Distributed Inertial Proximal Neurodynamic Approach for Sparse Recovery on Directed Networks

You Zhao, Xing He, *Member, IEEE*, Mingliang Zhou, Junzhi Yu, *Fellow, IEEE*, and Tingwen Huang, *Fellow, IEEE*

***Abstract*—This article investigates a fully distributed inertial neurodynamic approach for sparse recovery. The approach is based on proximal operators and inertia items. It aims to solve the L_1-norm minimization problem with consensus and linear observation constraints over directed communication networks. The proposed neurodynamic approach has the advantages of only requiring the communication network to be directed and weight-balanced, does not involve a central processing node and global parameters, which means that no single node can access the entire network and observe it at any time, so it is fully distributed. To effectively deal with the nonsmooth objective function, L_1-norm, the proximal operator method is used here. For efficiently handling linear observation and consensus constraints, a primal-dual method is applied to the inertial dynamic system. With the aid of maximal monotone operator theory and Baillon–Haddad lemmas, it reveals that the trajectories of our approach can converge to consensus solution at the optimal solution, provided that the distributed parameters satisfy technical conditions. In addition, we aim to demonstrate the weak convergence of the trajectories in our proposed neurodynamic approach toward the zeros of the optimal operator in Hilbert space, using Opial's lemma. Finally, comparative experiments on sparse signal and image recovery confirm the efficiency and effectiveness of our proposed neurodynamic approach.**

***Index Terms*—Distributed inertial neurodynamic approach, maximal monotone operator, proximal operator, sparse recovery, weak convergence.**

Manuscript received 24 February 2024; accepted 28 May 2024. Date of publication 31 July 2024; date of current version 18 September 2024. This work was supported in part by the Natural Science Foundation of China under Grant 62176218 and Grant 62176027; in part by the Fundamental Research Funds for the Central Universities under Grant XDJK2020TY003; and in part by the Chongqing Talent under Grant cstc2024ycjh-bgzxm0082. This article was recommended by Associate Editor S. Agaian. *(Corresponding author: Xing He.)*

You Zhao and Xing He are with the College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China (e-mail: Zhaoyou1991sdtz@163.com; hexingdoc@swu.edu.cn).

Mingliang Zhou is with the School of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: mingliangzhou@cqu.edu.cn).

Junzhi Yu is with the State Key Laboratory for Turbulence and Complex Systems, Department of Advanced Manufacturing and Robotics, College of Engineering, Peking University, Beijing 100871, China, and also with the Nanchang Innovation Institute, Peking University, Nanchang 330224, China (e-mail: junzhi.yu@ia.ac.cn).

Tingwen Huang is with the Science Program, Texas A&M University at Qatar, Doha, Qatar (e-mail: tingwen.huang@qatar.tamu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2024.3408473.

Digital Object Identifier 10.1109/TSMC.2024.3408473

I. Introduction

SPARSE representation serves as a fundamental data science method that is used to solve a breadth of problems in science and engineering. Sparse recovery, as an important scientific problem in sparse representation, has emerged as one of research hotspots in signal processing, data analysis and pattern recognition [1], [2], facial expression recognition [3] and fault diagnosis [4]. Mathematically, sparse recovery can be phrased as a constrained nonconvex optimization problem as follows:

$$\min_{\mathrm{x}\in\mathbb{R}^n} \|\mathrm{x}\|_0, \;\; \text{s.t. } \mathrm{A}\mathrm{x} = b \tag{1}$$

where $\|\mathrm{x}\|_0$ is called L_0-norm (pseudo-norm), $\mathrm{A} \in \mathbb{R}^{m\times n}$, $(m \ll n)$, and $b \in \mathbb{R}^m$. Owing to its nonconvexity, it is NP-hard to obtain a globally optimal solution of the problem (1). It is known from compressed sensing theory that when the observation matrix A satisfies certain conditions (such as, $\|x\|_0 = S$ and every group of $2S$ columns of A is linearly independent, or matrix A fulfills the restricted isogeometric property (RIP) condition), problem (1) allows for convex relaxation as L_1-norm minimization problem [5]

$$\min_{\mathrm{x}\in\mathbb{R}^n} \|\mathrm{x}\|_1, \;\; \text{s.t. } \mathrm{A}\mathrm{x} = b. \tag{2}$$

Problem (2) is convex but nonsmooth that has been widely exploited for signal and image processing [6], sparse coding [7]. To resolve the problem (2), many prominent centralized discrete numerical algorithms have been investigated, including the proximal point algorithm [8], projection gradient algorithm [9], and orthogonal matching pursuit (OMP) approach [10].

In addition, an enormous number of neurodynamic approaches [11], [12], [13], [14], [15] have been proposed to conduct sparse recovery in a centralized manner due to their advantage of physical simulation system implementation and the possibility of using Lyapunov stability theory in convergence analysis. Inspired by the projection neural networks (PNNs), various continuous and discrete-time neurodynamics approaches have been studied by Liu and Wang [11] for sparse recovery by tackling the problem (2). With a combination of the local competitive algorithm (LCA) and the Lagrangian programming neural network (LPNN), Feng et al. [12] proposed a LPNN-LCA for addressing problem (2) but only provided a local convergence property of LPNN-LCA. Later, Wang et al. [13] further investigated an improved LPNN-LCA framework and gave its global asymptotic stability property by means of the projection theorem. Recently, in virtue of the sliding mode techniques, He et al. proposed two novel neurodynamic approaches for sparse recovery via minimizing the problem (2), and they have finite-time [14] and fixed-time [15], [16] convergence rates, respectively.

In recent years, with the ever-increasing size of optimization problems, distributed neurodynamic approaches have gained more attractiveness, and are even necessary under some certain scenarios. Compared to centralized methods, a distinguishing feature of distributed methods is that they require a communication network to exchange information. Therefore, the performance of distributed dynamic approaches is highly dependent on the communication network structure. Recently, numerous distributed neurodynamic approaches have been investigated for addressing optimization problems in distributed manner over undirected or directed communication networks [17], [18], [19], [20], [21], [22]. Liu and Wang [17] devised a distributed multiagent neurodynamic approach with the help of differential inclusion method (i.e., subgradient-based method) for nonsmooth constrained convex optimization problems over undirected graphs. Two distributed continuous-time projection algorithms (DCTAs) based on subgradients and derivative feedback items have been proposed in [18] to address nonsmooth convex problems, which have set and equality constraints over undirected graphs. Le et al. [19] developed a distributed neurodynamic approach using the differential inclusion method to tackle nonsmooth convex optimization problems equipped with both coupled equality and inequality constraints under undirected networks. Zhang et al. [38], [39] proposed two novel distributed optimization algorithms with event-triggered strategy for solving two-layered model problems in power grids. Yi et al. [20] investigated a distributed inertial (second-order) dynamical approach for resolving smooth convex problems with the consensus of constraints over undirected graphs. Later, Wei et al. [21] proposed a distributed smooth double proximal primal-dual continuous-time approach to address distributed nonsmooth optimization problems with consensus constraints under undirected graph. In the combination with proximal operator and Lagrangian methods, Wang et al. [22] proposed a novel distributed inertial proximal-gradient continuous-time approach with derivative feedback over undirected graphs. This approach is designed to solve nonsmooth convex optimization problems that are equipped with consensus constraints. On the basis of differential inclusion method, Gharesifard and Cortés [23] proposed a distributed continuous-time approach for distributed unconstrained nonsmooth convex problems over directed graphs. Jiang et al. [24] investigated a distributed continuous-time approach by using differential inclusion method, which is used to deal with an approximate nonsmooth distributed constrained (affine equality and inequality constraints) optimization problem under directed graph. Wang et al. [25] investigated a second-order primal-dual projected dynamical approach to effectively address distributed smooth convex optimization problems subject to consensus and set constraints over directed networks. Wang et al. [26] addressed the output consensus problem for heterogeneous multiagent systems subject to input saturation constraints in the presence of Markovian randomly switching topologies. Furthermore, a fully distributed antiwindup control protocol was proposed in [27] to address the intelligent interconnected electric vehicle platooning problem, featuring switch topologies and input saturation capability.

Recently, Zhao et al. [28] made a pioneering effort to explore the distributed version of problem (2) over undirected graphs. They proposed two distributed neurodynamic approaches specifically tailored to tackle the problem (2) for recovering sparse signals. Later, Xu et al. [29] designed a novel distributed neurodynamic approach with two-layer structure for solving the distributed version of problem (2) over undirected graphs. Depending on projection and primal-dual dynamical methods, Zhao et al. [30] designed two distributed neurodynamic approaches in continuous and discrete-time to solve distributed version of the problem (2) for sparse recovery on the undirected graph. Notably, the distributed neurodynamic approaches [28], [29], [30] referred to above can be used effectively to tackle the problem (2) in a distributed manner but all require the communication topology to be undirected. However, in practical applications, the communication of networks requires directed graphs. The neurodynamic approaches [28], [29], [30] are designed with first-order dynamical systems that have slow convergence rates. The above neurodynamic approaches [28], [29], [30] only provide the asymptotic convergence properties. In addition, the subgradient-based distributed neurodynamic approaches [23] and [24] can deal with nonsmooth constrained convex optimization problems under directed graphs, however the issue of subgradient selection in sparse recovery a challenge for these above proposed approaches.

Motivated by the discussion above, we investigate novel distributed neurodynamic methods with inertial term for sparse recovery by solving the problem (2) over directed graphs. Furthermore, this article presents four major contributions, which are summarized as follows.

1) A novel fully distributed inertial proximal neurodynamic approach (DIProx-NA) is proposed to address the problem (2) in a distributed way for sparse recovery. Compared with the existing works [28], [29], [30], the DIProx-NA is an inertial dynamical system. Moreover, the communication graph can be directed in here, and it is more general as compared to the undirected graph.
2) To deal with the nonsmooth objective function (i.e., L_1-norm), we utilize the proximal operators instead of the subgradients (i.e., differential inclusion method) [17], [18], [23], and [24], which can effectively address the challenge of selecting an appropriate subgradient at a nondifferentiable point. Furthermore, selecting a fixed subgradient is not optimal since this method does not guaranteed that the optimal solution is a stable equilibrium point.
3) Instead of using global parameters in the algorithms designing [20], [21], [23], [24], and [25], the parameters of our DIProx-NA are fully distributed. Additionally, we

also derive a sufficient condition for all the distributed parameters in DIProx-NA.

