Differential Inclusion Neural Network for Compressed Sensing

Zixin Liu[1,2]
[1] School of Electronic Engineering
Beijing University of Posts and Telecommunications
Beijing, P.R.China, 102209
[2] Department of Mathematics and Statistics
Guizhou University of Finance and Economics
Guiyang, P.R.China, 550025
*E-mail: xinxin905@163.com

Yuanan Liu
School of Electronic Engineering
Beijing University of Posts and Telecommunications
Beijing, P.R.China, 102209
E-mail: yuliu@bupt.edu.cn

Abstract—**The issue on neural network method to solve compressed sensing problem is concerned. Combined with optimization technique, nonsmooth analysis theory, differential inclusion theory, and set-valued analysis method, a classical approximate compressed sensing model with dense noise is transformed into a differential inclusion neural network model. On the basis of existence and stable theory, some existence and stability results are also given.**

Keywords—compressed sensing; differential inclusion theory; nonsmooth analysis; neural network; set-valued analysis

I. INTRODUCTION

In recent years, compressed sensing (CS) theory [1.2] arouses many domestic and foreign scholars' research interest. It is widely applied to many different research fields, such as various compression imaging, MRI and CT, face recognition, variable selection, biological computing, remote sensing, LASSO and so on [3-5]. A typical compressed sensing model under-determined linear system of equations can be found by solving the following so-called " l_0 -norm" minimization problem, i.e.,

$$\min_{X \in R^n} \| X \|_0 \quad s.t. \quad AX = b, \qquad (1)$$

where $A \in R^{k \times n}(k << n), b \in R^k$, R^n is a n-dimensional vector space; $\| X \|_0$ denotes the number of nonzero components in X. As pointed out in [6], although problem (1) is widely applied in many different science fields, it is NP-hard, this makes the theoretical research for problem (1) becomes very difficult. Recently, in order to overcome this problem, many different new approximate models for (1) are established. Among all of these approximate models, basis pursuit (BP) is the most popular one. It becomes the best alternative model for problem (1), since the convex envelope of $\| X \|_0$ is $\| X \|_1$, where $\| X \|_1 = \sum_i | x_i |$ is the l_1 norm of X. "Basic Pursuit problem" [7] was proposed as follows

$$\min_{X \in R^n} \| X \|_1 \quad s.t. \quad AX = b. \qquad (2)$$

If term b in problem (2) is contaminated by small dense noise, a natural approach to relax the constraint in (2) yields the following model

$$\min_{X \in R^n} \| X \|_1 \quad s.t. \quad \| AX - b \|_2 < \varepsilon, \qquad (3)$$

where $\| X \|_2$ is Euclidean norm of X, and ε is non-negative parameter. By convex analysis theory, problem (3) can be equivalently transformed into Lagrangian version, i.e., the following $l_1 - l_2$ minimization

$$\min_{X \in R^n} \frac{1}{2} \| b - AX \|_2^2 + \tau \| X \|_1, \qquad (4)$$

where $\tau > 0$ is a positive constant.

Obviously, problem (4) is a convex optimization problem, but not a smooth one, since term $\| X \|_1 = \sum_i | x_i |$ is not differentiable. In order to solve this approximate compressed sensing problem (4) with dense noise, many different methods are proposed, such as smooth method, subdifferential method, filling method, and so on. In what follows, this paper attempts to utilize differential inclusion neural network to handle this problem.

II. DIFFERENTIAL INCLUSION NEURAL NETWORK MODEL DESCRIPTION

Since problem (4) is a nonsmooth optimization problem in essence. Thus we can utilize the relationship between nonsmooth optimization and dynamic system to establish new differential inclusion neural network for model (4). Before proceed, we first introduce some basic definitions and classical results as follows.

A. Basic Definitions

Definition 1. Suppose $\Omega \subset R^n$, then $F: x \to F(x)$ is called a set-valued function from $\Omega \to R^n$, if for each point

$x \in \Omega$, the corresponding image set $F(x)$ is a nonempty and closed set satisfying $F(x) \subset R^n$.

