td></td>
</tr>
</tbody>
</table>

**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 = $\text{randn}(K, N)$; for $j = 1:N$

$A(:, j) = A(:, j)/\text{norm}(A(:, j))$;

end

noise = $\sigma*\text{randn}(K, L)$;

$b = A*S-noise$;

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

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

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

$$X = \{x \in \mathbb{R}^n : Ax = b(l)\}$$

$$P_X(x) = x - A^T(AA^T)^{-1}(Ax - b(l))$$

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 Abubgarbieh [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 $\sigma^2$. 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
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, lbph, svi, lcp, pgleason, 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:

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

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

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

(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 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...
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 with a random initial point in \( \mathcal{X} \).

### 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

\[
\kappa(Q) = \frac{\lambda_{\text{max}}(Q)}{\lambda_{\text{min}}(Q)}
\]

where \( \lambda_{\text{max}}(Q) \) and \( \lambda_{\text{min}}(Q) \) 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

\[
\min_{x \in \mathcal{X}} \kappa(Q(x)) = \kappa \left( I + \sum_{i=1}^{n} x_i Q_i \right)
\]

s.t. \( x \in \mathcal{X} = \{ x : 0 \leq x_i \leq 1, i = 1, 2, \ldots, n \} \)

where \( n = 100, I \) is the 5 \( \times \) 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

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

With a random initial point in \( \mathcal{X} \), Fig. 8 presents the convergence of \( \lambda_{\text{max}}(Q(x(t))) \) and \( \lambda_{\text{min}}(Q(x(t))) \) 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

\[
\min_{x \in \mathcal{X}} \kappa(Q(x)) = \kappa \left( I + \sum_{i=1}^{n} x_i Q_i \right) + \lambda \sum_{i=1}^{n} |x_i|^p
\]

s.t. \( x \in \mathcal{X} = \{ x : 0 \leq x_i \leq 1, i = 1, 2, \ldots, n \} \).

The objective function of (44) is the combining of a nonsmooth, nonconvex function, and a non-Lipschitz function.
We use the SNN modeled by
\[
\dot{x}(t) = -x(t) + \lambda \left[ x(t) - \nabla_s f(x(t), \mu(t)) \right] - \lambda \nabla_s \theta^p (x(t), \mu(t))
\]  
(45)
where \( \theta \) is defined as in (11), \( \lambda \) is a positive parameter.

Table V shows the numerical results for different values of \( \lambda \) in (45). From the results in Table V, we see that as \( \lambda \) 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 \( \lambda \) 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 stationarity 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
\[
|\nabla_x \theta^p (d_i^T x, \mu)| \leq \left\{ \begin{array}{ll}
0, & \text{if } |d_i^T x| > \mu \\
1, & \text{if } |d_i^T x| \leq \mu.
\end{array} \right.
\]
Thus, from (11), (15), and (17), we obtain
\[
|\nabla_x \varphi (\theta^p (d_i^T x, \mu))| \leq p \|d_i\| \mu^{-1}.
\]
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
\[
\nabla_s^2 \theta^p (s, \mu) > 0 \ \forall s \in (-\mu, \mu)
\]  
(47)
which means that \( \theta^p (s, \mu) \) is strongly convex in \( s \) on \([-\mu, \mu] \).
For any fixed \( \mu \), 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 \( \varphi \) is globally Lipschitz, we obtain 2).
3) First, we prove the Lipschitz property of \( \nabla_x \theta^p (s, \mu) \) in \( s \) for any fixed \( \mu > 0 \), which can be derived by the global boundedness of the Clarke generalized gradient of \( \nabla_x \theta^p (s, \mu) \). \( \nabla_x \theta^p (s, \mu) \) is not differentiable only when \( |s| = \mu \). From (7), we only need to prove the global boundedness of the gradient of \( \nabla_x \theta^p (s, \mu) \) when \( |s| \neq \mu \).
Fix \( \mu > 0 \). When \( |s| > \mu \), we obtain
\[
|\nabla_s^2 \theta^p (s, \mu)| = |p(p-1)|s|^{p-2}| \leq p \mu^{-2}.
\]
When \( |s| < \mu \), we have
\[
\nabla_s^2 \theta^p (s, \mu) \leq p \left( \frac{|s|^2}{2 \mu} + \frac{\mu}{2} \right) \left( \frac{1}{\mu} \right) \leq 2p \mu^{-2}.
\]  
(48)
From (47), then
\[
|\nabla_s^2 \theta^p (s, \mu)| \leq 2p \mu^{-2} \ \forall s \in (-\mu, \mu).
\]  
(49)
Combining (7), (48), and (49), we have
\[
|\xi| \leq 2p \mu^{-2} \ \forall \xi \in \mathbb{R}, \ \xi \in \partial \varphi (\nabla_x \theta^p (s, \mu))
\]
which implies that
\[
|\nabla_x \theta^p (s, \mu) - \nabla_x \theta^p (r, \mu)| \leq 2p \mu^{-2} |s-r| \ \forall s, r \in \mathbb{R}.
\]
Therefore, for any \( x, z \in \mathbb{R}^n \), we get
\[
\|\nabla_x \theta^p (d_i^T x, \mu) - \nabla_x \theta^p (d_i^T z, \mu)\| \leq 2p \mu^{-2} \|d_i^T \| \|x - z\|
\]  
(50)
4) From (13), we have
\[
\|\nabla_x \theta^p (d_i^T x, \mu)\| \leq 2p \|d_i\| \mu^{-1}
\]  
(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 \( \nabla \varphi (\theta^p (d_i^T x, \mu)) \) in \( x \) for any fixed \( \mu \).
APPENDIX B
PROOF OF PROPOSITION 3.2
Denote $x_t, \hat{x}_t \in AC[0, \infty)$ two solutions of the SNN with initial point $x_0$. From Theorem 3.1, there is a $p > 0$ such that $\|x_t\| \leq \rho$ and $\|x_t\| \leq \rho$, $\forall t \geq 0$. 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 $x_{t+\delta} \neq \hat{x}_{t+\delta}$.

Denote $k(x, \mu) = -x + P_N \left[ x - \nabla_x \hat{f}(x, \mu) - \sum_{i=1}^{r} \nabla_x \phi(d_i^T x, \mu) \right].$

Form Propositions 2.2 and 3.1, when $\forall a \leq b$, $\hat{x}_a \neq x_b$, and $x$ is locally Lipschitz, $k(\cdot, \mu)$ is locally Lipschitz for any fixed $\mu > 0$, which means that there exists an $L_{\mu} > 0$ such that

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

Since $x_t, \hat{x}_t$, and $\mu_t$ 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

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

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


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.