4) By means of Opial's lemma, we exhibit the trajectories of DIProx-NA asymptotically and weakly converge to the optimal solution set in Hilbert space. As far as we know, this is the first study that guarantees weak convergence of the trajectories derived from distributed neurodynamic approaches.

This article is constructed as follows. Necessary preliminaries are offered in Section II. In Section III, a distributed optimization problem of the problem (2) under directed graphs with weight-balanced is studied. In Section IV, the DIProx-NA is developed, and the existence, uniqueness, global convergence, and weak convergence of solutions are thoroughly discussed. Additionally, signal and image recovery examples are presented in Section V to showcase the effectiveness of DIProx-NA. Finally, this article concludes with a summary of the findings in Section VI.

Notations: For column vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, x^T is the transpose of x. For $x_1, \ldots, ., x_n \in \mathbb{R}^m$, $x = \text{col}(x_1, \ldots, x_n) \in \mathbb{R}^{nm}$ is the stacked column of them. $\otimes$ denotes the Kronecker product. $\mathrm{I}_{\mathbb{R}^m}$ is an indicator function, i.e., $\mathrm{I}_{\mathbb{R}^m}(u) = 0$, if $u \in \mathbb{R}^m$ and $\mathrm{I}_{\mathbb{R}^m}(u) = +\infty$, if $u \notin \mathbb{R}^m$. Denote $I_n \in \mathbb{R}^{n\times n}$ as a $n \times n$ identity matrix. $A \succeq 0$ ($A \succ 0$) represents that the symmetric matrix A is positive semidefinite (definite). $A \succeq B$, $A \succ B$ hold if $A - B \succeq 0$, $A - B \succ 0$ are satisfied, respectively. For matrices $A_1 \in \mathbb{R}^{p_1\times q_1}$ and $A_2 \in \mathbb{R}^{p_2\times q_2}$, $\text{bldiag}\{A_1, A_2\} \in \mathbb{R}^{(p_1+p_2)\times(q_1+q_2)}$ means the block diagonal matrix of A_1 and A_2; $\delta_{\max}(A_1)$ denotes the maximum singular (eigenvalue) value of A_1 ($p_1 = q_1$). $\bar{\beta} = \max_{1\leqslant i\leqslant p}\{\beta_i\}$, $\bar{\alpha} = \max_{1\leqslant i\leqslant p}\{\alpha_i\}$, $\bar{\theta} = \max_{1\leqslant i\leqslant p}\{\theta_i\}$, $\bar{\gamma} = \max_{1\leqslant i\leqslant p}\{\gamma_i\}$, $\underline{\beta} = \min_{1\leqslant i\leqslant p}\{\beta_i\}$, $\underline{\alpha} = \min_{1\leqslant i\leqslant p}\{\alpha_i\}$, $\underline{\theta} = \min_{1\leqslant i\leqslant p}\{\theta_i\}$, $\underline{\gamma} = \min_{1\leqslant i\leqslant p}\{\gamma_i\}$. $\boldsymbol{\gamma}_\mu$ denotes the parameter matrix $\boldsymbol{\gamma}$ of the multiplication with respect to variable μ. Define Id as an identity operator. For operator $H : \mathbb{R}^{2mn+p} \rightrightarrows \mathbb{R}^{2mn+p}$, we say H is nonexpansive if $\|H(z) - H(\hat{z})\| \leqslant \|z - \hat{z}\|$ $\forall z, \hat{z} \in \mathbb{R}^{2mn+p}$ holds, and is contractive if it satisfies $\|H(z) - H(\hat{z})\| < \|z - \hat{z}\|$ $\forall z, \hat{z} \in \mathbb{R}^{2mn+p}$. $\textbf{zer}\ H = \{z \in \mathbb{R}^{2mn+p} | 0 \in H(z)\}$. $\mathsf{L}^q([0, +\infty)), q > 0$ is a set of q-times Lebesgue integrable function on region $[0, +\infty)$. Note that the symbols 0 and 1, respectively, are a scalar or a vector, which can be obtained depending on the context of this article.

II. Preliminaries

A. Restricted Isometry Property

Definition 1 [36]: For any signal x with S-sparsity, we claim the matrix A fulfills RIP condition, if

$$(1 - \sigma_S)\|x\|^2 \leq \|A\ x\|^2 \leq (1 + \sigma_S)\|x\|^2$$

where $\sigma_S \in (0, 1)$.

B. Convex Functions and Subdifferential

Definition 2 [37]:

1) $\mathfrak{g} : \mathbb{R}^n \to \mathbb{R}$ is a nonsmooth and convex function, then it fulfills

$$\mathfrak{g}(\mathfrak{x}) \geqslant \mathfrak{g}(\mathfrak{y}) + \xi^T(\mathfrak{x} - \mathfrak{y})\ \forall\ \mathfrak{x}, \mathfrak{y} \in \mathbb{R}^n \tag{3}$$

where ξ means the subgradient of $\mathfrak{g}$ at $\mathfrak{y}$.

2) The subdifferential $\partial\mathfrak{g}(\mathfrak{x})$ is a set that encompasses all subgradients in the following way:

$$\partial\mathfrak{g}(\mathfrak{x}) = \left\{\xi \in \mathbb{R}^n | \mathfrak{g}(\mathfrak{x}) - \mathfrak{g}(\mathfrak{y}) \geqslant \xi^T(\mathfrak{x} - \mathfrak{y})\ \forall \mathfrak{x} \in \mathbb{R}^n\right\}. \tag{4}$$

C. Proximal Operator

Definition 3: If $\mathfrak{g}$ is proper, lower semi-continuous and convex, the proximal operator $\textbf{prox}_{\mathfrak{g}}(u)$ is denoted as

$$\begin{aligned}\textbf{prox}_{\mathfrak{g}}(u) &= \arg\min_{v\in\mathbb{R}^n}\left\{\mathfrak{g}(v) + \frac{1}{2}\|v - u\|^2\right\}\\ &= (I_n + \partial\mathfrak{g})^{-1}u\end{aligned} \tag{5}$$

where $\|\cdot\|$ is the Euclidean norm.

Lemma 1 [32]: Suppose $\mathfrak{g}$ is a proper, lower semi-continuous, and convex function. Then, the proximal operator $\text{prox}_{\mathfrak{g}}$ is a unique nonexpansive mapping. In simpler terms, for any $u, v \in \mathbb{R}^n$, we have the following inequality:

$$\left\|\textbf{prox}_{\mathfrak{g}}(u) - \textbf{prox}_{\mathfrak{g}}(v)\right\| \leqslant \|u - v\|. \tag{6}$$

Lemma 2 [32]: If $\mathfrak{g}(v) = \|v\|_1$, the proximal operator $\text{prox}_{\mathfrak{g}}(u)$ has a closed-form solution, i.e.,

$$\begin{aligned}\textbf{prox}_{\|\cdot\|_1}(u) &= \left[\textbf{prox}_{\|\cdot\|_1}(u_1), \ldots, \textbf{prox}_{\|\cdot\|_1}(u_n)\right]\\ \textbf{prox}_{\|\cdot\|_1}(u_i) &= \text{sign}(u_i)\max\{|u_i| - 1, 0\}.\end{aligned} \tag{7}$$

D. Algebraic Graph Theory

Define a directed graph of multiagents by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = (v_1, \ldots, v_n)$ denotes the agents or vertexes' sets, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the set of edges consisting of the communication among all agents, and $\mathcal{A} = \{\mathrm{a}_{ij}\} \in \mathbb{R}^{n\times n}$ denotes the adjacency matrix. An edge $e_{ij} \in \mathcal{E}$ indicates that agent i can get information from agent j. Note that $\mathrm{a}_{ij} > 0$ if $e_{ij} \in \mathcal{E}$; $\mathrm{a}_{ij} = 0$, otherwise. If agent i and agent j has a directed path, it is represented by a sequence of edges in the form $(i, i_1), (i_1, i_2) \cdots (i_k, j)$. A digraph is called strongly connected if there exists a directed path for any pair of different agents. The in-degree and out-degree of agent i are marked by $d_{\text{in}}(i) = \sum_{j\in\mathcal{N}_i}\mathrm{a}_{ij}$ and $d_{\text{out}}(i) = \sum_{j\in\mathcal{N}_i}\mathrm{a}_{ji}$, respectively, where $\mathcal{N}_i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$. For digraph $\mathcal{G}$, let $D_{\text{in}} = \text{diag}\{d_{\text{in}}(1), \ldots, d_{\text{in}}(n)\}$ be the in-degree matrix, then, its Laplacian matrix is $L_n = D_{\text{in}} - \mathcal{A}$, and it satisfies $L_n 1 = 0$. The digraph $\mathcal{G}$ is weight-balanced if and only if $d_{\text{in}}(i) = d_{\text{out}}(i)$, which directly implies $1^T L_n = 0$.

E. Auxiliary Lemmas of Convergence

The following lemmas play a critical part in proving the convergence of the trajectories of our proposed DIProx-NA in this article.

Lemma 3 [33]: Let $F(t), t \in [0, +\infty)$ be locally absolutely continuous and bounded below, then there is $\Psi(t) \in \mathsf{L}^1([0, +\infty))$, i.e., $\int_0^{+\infty}\Psi(t)dt < +\infty$, such that

$$\frac{d}{dt}F(t) \leqslant \Psi(t)\ \forall\ t \in [0, +\infty). \tag{8}$$

Subsequently, the $\lim_{t\to+\infty}F(t) \in \mathbb{R}$ exists.

Lemma 4 [33]: If $1 \leqslant q < \infty$, $1 \leqslant r \leqslant \infty$ and F is locally absolutely continuous, $F(t) \in \mathsf{L}^q([0, +\infty))$, $\Psi(t) \in \mathsf{L}^r([0, +\infty))$, and

$$\frac{d}{dt}F(t) \leqslant \Psi(t)\ \forall\ t \in [0, +\infty). \quad (9)$$

Then, there exists $\lim_{t\to+\infty} F(t) = 0$.