Definition 2. A function $f(x): R^n \to R$ is said to be Lipschitz near $x \in R^n$, if there exist positive number δ and ε such that for any $x_1, x_2 \in R^n$ satisfies $\| x_1 - x \| < \delta$ and $\| x_2 - x \| < \delta$, we have $| f(x_1) - f(x_2) | \le \varepsilon \| x_1 - x_2 \|$. If $f(x): R^n \to R$ is Lipschitz near any point $x \in R^n$, then $f(x)$ is also said to be locally Lipschitz in R^n.

Definition 3. If $f(x): R^n \to R$ is Lipschitz near point x, then the generalized directional derivative of $f(x)$ at point x in the direction $\upsilon \in R^n$ is given by

$$f^0(x; \upsilon) = \limsup_{\substack{y \to x \\ s \to 0^+}} \frac{f(y + s\upsilon) - f(y)}{s}.$$

Definition 4. The Clarke's generalized gradient of $f(x)$ at point x is defined as

$$\partial f(x) = \{ y \in R^n : f^0(x; \upsilon) \ge y^T \upsilon, \forall \upsilon \in R^n \}.$$

When $f(x): R^n \to R$ is locally Lipschitz in R^n, and differentiable for almost all $x \in R^n$, then the Clarke's generalized gradient of $f(x)$ at point x is equivalent to

$$\partial f(x) = K \{ \lim_{n \to \infty} \partial f(x_n) : x_n \to x, x_n \notin N, x_n \notin E \},$$

where $K(\cdot)$ denotes the closure of convex hull, $N \subset R^n$ is an arbitrary set with measure zero, and $E \subset R^n$ is the set points where $f(x)$ is not differentiable.

B. Classical Results Introduction of Projection Neural Network for Nonsmooth Optimization Subject to Linear Equality and Bound Constraint

Consider the following general nonsmooth optimization problem

$$\text{minimize } f(x)$$
$$\text{subject to } Ax = b \qquad (5)$$
$$x \subset \Omega$$

where $x = (x_1, x_2, \cdots, x_n)^T \in R^n$, $f(x): R^n \to R$ is an objective function which is not necessarily convex and smooth. By using optimization technique, nonsmooth analysis theory, differential inclusion theory, set-valued analysis method, and Clarke's subdifferential results, Liu and Wang in [8] established the following projection neural network to solve problem (5)

$$\varepsilon \frac{dy}{dt} \in -MP_\Omega(y) - (I - M)(y - P_\Omega(y)) + \partial f((I - M)P_\Omega(y) + q) + q \quad (6)$$
$$x = P_\Omega(y)$$

where $M = A^T (AA^T)^{-1} A, q = A^T (AA^T)^{-1} b$, $\partial f(x)$ is the Clarke's generalized gradient of $f(x)$, and $P_\Omega(\cdot)$ is a projection operator defined by

$$P_\Omega(u) = \arg\min_{\upsilon \in \Omega} \| u - \upsilon \|.$$

Specially, when problem (5) degenerates into the following form

$$\text{minimize } f(x)$$
$$\text{subject to } Ax = b. \qquad (7)$$

then problem (7) can be solved by the following projection neural network

$$\varepsilon \frac{dx}{dt} \in -Mx - (I - M)\partial f((I - M)x + q) + q. \quad (8)$$

When problem (5) degenerates into the following form

$$\text{minimize } f(x)$$
$$x \subset \Omega \qquad (9)$$

then problem (9) can be solved by the following neural network

$$\varepsilon \frac{dy}{dt} \in -y + P_\Omega(y) - \partial f(P_\Omega(y)) \qquad (10)$$
$$x = P_\Omega(y).$$

Fact 1 (existence result):

Assume that the objective function $f(x)$ in problem (5), (7) and (9) is pseudoconvex on the feasible region. Then $x^* \in R^n$ is an optimal solution of problem (5), (7) and (9) if and only if there exists an equilibrium point $y^* \in R^n$ for system (6), (8) and (10) such that $x^* = P_\Omega(y^*)$.

Fact 2 (stability results):

(1) If function $f(x)$ in problem (5) is convex on Ω, then for any initial value y_0, the output vector of neural network (6) is globally convergent to the optimal solution set M if state vector $y(t)$ is bounded and

$$\forall x \in \Theta, (x - x^*)^T (\gamma - \gamma^*) = 0 \Leftrightarrow x \in M$$

where $x^* \in M, \gamma \in \partial f(x), \gamma^* \in \partial f(x^*)$.