The following lemma is called as the continuous version of Opial's lemma.

Lemma 5 [34] (Opial's Lemma): Let $\mathbb{S} \subseteq \mathbb{R}^n$ be a nonempty set and $x(t) \in \mathbb{R}^n$ be a given operator. The following statements are true:

1) $\lim_{t\to+\infty} \|x(t) - x^*\|$ exists $\forall\ x^* \in \mathbb{S}$;
2) every weak sequential cluster point of the map x belongs to $\mathbb{S}$, then there exists $x_\infty \in \mathbb{S}$ such that $x(t)$ converges weakly to x_∞ as $t \to +\infty$.

Lemma 6 (Gronwall's Inequality): Let $\pi : [a, b] \to \mathbb{R}$ be continuous and $\varpi : [a, b] \to \mathbb{R}$ be continuous and nonnegative. If a continuous function $\mathfrak{x} : [a, b] \to \mathbb{R}$ satisfies

$$\mathfrak{x}(t) \leqslant \pi(t) + \int_a^t \varpi(s)\mathfrak{x}(s)ds \quad (10)$$

for $a \leqslant t \leqslant b$, then on the same interval

$$\mathfrak{x}(t) \leqslant \pi(t) + \int_a^t \pi(s)\varpi(s)e^{\int_s^t \varpi(\tau)d\tau}ds. \quad (11)$$

In particular, if $\pi(t) \equiv \pi$ is a constant, then, one has

$$\mathfrak{x}(t) \leqslant \pi e^{\int_s^t \varpi(\tau)d\tau}. \quad (12)$$

If, in addition, $\varpi(t) \equiv \varpi \geqslant 0$ is a constant, then

$$\mathfrak{x}(t) \leqslant \pi e^{\varpi(t-a)}. \quad (13)$$

III. Problem Transformation

It is noteworthy that the problem (2) does not adhere to the standard format of a distributed optimization problem. In order to tackle this problem using a distributed approach, we first transform it into a distributed version. In this section, we will discuss how to turn the problem (2) into a distributed constrained consensus problem under directed graphs.

Assumption 1: The communication graph $\mathcal{G}$ of multiagents is connected and weight-balanced.

Remark 1: Note that the Assumption 1 implies that $L_p 1 = 0$, $1^T L_p = 0$ of a p-agents network. Consequently, the condition $x_1 = x_2 = \cdots = x_p \in \mathbb{R}^n$ holds if and only if $Lx = L_p \otimes I_n x = 0$, where $x = \mathrm{col}(x_1, x_2, \ldots, x_p) \in \mathbb{R}^{pn}$.

Assumption 2 [28]: The matrix A has a full row-rank.

Taking the matrix A by rows, we split it as shown in Fig. 1, where $A_i \in \mathbb{R}^{m_i \times n}$ represents the ith row subblock in matrix A, then, one has $\sum_{i=1}^p m_i = m$. Furthermore, the observed value b can be decomposed as $b = (b_1^T, \ldots, b_p^T)^T \in \mathbb{R}^m$. By the properties of row splitting of matrix A and $\mathcal{G}$ in Assumption 1, the equation $Ax = b$ can be equivalent to

$$Ax = b, Lx = 0 \in \mathbb{R}^{pn} \quad (14)$$

where $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times pn}$, $x = \mathrm{col}(x_1, \ldots, x_p) \in \mathbb{R}^{pn}$ (see Fig. 2), $L = L_p \otimes I_n$, and $L_p \in \mathbb{R}^{p\times p}$.

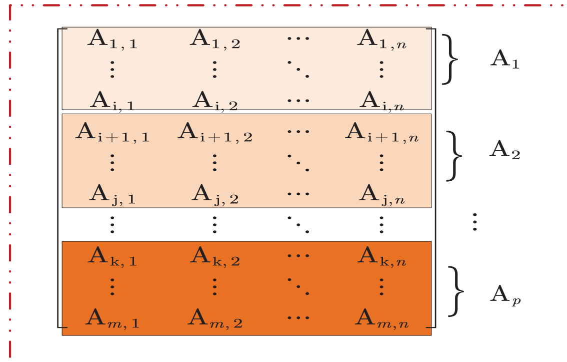

Fig. 1. A is divided into p blocks by rows.

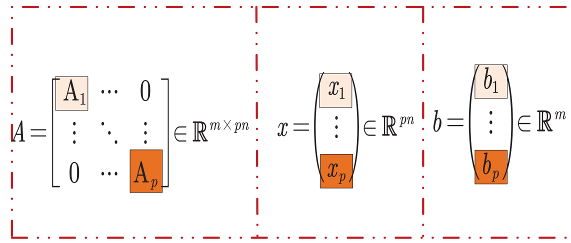

Fig. 2. Matrix A, x, and b.

Lemma 7: Assume that Assumptions 1 and 2 are fulfilled, according to the condition (14), problem (2) can be equivalently formulated as the following distributed constrained optimization problem:

$$\min_{x\in\mathbb{R}^{pn}} \|x\|_1, \text{ s.t. } Ax = b, Lx = 0. \quad (15)$$

Proof: The proof is inspired by Theorem 3 in [28]. Note that the directed graph $\mathcal{G}$ is weight-balanced, then, $1^T L = 0$ and $L1 = 0$. Define $\bar{x} = 1 \otimes x \in R^{pn}$ and $Ax = b$, one has

$$\begin{aligned} L\bar{x} &= (L_p \otimes I_n)\bar{x} = (L_p \otimes I_n)(1 \otimes x) \\ &= (L_p 1) \otimes (I_n x) = 0, \\ A\bar{x} &= \sum_{k=1}^p A_{mk}\bar{x}_k = \sum_{k=1}^p A_{mk}x = Ax = b. \end{aligned}$$

Conversely, based on the properties of Kronecker product, it has that $L\bar{x} = (L_p \otimes I_n)\bar{x} = (L_p \otimes I_n)vec(\mathcal{X}) = vec(I_n \mathcal{X} L_p)$ with $vec(ABC) = (C^T \otimes A)vec(B)$, $\mathcal{X} = (x_1, \ldots, x_p) \in \mathbb{R}^{n\times p}$ and $\bar{x} = vec(\mathcal{X})$. Since $L\bar{x} = 0$, such that $I_n \mathcal{X} L_p = 0$ and $L_p \mathcal{X}^T = 0$, and 1 is the right eigenvector of L_p which corresponds to a unique zero eigenvalue due to $\mathcal{G}$ is weight-balanced. Therefore, $\mathcal{X}^T = \alpha^T \otimes 1$, where $\varrho = (\varrho_1, \ldots, \varrho_n)^T \in \mathbb{R}^n$ is a column vector. It follows that $\mathcal{X} = \varrho \otimes 1^T$. Then each column vector $\bar{x}_i$ of $\bar{x}$ satisfies $\bar{x}_i = \varrho$. Therefore, we have $\bar{x}_i = \alpha$. Therefore, the proof is thereby completed. ■

IV. Distributed Inertial Proximal Neurodynamic Approach for Problem (15) Under Directed Graphs

To tackle the problem (15) in a fully distributed manner under directed graphs, we present a DIProx-NA as follows.

For agent i, $(i = 1, \ldots, p)$, the DIProx-NA is

$$\begin{cases} \ddot{x}_i(t) + \gamma_i \dot{x}_i(t) + \theta_i\Big(x_i(t) - \mathbf{Prox}_{(\|\cdot\|_1)_i}\Big(x_i(t) \\ \quad -\alpha_i \sum_{j\in\mathcal{N}_i} a_{ij}\big(x_i(t) - x_j(t)\big) - \beta_i A_i^T(A_i x_i(t) - b_i) \\ \quad - A_i^T \lambda_i(t) - \sum_{j\in\mathcal{N}_i} a_{ji}\big(\mu_i(t) - \mu_j(t)\big)\Big)\Big) = 0 \\ \ddot{\lambda}_i(t) + \gamma_i \dot{\lambda}_i(t) - \theta_i(A_i x_i - b_i) = 0\} \\ \ddot{\mu}_i(t) + \gamma_i \dot{\mu}_i(t) - \theta_i \sum_{j\in\mathcal{N}_i} a_{ij}\big(x_i(t) - x_j(t)\big) = 0 \end{cases} \tag{16}$$

where $x_i \in \mathbb{R}^n$ is the primal variable and $\lambda_i \in \mathbb{R}^{m_i}$, $\mu_i \in \mathbb{R}^n$ are dual (Lagrange multiplier) variables, and γ_i, θ_i, α_i, and $\beta_i > 0$, and how to choose them is discussed in detail later (see Theorem 3).

The compact form of DIProx-NA is

$$\begin{cases} \ddot{x}(t) + \boldsymbol{\gamma}\dot{x}(t) + \boldsymbol{\theta}\big(x(t) - \mathbf{Prox}_{\|\cdot\|_1}(x(t) - \boldsymbol{\alpha} L x(t) \\ \quad - \boldsymbol{\beta} A^T(Ax(t) - b) - A^T\lambda(t) - L^T\mu(t))\big) = 0 \\ \ddot{\lambda}(t) + \boldsymbol{\gamma}\dot{\lambda}(t) = \boldsymbol{\theta}(Ax(t) - b) \\ \ddot{\mu}(t) + \boldsymbol{\gamma}\dot{\mu}(t) = \boldsymbol{\theta} L x(t) \end{cases} \tag{17}$$

where $x \in \mathbb{R}^{pn}$ is the primal variable and $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^{pn}$ are dual (Lagrange multiplier) variables, $\boldsymbol{\gamma} = \text{bldiag}\{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p\} \in \mathbb{R}^{pn\times pn}$, $\boldsymbol{\theta} = \text{bldiag}\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p\} \in \mathbb{R}^{pn\times pn}$. $\boldsymbol{\gamma}_i = \gamma_i I_n$ and $\boldsymbol{\theta}_i = \theta_i I_n$ if γ_i and θ_i are the multipliers of the variables $x(t)$ and $\mu(t)$. Moreover, $\boldsymbol{\gamma}_i = \gamma_i I_{m_i}$ and $\boldsymbol{\theta}_i = \theta_i I_{m_i}$ if bothγ_i and θ_i are the multipliers of the variable $\lambda(t)$, $\boldsymbol{\alpha} = \text{bldiag}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p\} \in \mathbb{R}^{pn\times pn}$, $\boldsymbol{\alpha}_i = \alpha_i I_{m_i}$, $\boldsymbol{\beta} = \text{bldiag}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p\} \in \mathbb{R}^{pn\times pn}$ and $\boldsymbol{\beta}_i = \beta_i I_n$, $i = 1, \ldots, p$.

In the following theorem a relation is provided between the optimal solution of the problem (15) and the equilibrium point of DIProx-NA (16).

Remark 2: In the problem (2) or problem (13), the L_1-norm (objective function) is convex but nonsmooth, the subgradient-based method may be a good choice, however it suffers from the difficulty of subgradient selection at the nondifferentiable point, i.e., there exists no exact single-valued gradient of L_1-norm at 0.

For example:

1) let $A = [1, (3/4)]$, $b = 3$, the optimal sparse solution is $x^* = (3, 0)^T$. According to the KKT condition of problem (2), one has $\partial|x_1| + \lambda = 0$; $\partial|x_2| + (3/4)\lambda = 0$; $(x_1 + (3/4)x_2) = 3$, which become $\partial|x_1^*| = -\lambda = 1$, $\partial|x_2^*| = -(3/4)\lambda = (3/4)$, (i.e., we should choose a suitable subgradient $\partial|x_2| = (3/4)$) at the optimal solution $x^* = (3, 0)^T$ and
2) let $A = [1, (3/5)]$, $b = 3$, we should choose a suitable subgradient $\partial|x_2| = (3/5)$ at the optimal solution $x^* = (3, 0)^T$.

Similarly, in the case where an optimal point contains multiple zero elements, using a fixed subgradient is not suitable.

To overcome this issue of fixed subgradient selection, an approximation operator of L_1-norm, i.e., soft threshold is used to adaptive capture the subgradient of L_1-norm in DIProx-NA (15), which can effectively avoid the subgradient selection problem.

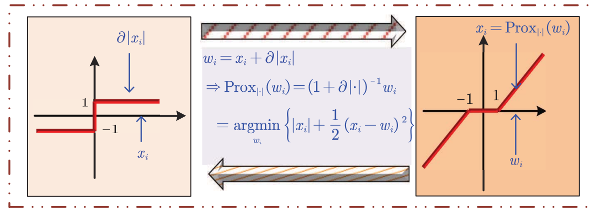

Fig. 3. Relationship between subgradient $\partial|x_i|$ and proximal operator $\text{Prox}_{|\cdot|}(w_i)$.

Denote $\partial\|x\|_1 = (\partial|x_1|, \ldots, \partial|x_{pn}|)^T$, where

$$\partial\|x_i\|_1 = \begin{cases} 1, & x_i > 0 \\ [-1, 1], & x_i = 0, i = 1, \ldots, pn \\ -1, & x_i < 0. \end{cases}$$

Let $w_i = x_i + \partial\|x_i\|_1$, i.e., $w_i = \begin{cases} x_i + 1, & x_i > 0 \\ [-1, 1], & x_i = 0, \\ x_i - 1, & x_i < 0 \end{cases}$, thus, we can obtain that $x_i = \begin{cases} w_i - 1, w_i > 1 \\ 0, w \in [-1, 1] \\ w_i + 1, w_i < -1 \end{cases}$ (see Fig. 3).

Note that $x_i = \text{Prox}_{|\cdot|}(w_i) = \arg\min_{w_i}\{|x_i| + (1/2)(x_i - w_i)^2\} = (1 + \partial\|\cdot\|_1)^{-1} w_i$ adaptively updates according to the variable w_i. Hence, it is possible to address the issue of subgradient selection with the use of the projection operator, i.e., our proposed DIProx-NA is suitable for the considered problem (13).

Theorem 1: If Assumptions 1 and 2 hold and $x^* \in \mathbb{R}^{pn}$ be the optimal solution for the problem (15) if and only if there exists $\mu^* \in \mathbb{R}^{pn}$ and $\lambda^* \in \mathbb{R}^{pn}$, such that $\text{col}(x^*, \lambda^*, \mu^*)$ is the equilibrium point of DIProx-NA (16).

Proof: *Necessity*: x^* is optimal of problem (15) from the Karush–Kuhn–Tucker (KKT) conditions that if and only if there exist λ^* and μ^* that fulfill the following conditions:

$$\partial\left\|x^*\right\|_1 + A^T\lambda^* + L^T\mu^* \ni 0 \tag{18a}$$
$$Ax^* = b, Lx^* = 0. \tag{18b}$$

Add x^* to both sides of (18a) and arrange it, one has

$$x^* + \partial\left\|x^*\right\|_1 \ni x^* - A^T\lambda^* - L^T\mu^*. \tag{19}$$

The conditions $Ax^* = b$ (i.e., $A^T(Ax^* - b) = 0$), $Lx^* = 0$ and $\boldsymbol{\alpha}$, $\boldsymbol{\beta} \succ 0$ imply that $\boldsymbol{\alpha} Lx^* + \boldsymbol{\beta} A^T(Ax^* - b) = 0$, which is added in the right-hand of (19) and yields

$$\begin{aligned} x^* + \partial\left\|x^*\right\|_1 &\ni x^* - A^T\lambda^* - L^T\mu^* \\ &\quad -\boldsymbol{\alpha} Lx^* + \boldsymbol{\beta} A^T\big(Ax^* - b\big) \end{aligned} \tag{20}$$

then, by the aid of condition (5), (19) becomes

$$\begin{aligned} x^* = \mathbf{prox}_{\|\cdot\|_1}\big(&x^* - A^T\lambda^* - L^T\mu^* \\ &- \boldsymbol{\alpha} Lx^* - \boldsymbol{\beta} A^T\big(Ax^* - b\big)\big). \end{aligned} \tag{21}$$

Combining (18b) with (21), one obtains that the optimal solution of problem (15) is an equilibrium of (16), i.e., (17).

Sufficiency: Let $\mathrm{col}(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^{2pn+m}$ is an equilibrium point of (16), i.e., (17), then $\mathrm{col}(\bar{x}, \bar{\lambda}, \bar{\mu})$ satisfies

$$\begin{aligned}\bar{x} = \mathbf{prox}_{\|\cdot\|_1}(\bar{x} - A^T\bar{\lambda} - L^T\bar{\mu}\\ -\boldsymbol{\alpha}L\bar{x} - \boldsymbol{\beta}A^T(A\bar{x} - b)) \quad &(22a)\\ L\bar{x} = 0, A\bar{x} = b. \quad &(22b)\end{aligned}$$

From (5) and, (21) can be written as

$$\bar{x} + \partial\|\bar{x}\|_1 \ni \bar{x} - A^T\bar{\lambda} - L^T\bar{\mu} - \boldsymbol{\alpha}L\bar{x} + \boldsymbol{\beta}A^T(A\bar{x} - b).$$

Note that, the conditions $L\bar{x} = 0$, $A\bar{x} = b$ in (22b) and $\boldsymbol{\alpha} \succ 0$, $\boldsymbol{\beta} \succ 0$ imply $\boldsymbol{\alpha}L\bar{x} = 0$, $\boldsymbol{\beta}A^T(A\bar{x} - b) = 0$. Therefore, (22) can be reduced to

$$\begin{aligned}\partial\|\bar{x}\|_1 + A^T\bar{\lambda} + L^T\bar{\mu} \ni 0\\ L\bar{x} = 0, A\bar{x} = b. \quad (23)\end{aligned}$$

which coincides with the KKT conditions of problem (15), thus the equilibrium point $\mathrm{col}(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^{2pn+m}$ is the optimal solution of the optimization problem (15) due to the convexity of problem (15). Therefore, the *Sufficiency* holds. In summary, the proof is completed, i.e., Theorem 1 holds. ■

Remark 3: The choice of parameter for the soft thresholding depends on the coefficients in front of the objective function $\|x\|_1$ in (15). It is worth noting that multiplication of a non-negative constant in front of the objective function $\|x\|_1$ in optimization problem (15) does not affect the its optimal solution, that is, the soft threshold can be chosen as $\rho > 0$, i.e., the objective function in optimization problem (15) is multiplied in front by a parameter $\rho > 0$. However, when a specific parameter $\rho > 0$ is chosen as the soft threshold, each agent must adopt the global parameter $\rho > 0$ as the threshold for computation. This destroys the fully distributed characteristic of the neurodynamic approach.. In this article, there is no specific parameter set in front of the objective function $\|x\|_1$ in the optimization problem (15) (the default is 1), accordingly every agent can choose its own soft threshold parameter 1 according to its own objective function (without a specific setting of the global parameter), and there is no global parameter used, so it can be effective to realize the fully DIProx-NA.

Next, we show the convergence analysis of DIProx-NA (17). However, before performing convergence analysis of DIProx-NA (17), we need to give two important lemmas that act as an key role in convergence analysis.

Let $z = \mathrm{col}(x, \lambda, \mu) \in \mathbb{R}^{2pn+m}$, then, the DIProx-NA (17) is equivalent to

$$\ddot{z}(t) + \boldsymbol{\Gamma}\dot{z}(t) + \boldsymbol{\Theta}\big(z(t) - \mathbf{Prox}_{\mathfrak{f}}(z(t) - G(z(t)))\big) = 0$$

where

$$\mathbf{Prox}_{\mathfrak{f}} = \begin{bmatrix}\mathbf{Prox}_{\|\cdot\|_1}\\ \mathbf{Prox}_{\mathrm{I}_{\mathbb{R}^m}}\\ \mathbf{Prox}_{\mathrm{I}_{\mathbb{R}^{pn}}}\end{bmatrix} : \mathbb{R}^{2pn+m} \to \mathbb{R}^{2pn+m}$$

$$G(z(t)) = \begin{pmatrix}\boldsymbol{\alpha}Lx(t) + \boldsymbol{\beta}A^T(Ax(t) - b) + A^T\lambda(t) + L^T\mu(t)\\ -(Ax(t) - b)\\ -Lx(t)\end{pmatrix}.$$

$\boldsymbol{\Gamma} = \mathrm{bldiag}(\boldsymbol{\gamma}_x, \boldsymbol{\gamma}_\lambda, \boldsymbol{\gamma}_\mu) \in \mathbb{R}^{(2pn+m)\times(2pn+m)}$ and $\boldsymbol{\Theta} = \mathrm{bldiag}(\boldsymbol{\theta}_x, \boldsymbol{\theta}_\lambda, \boldsymbol{\theta}_\mu) \in \mathbb{R}^{(2pn+m)\times(2pn+m)}$.