(2) If function $f(x)$ in problem (7) is pseudoconvex on the equality constraint set, then for any initial value x_0, the state

vector of neural network (8) is globally convergent to the optimal solution set M

$$\lim_{t \to \infty} dist(x(t), M) = 0.$$

(3) If function $f(x)$ in problem (9) is pseudoconvex on the bounded constraint set Ω, then for any initial value y_0, the output vector of neural network (10) is globally convergent to the optimal solution set M if the output vector $y(t)$ is bounded

$$\lim_{t \to \infty} dist(x(t), M) = 0.$$

C. Differential Inclusion Neural Network for Compressed Sensing

From definition 4, for l_1 norm $\| \cdot \|_1$, the subdifferential of $\| x \|_1$ at x^* is $\partial \| x \|_1$, and the ith component is defined as

$$\partial_i \| x^* \|_1 = \begin{cases} -1 & x_i^* > 0 \\ [-1,1] & x_i^* = 0 \\ 1 & x_i^* < 0 \end{cases}. \qquad (11)$$

Thus

$$\partial \| (I - M)x + q \|_1 = (I - M)^T \partial \| x \|_1.$$

From (7) and (8), the following differential inclusion neural network can be established to solve problem (2)

$$\varepsilon \frac{dx}{dt} \in -Mx - (I - M)(I - M)^T \partial \| x \|_1 + q. \qquad (12)$$

Since M is a projection matrix, this means that $M^2 = M$. Utilizing the property of $\partial \| x \|_1$, Liu and Wang further simplified model (12) as following projection neural network [9]

$$\varepsilon \frac{dz}{dt} = -Mz - (I - 2M)P_\Omega(z) + q, \qquad (13)$$

$$x = z - P_\Omega(z),$$

where $\Omega = \{x \in R^n \mid -1 \le x_i \le 1, i = 1,2,\cdots,n\}$,

$$P_\Omega(z) = \arg\min_{\gamma \in \Omega} \| z - \gamma \|_2, \text{ and}$$

$$P_\Omega(z_i) = \begin{cases} 1 & z_i > 1 \\ z_i & -1 \le z_i \le 1 \\ 1 & z_i < -1 \end{cases}.$$

Similarly, for problem (4), when $\Omega = R^n$, it yields that

$$x = P_\Omega(x) = P_{R^n}(x).$$

From (9) and (10), the following differential inclusion neural network can be established to solve problem (4)

$$\varepsilon \frac{dx}{dt} \in -\partial f(x), \qquad (14)$$

where

$$\partial f(x) = A^T Ax - A^T b + \partial \| x \|_1.$$

Set x^* is an equilibrium point of system (14), $\gamma^* \in \partial \| x \|_1$, From (11) and (14), it obtains

$$\gamma^* = P_\Omega(\gamma^* + x^*),$$

$$\partial f(x^*) = A^T Ax^* - A^T b + \partial \| x^* \|_1 = 0$$

where $\Omega = \{x \in R^n \mid -1 \le x_i \le 1, i = 1,2,\cdots,n\}$.

Let $z^* = \gamma^* + x^*$, obviously, $x^* = z^* - P_\Omega(z^*)$, and $\gamma^* = P_\Omega(z^*)$, we have

$$A^T A(z^* - P_\Omega(z^*)) - A^T b + \gamma^* = 0$$

Namely,

$$A^T A(z^* - P_\Omega(z^*)) - A^T b + P_\Omega(z^*) = 0 \qquad (15)$$

Thus, differential inclusion neural network (14) can be equivalently rewritten into the following projection neural network

$$\frac{dz}{dt} = -A^T Az + (A^T A - I)P_\Omega(z) + A^T b \qquad (16)$$

$$x = z - P_\Omega(z)$$

D. Existence and Stability Analysis

From problem (2) and (4), one can see that although function $\| X \|_1$ and $\frac{1}{2} \| b - AX \|_2^2 + \tau \| X \|_1$ are both nonsmooth function, they are convex on R^n. Thus they both satisfy the conditions in fact 1 and fact 2. Hence the following existence results and stability results hold naturally.