Furthermore, defining two operators $T(z) = \mathbf{Prox}_{\mathfrak{f}}(z - G(z))$ and $H(z) = z - T(z)$. Then, the DIProx-NA (17) becomes

$$\ddot{z}(t) + \boldsymbol{\Gamma}\dot{z}(t) + \boldsymbol{\Theta}(H(z(t))) = 0. \quad (24)$$

Lemma 8: If $2\delta_{\max}(A) + \bar{\beta}\delta_{\max}(A^TA) + \sqrt{(\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL)} \in (0, 2]$, i.e., [(34) and (35) hold], then the operator T is nonexpansive.

Proof: Because the proximal operator $\mathbf{Prox}_{\mathfrak{f}}$ in Lemma 1 is nonexpansive, one has

$$\begin{aligned}&\|T(z) - T(\hat{z})\|^2\\ &= \|\mathbf{Prox}_{\mathfrak{f}}(z - G(z)) - \mathbf{Prox}_{\mathfrak{f}}(\hat{z} - G(\hat{z}))\|^2\\ &\leqslant \|z - G(z) - \hat{z} + G(\hat{z})\|^2\\ &= \|z - \hat{z}\|^2 + \|G(\hat{z}) - G(z)\|^2\\ &\quad + 2(z - \hat{z})^T(G(\hat{z}) - G(z)). \quad (25)\end{aligned}$$

For $\|G(\hat{z}) - G(z)\|$, we have

$$\begin{aligned}&\|G(\hat{z}) - G(z)\|\\ &= \left\|\begin{matrix}\boldsymbol{\alpha}L(x - \hat{x}) + \boldsymbol{\beta}A^TA(x - \hat{x})\\ -A(x - \hat{x})\\ -L(x - \hat{x})\end{matrix}\right.\\ &\qquad \left.\begin{matrix}+A^T(\lambda - \hat{\lambda}) + L^T(\mu - \hat{\mu})\\ 0\\ 0\end{matrix}\right\|\\ &\leqslant \left\|\begin{matrix}A^T(\lambda - \hat{\lambda}) + \boldsymbol{\beta}A^TA(x - \hat{x})\\ -A(x - \hat{x})\\ 0\end{matrix}\right\| \|z - \hat{z}\|\\ &\quad + \left\|\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L^T\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}\right\| \|z - \hat{z}\|\\ &\leqslant \left(2\delta_{\max}(A) + \bar{\beta}\delta_{\max}(A^TA) + \left\|\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L^T\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}\right\|\right)\\ &\quad \times \|z - \hat{z}\|. \quad (26)\end{aligned}$$

In addition, for $\left\|\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L^T\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}\right\|$, one has

$$\begin{aligned}\left\|\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L^T\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}\right\|^2 &= \left\|\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L^T\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}^T\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L^T\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}\right\|\\ &= \left\|\begin{bmatrix}L^T(\boldsymbol{\alpha}^2)L^T - L^2 & 0 & L^T\boldsymbol{\alpha}L^T\\ 0 & 0 & 0\\ -L^T\boldsymbol{\alpha}L & 0 & -(L^T)^2\end{bmatrix}\right\|\\ &\leqslant (\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL) \quad (27)\end{aligned}$$

which further implies

$$\left\|\begin{bmatrix}\boldsymbol{\alpha}L & 0 & L\\ 0 & 0 & 0\\ -L & 0 & 0\end{bmatrix}\right\| \leqslant \sqrt{(\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL)}. \quad (28)$$

Substituting (28) in (26), one obtains

$$\|G(z) - G(\hat{z})\| \leqslant \Big(2\delta_{\max}(A) + \bar{\beta}\delta_{\max}(A^TA) + \sqrt{(\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL)}\Big)\|z - \hat{z}\| \tag{29}$$

which implies that $\|G(z) - G(\hat{z})\|$ is Lipschitz continuous.

Note that $G(z)$ is monotonic of z since it fulfills

$$(G(z) - G(\hat{z}))^T(z - \hat{z}) = (x - \hat{x})^T\boldsymbol{\alpha}L(x - \hat{x}) + (x - \hat{x})^T A^T A(x - \hat{x}) \geqslant 0.$$

Based on Baillon–Haddad lemma [35], one has

$$\langle \hat{z} - z, G(\hat{z}) - G(z)\rangle \geqslant \frac{1}{2\delta_{\max}(A) + \bar{\beta}\delta_{\max}(A^TA) + \sqrt{(\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL)}} \times \|G(\hat{z}) - G(z)\|^2. \tag{30}$$

Moreover, note that if $2\delta_{\max}(A) + \bar{\beta}\delta_{\max}(A^TA) + \sqrt{(\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL)} \in (0, 2]$, one has

$$\|G(z) - G(\hat{z})\|^2 + 2\langle z - \hat{z}, G(\hat{z}) - G(z)\rangle \leqslant 0. \tag{31}$$

Inserting (31) into (25), we obtain

$$\|T(z) - T(\hat{z})\| \leqslant \|z - \hat{z}\| \tag{32}$$

i.e., the operator T is nonexpansive. ■

Remark 4: The operator T is nonexpansive and dependent on the parameters $\bar{\alpha}$, $\bar{\beta}$ and $\delta_{\max}(A^TA)$. Applying the basic matrix parametric inequality, we deduce that

$$\delta_{\max}(L^TL) = \|L\|^2 \leqslant \|L\|_1\|L\|_\infty \leqslant \left(\max_i \sum_{j=1}^n |a_{ji}|\right)\left(\max_i \sum_{j=1}^n |a_{ij}|\right) = 4(d_{\text{in}}^{\max})^2 \tag{33}$$

where $d_{\text{in}}^{\max} = \max_{i\in\mathcal{V}} d_{\text{in}}^{\max}(i)$ and the last equality holds due to $d_{\text{in}}^{\max} = d_{\text{out}}^{\max}$.

For parameters $\bar{\alpha}$ and $\bar{\beta}$. According to

$$(\bar{\alpha}^2 + \bar{\alpha} + 1)\delta_{\max}(L^TL) \leqslant (2 - 2\delta_{\max}(A) - \bar{\beta}\delta_{\max}(A^TA))^2$$

one has

$$0 < \bar{\alpha} \leqslant \frac{\sqrt{\frac{4(2-2\delta_{\max}(A)-\bar{\beta}\delta_{\max}(A^TA))^2}{\delta_{\max}(L^TL)} - 3} - 1}{2} \tag{34}$$

when the $\bar{\beta}$ satisfies

$$\left[\frac{2}{\delta_{\max}(A^TA)} - \frac{2\delta_{\max}(A)}{\delta_{\max}(A^TA)} + \frac{\sqrt{\frac{1}{\delta_{\max}(L^TL)}}}{\delta_{\max}(A^TA)}\right]^+ < \bar{\beta}. \tag{35}$$

Combining (33)–(35), we have the conclusion that if parameters $\bar{\alpha}$, $\bar{\beta}$ satisfy

$$\bar{\beta} \in \left(\left[\frac{2}{\delta_{\max}(A^TA)} - \frac{2\delta_{\max}(A)}{\delta_{\max}(A^TA)} + \frac{1}{2(d_{\text{in}}^{\max})\delta_{\max}(A^TA)}\right]^+, +\infty\right)$$
$$\bar{\alpha} \in \left(0, \frac{\sqrt{\frac{(2-2\delta_{\max}(A)-\bar{\beta}\delta_{\max}(A^TA))^2}{(d_{\text{in}}^{\max})^2} - 3} - 1}{2}\right) \tag{36}$$

the operator T is contractive, thus, T is nonexpansive.

Lemma 9: If operator T is nonexpansive, then operator $H = \text{Id} - T$ is maximal monotone and (1/2)-coercive, i.e.,

$$(H(z) - H(\hat{z}))^T(z - \hat{z}) \geqslant \frac{1}{2}\|H(z) - H(\hat{z})\|^2 \ \forall z, \hat{z} \in \mathbb{R}^{2pn+m} \tag{37}$$

where (37) implies operator H is 2-Lipschitz.

Proof: For any z, $\hat{z} \in \mathbb{R}^{2pn+m}$, by a direct calculation, we get

$$\begin{aligned}(H(z) - H(\hat{z}))^T(z - \hat{z}) - \frac{1}{2}\|H(z) - H(\hat{z})\|^2 &= (z - T(z) - \hat{z} + T(\hat{z}))^T(z - \hat{z}) - \frac{1}{2}\|z - \hat{z}\|^2 \\ &\quad - \frac{1}{2}\|T(z) - T(\hat{z})\|^2 + (z - \hat{z})^T(T(z) - T(\hat{z})) \\ &= \frac{1}{2}\|z - \hat{z}\|^2 - \frac{1}{2}\|T(z) - T(\hat{z})\|^2 \geqslant 0\end{aligned} \tag{38}$$

where the last inequality is satisfied from Lemma 7. ■

Theorem 2: There is a unique solution of $z(t)$ of DIProx-NA (24) with any initial values $z_0 \in \mathbb{R}^{2pn+m}$ and $\dot{z}_0 \in \mathbb{R}^{2pn+m}$.

Proof: Note that DIProx-NA (24) could be equivalently converted to the following first-order dynamical system:

$$\begin{aligned}\dot{\mathsf{z}}(t) &= \Lambda(\mathsf{z}(t)) \\ \Lambda(\mathsf{z}(t)) &= \begin{pmatrix} W(t) \\ \boldsymbol{\Gamma}W(t) + \boldsymbol{\Theta}(H(z(t)))\end{pmatrix}.\end{aligned} \tag{39}$$

Let $\mathsf{z}^1 = (z^1, W^1)$, $\mathsf{z}^2 = (z^2, W^2) \in \mathbb{R}^{4pn+2m}$ be two solutions of DIProx-NA (39) with the same initial values $z_0 \in \mathbb{R}^{2pn+m}$. Assume that there exist $\tilde{t} > 0$ and $\mathsf{z}^1(t) \neq \mathsf{z}^2(t)$, then $\mathsf{z}^1(t) \neq \mathsf{z}^2(t)$ for $\forall t \in [\tilde{t}, \tilde{t} + \upsilon]$ with $\upsilon > 0$. And denote

$$\Lambda(\mathsf{z}^\mathfrak{k}(t)) = \begin{pmatrix} W^\mathfrak{k}(t) \\ \boldsymbol{\Gamma}W^\mathfrak{k}(t) + \boldsymbol{\Theta}(H(z^\mathfrak{k}(t)))\end{pmatrix} \ \forall \mathfrak{k} = 1, 2.$$

First, for $\forall\ t \in [0, \tilde{t} + \upsilon]$, we show

$$\begin{aligned}\|\Lambda(\mathsf{z}^1(t)) - \Lambda(\mathsf{z}^2(t))\| &\leqslant \|W^1(t) - W^2(t)\| + \|\boldsymbol{\Gamma}W^1(t) - \boldsymbol{\Gamma}W^2(t)\| \\ &\quad + \|\boldsymbol{\Theta}(H(z^1(t))) - \boldsymbol{\Theta}(H(z^2(t)))\| \\ &\leqslant (1 + \bar{\gamma} + 2\bar{\theta})\|\mathsf{z}^1(t) - \mathsf{z}^2(t)\|.\end{aligned}$$

Setting $V(t) = (1/2)\|\mathsf{z}^1(t) - \mathsf{z}^2(t)\|^2$, then, it yields

$$\begin{aligned}\dot{V}(t) &= (\mathsf{z}^1(t) - \mathsf{z}^2(t))^T(\Lambda(\mathsf{z}^1(t)) - \Lambda(\mathsf{z}^2(t))) \\ &\leq (1 + \bar{\gamma} + 2\bar{\theta})\|\mathsf{z}^1(t) - \mathsf{z}^2(t)\|^2\end{aligned} \tag{40}$$

then, we get $\mathsf{z}^1(t) = \mathsf{z}^2(t)$ for $\forall t \in [0, \tilde{t} + \upsilon]$. Integrating the above inequality (40) from 0 to $t (\leqslant \tilde{t} + \upsilon)$, we have $\mathsf{z}^1(t) = \mathsf{z}^2(t)$ for $\forall\ t \in [0, \tilde{t} + \upsilon]$ by using Gronwall's inequality in Lemma 6. Therefore, a unique solution $z(t)$ exists, for $\forall\ t \in [0, \tilde{t} + \upsilon]$. Furthermore, from the theorem on the extension of a solution, $\mathsf{z}(t)$ is exists for $t \in [0, +\infty]$, thus, the proof is completed. ■

Theorem 3: If

$$\begin{aligned} &1):0 < \underline{\gamma} \leqslant \underset{1 \leqslant i \leqslant p}{\gamma_i} \leqslant \bar{\gamma} \\ &2):0 < \underline{\theta} \leqslant \underset{1 \leqslant i \leqslant p}{\theta_i} \leqslant \bar{\theta} \\ &3):\underline{\gamma} \geqslant \sqrt{\frac{2\bar{\theta}^2(1+\omega)}{\underline{\theta}}}, \omega > 0 \end{aligned} \tag{41}$$

holds, then one has:

1) $x(t), \lambda(t), \mu(t)$ generated by DIProx-NA (17) are bounded and $\dot{z}(t), \ddot{z}(t)$ and $H(z(t)) \in \mathsf{L}^2([t_0, +\infty))$;
2) $\lim_{t\to+\infty}\ddot{z}(t) = 0$, $\lim_{t\to+\infty}\dot{z}(t) = 0$, and $\lim_{t\to+\infty}H(z(t)) = 0$, which means DIProx-NA (17) is convergent.

Proof: 1) Let $z^* \in \mathbb{R}^{2pn+m}$ be the optimal solution of DIProx-NA (17), i.e., $0 \in Zer H$. Set $h_k(t) = (1/2)\|z_k(t) - z_k{}^*\|^2$, $h(t) = \mathrm{col}(h_1(t), \ldots, h_{2pn+m}(t))$, we have

$$\begin{aligned} \dot{h}_k(t) &= \left(z_k(t) - z_k{}^*\right)^T \dot{z}_k(t) \\ \ddot{h}_k(t) &= \|\dot{z}_k(t)\|^2 + \left(z_k(t) - z_k{}^*\right)^T \ddot{z}_k(t). \end{aligned} \tag{42}$$

It follows that, for any $k = 1, \ldots, 2pn + m$, one has:

$$\begin{aligned} &\ddot{h}_k(t) + \gamma_k(t)\dot{h}_k(t) \\ &\quad = \|\dot{z}_k(t)\|^2 + \left(z_k(t) - z_i{}^*\right)^T (\ddot{z}_k(t) + \gamma_k \dot{z}_k(t)). \end{aligned} \tag{43}$$

By summing (43) from $k = 1$ to $k = 2pn+m$ and combining with DIProx-NA (17), we have

$$1^T\ddot{h}(t) + 1^T\boldsymbol{\Gamma}(t)\dot{h}(t) + \left(z(t) - z^*\right)^T \boldsymbol{\Theta}(H(z(t))) = \|\dot{z}(t)\|^2. \tag{44}$$

Since $H(z^*) = 0$, then (43) becomes

$$\begin{aligned} &1^T\ddot{h}(t) + 1^T\boldsymbol{\Gamma}\dot{h}(t) + \left(z(t) - z^*\right)^T \boldsymbol{\Theta}\left(H(z(t)) - H\left(z^*\right)\right) \\ &\quad = \|\dot{z}(t)\|^2. \end{aligned} \tag{45}$$

According to Lemma 8 and $\boldsymbol{\Theta} \succeq \underline{\theta} I_{2np+m}$, we have $(z(t) - z^*)^T \boldsymbol{\Theta}(H(z(t)) - H(z^*)) \geqslant (\underline{\theta}/2)\|H(z(t))\|^2$, with $H(z^*) = 0$ it yields

$$1^T\ddot{h}(t) + 1^T\boldsymbol{\Gamma}(t)\dot{h}(t) + \tfrac{\underline{\theta}}{2}\|H(z(t))\|^2 \leqslant \|\dot{z}(t)\|^2. \tag{46}$$

From (24), (46) reduces to

$$1^T\ddot{h}(t) + 1^T\boldsymbol{\Gamma}\dot{h}(t) + \frac{\underline{\theta}}{2}\left\|\boldsymbol{\Theta}^{-1}(\ddot{z}(t) + \boldsymbol{\Gamma}\dot{z}(t))\right\|^2 \leqslant \|\dot{z}(t)\|^2 \tag{47}$$

and it implies

$$\begin{aligned} &1^T\ddot{h}(t) + 1^T\boldsymbol{\Gamma}\dot{h}(t) + \frac{\underline{\theta}}{2\bar{\theta}^2}\|\ddot{z}(t)\|^2 \\ &\quad + \left(\frac{\underline{\gamma}^2\underline{\theta}}{2\bar{\theta}^2} - 1\right)\|\dot{z}(t)\|^2 + \frac{\underline{\theta}}{\bar{\theta}^2}\ddot{z}(t)^T(\boldsymbol{\Gamma}\dot{z}(t)) \leqslant 0. \end{aligned} \tag{48}$$

Combining (48) with the following two conditions:

$$\ddot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t) = \frac{1}{2}\frac{d}{dt}\left(\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t)\right) \tag{49}$$

and

$$\frac{d}{dt}\left(1^T\boldsymbol{\Gamma}h(t)\right) = 1^T\boldsymbol{\Gamma}\dot{h}(t) \tag{50}$$

we get

$$\begin{aligned} &1^T\ddot{h}(t) + \frac{d}{dt}\left(1^T\boldsymbol{\Gamma}h(t)\right) + \frac{\underline{\theta}}{2\bar{\theta}^2}\|\ddot{z}(t)\|^2 \\ &\quad + \left(\frac{\underline{\gamma}^2\underline{\theta}}{2\bar{\theta}^2} - 1\right)\|\dot{z}(t)\|^2 + \frac{\underline{\theta}}{2\bar{\theta}^2}\frac{d}{dt}\left(\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t)\right) \leqslant 0. \end{aligned} \tag{51}$$

Since $\underline{\gamma} \geqslant \sqrt{[(2\bar{\theta}^2(1+\omega))/(\underline{\theta})]}$ with $\omega > 0$, i.e.,

$$\frac{\underline{\gamma}^2\underline{\theta}}{2\bar{\theta}^2} - 1 \geqslant \omega > 0. \tag{52}$$

Therefore, the inequality (51) yields

$$1^T\ddot{h}(t) + \frac{d}{dt}\left(1^T\boldsymbol{\Gamma}h(t)\right) + \frac{\underline{\theta}}{2\bar{\theta}^2}\frac{d}{dt}\left(\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t)\right) \leqslant 0 \tag{53}$$

which indicates that the function $1^T\dot{h}(t) + 1^T\boldsymbol{\Gamma}h(t) + (\underline{\theta}/2\bar{\theta}^2)\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t)$ is monotonically decreasing. Therefore, there is a positive constant M, such that

$$\begin{aligned} &1^T\dot{h}(t) + 1^T\boldsymbol{\Gamma}h(t) + \frac{\underline{\theta}}{2\bar{\theta}^2}\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t) \\ &\quad \leqslant 1^T\dot{h}(t_0) + 1^T\boldsymbol{\Gamma}h(t_0) + \frac{\underline{\theta}}{2\bar{\theta}^2}\dot{z}(t_0)^T\boldsymbol{\Gamma}\dot{z}(t_0) = M \end{aligned} \tag{54}$$

which further implies

$$1^T\dot{h}(t) + \underline{\gamma}1^T h(t) \leqslant M \tag{55}$$

since $1^T\dot{h}(t) + \underline{\gamma}1^T h(t) \leqslant 1^T\dot{h}(t) + 1^T\boldsymbol{\Gamma}h(t)$.