For system (12), (13), (14) and (16), the existence results can be described as follows.

Theorem1. Let $x^* \in R^n$ is an optimal solution of problem (2), if and only if there exists an equilibrium point $x^* \in R^n$ for differential inclusion neural network (12), or there exists an equilibrium point $x^* \in R^n$ for projection neural network (13) such that $x^* = z^* - P_\Omega(z^*)$.

Theorem2. Let $x^* \in R^n$ is an optimal solution of problem (4), if and only if there exists an equilibrium point $x^* \in R^n$ for differential inclusion neural network (14), or there exists an

equilibrium point $x^* \in R^n$ for projection neural network (16) such that $x^* = z^* - P_\Omega(z^*)$.

For system (12) and (13), the stability results can be described as follows.

Theorem3. For any initial value x_0, the state vector of differential inclusion neural network (12) is globally convergent to the optimal solution set M of problem (2)

$$\lim_{t \to \infty} dist(x(t), \mathrm{M}) = 0,$$

or for any initial value z_0 the state vector of projection neural network (13) is globally convergent to the equilibrium z^* such that

$$x^* = z^* - P_\Omega(z^*) \in \mathrm{M}.$$

From the expressions of (14) and (16), notice that $A^T A \geq 0$, similar to the proof in [10] and [11], it follows that state vector $x(t)$ of (14), and the state vector $z(t)$ of (16), are bounded. From fact 2, the stability results for system (14) and (16) can be described as follows

Theorem4. For any initial value x_0, the state vector of differential inclusion neural network (14) is globally convergent to the optimal solution set M of problem (4)

$$\lim_{t \to \infty} dist(x(t), \mathrm{M}) = 0,$$

or for any initial value z_0, the state vector of projection neural network (16) is globally convergent to the equilibrium z^* such that

$$x^* = z^* - P_\Omega(z^*) \in \mathrm{M}.$$

III. CONCLUSIONS

Applying nonsmooth and non-convex optimization method, this paper further researches the compressed sensing problem; Combined with optimization technique, nonsmooth analysis theory, differential inclusion theory, and set-valued neural network analysis method, some differential inclusion neural networks and projection neural networks for compressed sensing problem with or without dense noise are established. By using existence and stable theory of differential equation, some existence and stability results for our established neural network models are also given.

REFERENCES

[1] D.L. Donoho, "Compressed sensing", IEEE Trans. Inform. Theory, vol. 52 (4), pp. 1289‑1306, 2006.

[2] E.J. Candés, and M.B. Wakin, "An introduction to compressive sampling", IEEE Signal Process. Mag., vol. 25, pp. 21‑30, 2008.

[3] H. Jung, J. Ye, E. Kim, "Improved k-t blask and k-t sense using focus", Phys. Med. Biol., vol. 52 (11) , pp. 3201-3226, 2007.

[4] A.M. Andrés, S. Padovani, M. Tepper, et al., "Face recognition on partially occluded images using compressed sensing", Pattern Recognit. Lett. , vol. 36, pp. 235-242, 2014.

[5] R. Tibshirani, "Regression shrinkage and selection via the LASSO", J. R. Stat. Soc.: Ser. B, vol.58, pp. 267-288, 1996.

[6] B.K. Natarajan, "Sparse approximation solutions to linear systems", SIAM J. Sci. Comput. Vol. 24 (2), pp.227‑234, 1995.

[7] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput.,Vol. 20 , pp. 33‑61,1998.

[8] Q. S. Liu, and J. Wang, "A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints", IEEE Trans. Neural Network and Learning systems. vol. 24, no 5, pp.812-824, 2013.

[9] Q. S. Liu, and J. Wang, "L1-Minimization algorithms for sparse signal reconstruction based on a projection neural network", IEEE Trans. Neural Network and Learning systems. vol. 27, no.3, pp.698-707, 2016.

[10] Z.X. Liu, and N.F, Wang, "Neural Network to Solve Concave Games", International Journal of Computer Games Technology, vol. 2014, pp. 1-8, 2014.

[11] Jack K. Hale Sjoerd M. Verduyn Lunel, "Introduction to functional differential equations", Springer Science+Business Media LLC, New York, 1993.