By multiplying (55) with $e^{\underline{\gamma}t}$ and then integrating it from $t_0 \in [0, t)$ to t, we get

$$\begin{aligned} 1^T h(t) &\leqslant \frac{\left(1 - e^{\underline{\gamma}\left(\frac{t_0}{t}\right)}\right)M}{\underline{\gamma}} + e^{\underline{\gamma}t_0}1^T h(t_0) \\ &\leqslant \frac{M}{\underline{\gamma}} + e^{\underline{\gamma}t_0}1^T h(t_0) \end{aligned} \tag{56}$$

thus, it can be concluded that $1^T h(t)$ is bounded, implying that $x(t)$, $\lambda(t)$ and $\mu(t)$ are bounded.

Conversely, from (53), we obtain

$$\begin{aligned} &1^T\dot{h}(t) + \frac{\underline{\theta}}{2\bar{\theta}^2}\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t) \\ &\quad = \dot{z}(t)^T\left(z(t) - z^*\right) + \frac{\underline{\theta}}{2\bar{\theta}^2}\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t) \leqslant M \end{aligned} \tag{57}$$

which together with the boundedness of $z(t)$ implies $\dot{z}(t)$ is bounded, and furthermore, obtaining $1^T\dot{h}(t)$ is bounded.

Recalling (51), one has

$$\begin{aligned} &1^T\ddot{h}(t) + \frac{d}{dt}\left(1^T\boldsymbol{\Gamma}h(t)\right) + \frac{\underline{\theta}}{2\bar{\theta}^2}\|\ddot{z}(t)\|^2 \\ &\quad + \omega\|\dot{z}(t)\|^2 + \frac{\underline{\theta}}{2\bar{\theta}^2}\frac{d}{dt}\left(\dot{z}(t)^T\boldsymbol{\Gamma}\dot{z}(t)\right) \leqslant 0. \end{aligned} \tag{58}$$

Integrating (58) from t_0 to t, which gives us that there is a $N > 0$, such that

$$1^T\dot{h}(t) + 1^T\mathbf{\Gamma}h(t) + \frac{\theta}{2\bar{\theta}^2}\int_{t_0}^{t}\|\ddot{z}(s)\|^2 ds + \omega\int_{t_0}^{t}\|\dot{z}(s)\|^2 ds + \frac{\theta}{2\bar{\theta}^2}\dot{z}(t)^T\mathbf{\Gamma}\dot{z}(t) \leqslant 1^T\mathbf{\Gamma}h(t_0) + 1^T\dot{h}(t_0) + \frac{\theta}{2\bar{\theta}^2}\dot{z}(t_0)^T\mathbf{\Gamma}\dot{z}(t_0) = N \quad (59)$$

which complies with the boundedness of $1^T\dot{h}(t)$ and implies that $\dot{z}(t), \ddot{z}(t) \in \mathsf{L}^2([t_0, +\infty))$, i.e.,

$$\int_{t_0}^{+\infty}\dot{z}(t)dt < +\infty,$$
$$\int_{t_0}^{+\infty}\ddot{z}(t)dt < +\infty.$$

It further implies $H(z(t)) \in \mathsf{L}^2([t_0, +\infty))$, i.e.,

$$\int_{t_0}^{+\infty}\|H(z(s))\|^2 ds < +\infty.$$

2) One has

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{z}(t)\|^2\right) = \dot{z}(t)^T\ddot{z}(t) \leqslant \frac{1}{2}\|\dot{z}(t)\|^2 + \frac{1}{2}\|\ddot{z}(t)\|^2 \ \forall\ t \in [t_0, +\infty) \quad (60)$$

and it togethers with Lemma 4 gives that $\lim_{t\to+\infty}\dot{z}(t) = 0$.

Next, let us prove $\lim_{t\to+\infty}H(z(t)) = 0$. Writing DIProx-NA (24) at time t and $t+\epsilon$ with $\epsilon > 0$, and setting $y_\epsilon(t) = (1/\epsilon)(\dot{z}(t+\epsilon) - \dot{z}(t))$, $\ddot{y}_\epsilon(t) = (1/\epsilon)(\ddot{z}(t+\epsilon) - \ddot{z}(t))$ and $\Upsilon_\epsilon(t) = (1/\epsilon)(H(z(t+\epsilon)) - H(z(t)))$, we have

$$\dot{y}_\epsilon(t) + \mathbf{\Gamma}y_\epsilon(t) = \Theta(\Upsilon_\epsilon(t)). \quad (61)$$

Since the operator H is 2-Lipschitz in Lemma 8, then

$$\Upsilon_\epsilon(t) \leqslant \frac{2}{\epsilon}\|z(t+\epsilon) - z(t)\| \leqslant 2\sup_{s\in[t,+\infty)}\|\dot{z}(s)\| = g(t). \quad (62)$$

Note that $\lim_{t\to+\infty}\dot{z}(t) = 0$, we have $\lim_{t\to+\infty}g(t) = 0$. Integrate (61) from t_0 to t and we can get

$$\|y_\epsilon(t)\| \leqslant \left(e^{\mathbf{\Gamma}t}\mathbf{\Gamma}\right)^{-1}\int_{t_0}^{t}e^{\mathbf{\Gamma}s}\Upsilon_\epsilon(s)ds + \left(e^{\mathbf{\Gamma}t}\right)^{-1}e^{\mathbf{\Gamma}t_0}\|y_\epsilon(t_0)\| \quad (63)$$

which implies $\lim_{t\to+\infty}(\sup_{\epsilon>0}\|y_\epsilon(t)\|) = 0$ as $t \to +\infty$. Since $\|\ddot{z}(t)\| \leqslant \sup_{\epsilon>0}\|y_\epsilon(t)\|$, we deduce that $\lim_{t\to+\infty}\ddot{z}(t) = 0$.

Furthermore, for any $t \in [t_0, +\infty)$, one has

$$\frac{d}{dt}\left(\frac{1}{2}\|H(z(t))\|^2\right) = H(z(t))^T\frac{d}{dt}H(z(t)) \leqslant \frac{1}{2}\|H(z(t))\|^2 + \frac{2}{2}\|\dot{z}(t)\|^2 \leqslant \frac{1}{2}\left\|\mathbf{\Theta}^{-1}(\ddot{z}(t) + \mathbf{\Gamma}\dot{z}(t))\right\|^2 + \|\dot{z}(t)\|^2 \leqslant \frac{1}{\underline{\theta}^2}\|\ddot{z}(t)\|^2 + \left(1 + \frac{\bar{\gamma}}{\underline{\theta}^2}\right)\|\dot{z}(t)\|^2 \quad (64)$$

in which the first inequality is valid because the Lipschitz property of operator H in Lemma 8 and the second inequality is satisfied from the conditions 1) and 2) in (41). ■

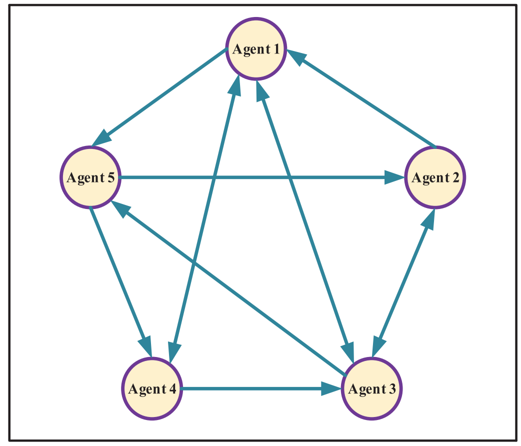

Fig. 4. Communication digraph of five agents.

Remark 5: In comparison with existing related distributed dynamical approaches, our DIProx-NA has the following advantages.

1) *The Local Objective Functions Are All Nonsmooth, Convex Functions:* Unlike the distributed dynamical approaches in [21] and [40], which rely on the strong convexity of the local objective function, and only one nonsmooth term of local objective function was considered in [25] and [41], the DIProx-NA proposed in this article can be used for locally convex, nonsmooth L_1-norm minimization problems.
2) *Avoiding the Subgradient Selection Dilemma:* Instead of the subgradients (i.e., differential inclusion method) [17], [18], [23], and [24], which can effectively address the challenge of selecting an appropriate subgradient at a nondifferentiable point. Furthermore, selecting a fixed subgradient is not optimal since this method does not guaranteed that the optimal solution is a stable equilibrium point (refer to Remark 2).
3) *Fully Distributed:* Instead of using global parameters in distributed continuous time approaches in [20], [21], [23], [24], and [25], the parameters in our DIProx-NA are fully distributed. Additionally, we also derive a sufficient condition for all the distributed parameters in DIProx-NA.
4) *Directed Graph:* In contrast to distributed projection neurodynamic approaches for problem (18) in [28], [29], and [30], which require the communication topology to be an undirected graph, the DIProx-NA proposed in this article only requires the communication graph to be directed, which is more generalizable.
5) *Weak Convergence:* Moreover, by means of Opial's lemma, we prove that the trajectory of DIProx-NA in Hilbert space asymptotically and weakly converges to the set of optimal solutions. To the best of our knowledge, this is the first attempt to investigate the trajectories of distributed neurodynamics approaches with weak convergence.

Theorem 4: The trajectory of $z(t)$ of (24), i.e., DIProx-NA (17) converges weakly to an element in **Zer** H, i.e., the optimal solution set of problem (15) on real Hilbert space $\mathcal{H}$.

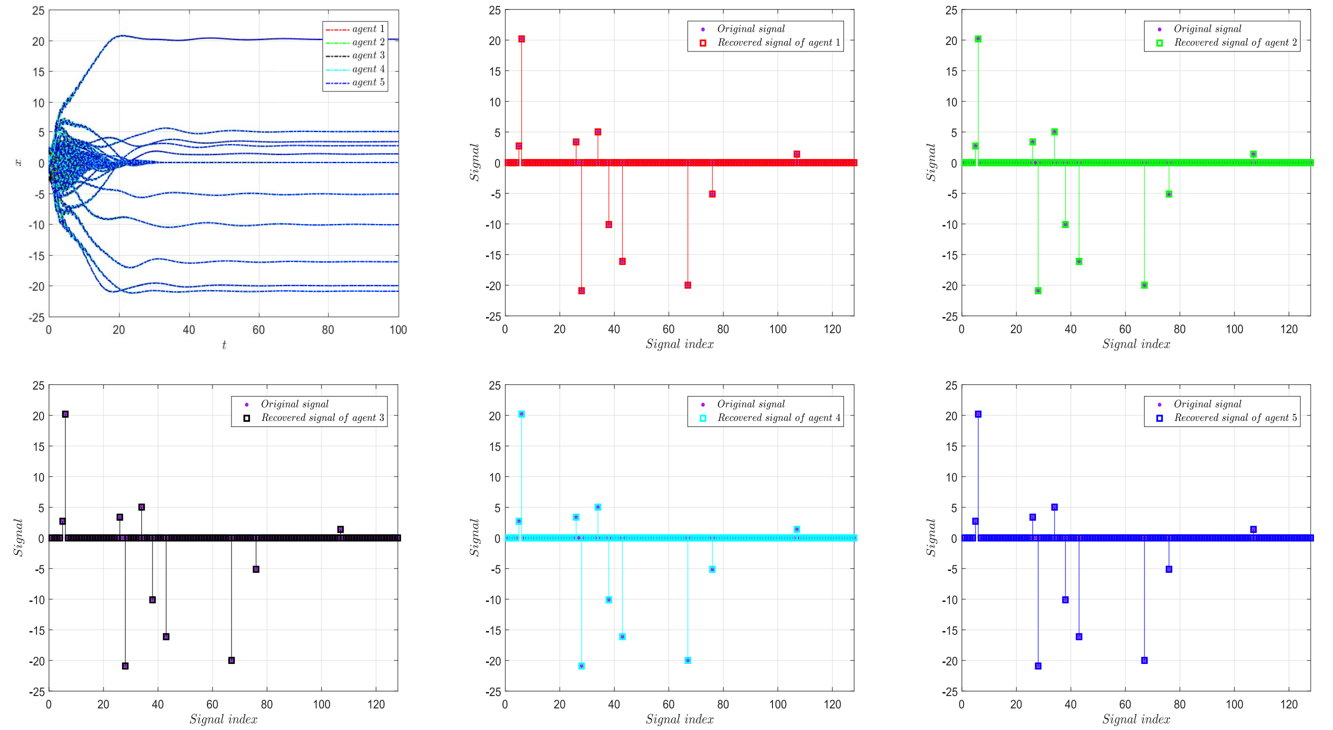

Fig. 5. *Top*: (Left) States transition curves of $x(t)$ to DIProx-NA (24). (Middle) Recovered signal of agent 1 in DIProx-NA. (Right) Recovered signal of agent 2 in DIProx-NA. *Bottom*: (Left) Recovered signal of agent 3 in DIProx-NA. (middle) Recovered signal of agent 4 in DIProx-NA. (Right) Recovered signal of agent 5 in DIProx-NA.

Proof: This proof is derived by the fact that both conditions in Lemma 5.

It is notable that the conclusions of this article still hold when the space of optimized variables is a real Hilbert space $\mathcal{H}$. In this case, the optimal problem (15) becomes

$$\min_{x\in\mathcal{X}}\|x\|_1, \ \text{ s.t. } Ax=b, Lx=0 \tag{65}$$

where $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are real Hilbert spaces, $\|x\|_1 : \mathcal{X}\to\mathbb{R}$ is a convex function, and A : $\mathcal{H}\to\mathcal{Y}$, $L^\star : \mathcal{H}\to\mathcal{Z}$ are linear continuous operators and $0\in\mathcal{Z}$. Moreover, $L^\star : \mathcal{H}\to\mathcal{Z}$ denotes the adjoint operator of Laplacian matrix L.

The corresponding DIProx-NA becomes

$$\begin{cases}\ddot{x}(t)+\boldsymbol{\gamma}\dot{x}(t)+\boldsymbol{\theta}\big(x(t)-\mathbf{Prox}_{\|\cdot\|_1}(x(t)-\boldsymbol{\alpha}Lx(t)\\ \qquad -\boldsymbol{\beta}A^\star(Ax(t)-b)-A^\star\lambda(t)-L^\star\mu(t))\big)=0\\ \ddot{\lambda}(t)+\boldsymbol{\gamma}\dot{\lambda}(t)=\boldsymbol{\theta}(Ax(t)-b)\\ \ddot{\mu}(t)+\boldsymbol{\gamma}\dot{\mu}(t)=\boldsymbol{\theta}Lx(t).\end{cases} \tag{66}$$

The conclusions of Lemmas 5–8 and Theorems 2 and 3 above are satisfied, which needs to do the computationally equivalent substitution, by replacing the transpose multiplication with the inner product operation $\mathfrak{a}^T\mathfrak{b} = <\mathfrak{a},\mathfrak{b}>$, $\|\mathfrak{a}\| = \sqrt{<\mathfrak{a},\mathfrak{a}>}$ (we still use the symbol $\|\mathfrak{a}\|^2$ to denote the inner product $<\mathfrak{a},\mathfrak{a}>$ in a real Hilbert space $\mathcal{H}$) and use the property that the inner product space is complete.

For 1) in Lemma 5, from (53), we have a similar result that $\varphi(t) =< 1,\dot{h}(t) > + < 1,\underline{\gamma}h(t) > +(\underline{\theta}/2\bar{\theta}^2) < \dot{z}(t),\boldsymbol{\Gamma}\dot{z}(t) >$ is monotonically decreasing, which means $\lim_{t\to+\infty}\Psi(t)$ exists.

According to Theorem 3, one also has $\lim_{t\to+\infty}\dot{z}(t)=0$. It combines with $< 1,\dot{h}(t) >=< \dot{z}(t), z(t)-z^* >$ and the boundedness of $z(t)$, we have $\lim_{t\to+\infty} < 1,\dot{h}(t) >= 0$. Then, one has

$$\lim_{t\to+\infty}\Psi(t)=\underline{\gamma}\lim_{t\to+\infty}\langle 1,h(t)\rangle \ \textit{exists}. \tag{67}$$

As a result, we get the existence of $\lim_{t\to+\infty}\|z(t)-z^*\| = \lim_{t\to+\infty} < z(t)-z^*, z(t)-z^* >$ (i.e., $\lim_{t\to+\infty}\|z(t)-z^*\|^2$ or $\lim_{t\to+\infty}\|x(t)-x^*\|^2+\|\lambda(t)-\lambda^*\|^2+\|\mu(t)-\mu^*\|^2$).

To prove 2) in Lemma 5, let z be a weak sequential cluster point of z, i.e., a sequence exists $t_n\to+\infty(n\to+\infty)$ which weakly converges to z. Since operator H is maximally monotone, then for the weak-strong topology of the product space $\mathcal{H}\times\mathcal{H}$ ($\mathcal{H}$ is a real Hilber space), the graph of it is continuously closed. Since $\lim_{n\to+\infty}H(z(t_n))=0$, we deduce that $Hz=0$, hence $z\in Zer\ H$. ■

V. Experimental Simulation

The validity of DIProx-NA (17) in recovering sparse signals and images is illustrated in this section.

Sparse Signal Recovery: Selecting Gaussian matrix $A_{m\times n}$ as a measurement matrix with $m=64$ and $n=128$. Producing a random sparse signal of sparsity $S=10$, that is, there exist 10 elements that are nonzero. Splitting the matrices A and b into 5 segments $A=\text{bldiag}\{A_1,A_2,A_3,A_4,A_5\}\in\mathbb{R}^{64\times640}$ and $b=\text{col}(b_1,b_2,b_3,b_4,b_5)\in\mathbb{R}^{64}$. Every agent can access to local information of A_i, b_i and A_j, b_j where $j\in\mathcal{N}_i^{\text{in}}=\{j\in\mathcal{V}|e_{ij}\in\mathcal{E}\}$ (i.e., agent j is a in-neighbor). The directed communication graph is given in Fig. 4.

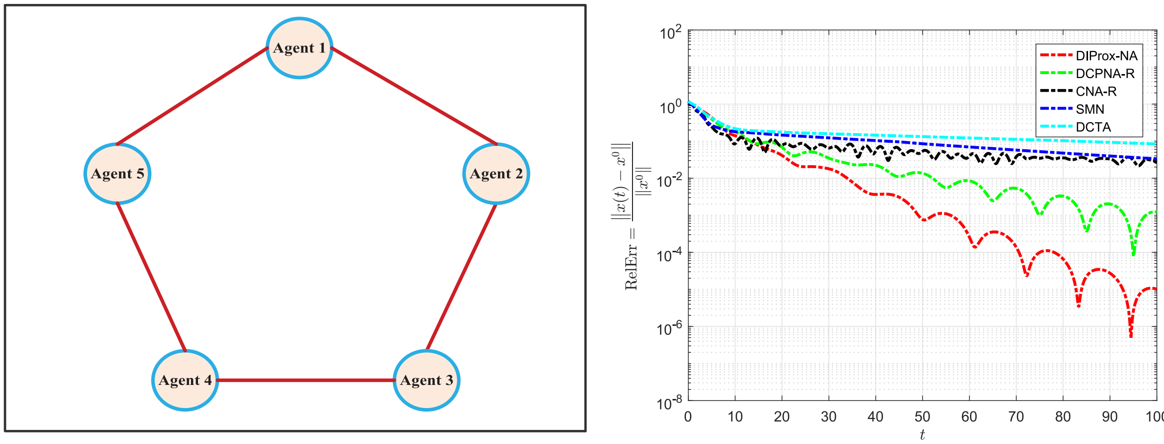

Fig. 6. (Left) Undirected communication digraph of five agents. (Right) Convergence rate with various distributed neurodynamic approaches in m=64, n=128, and S=10.

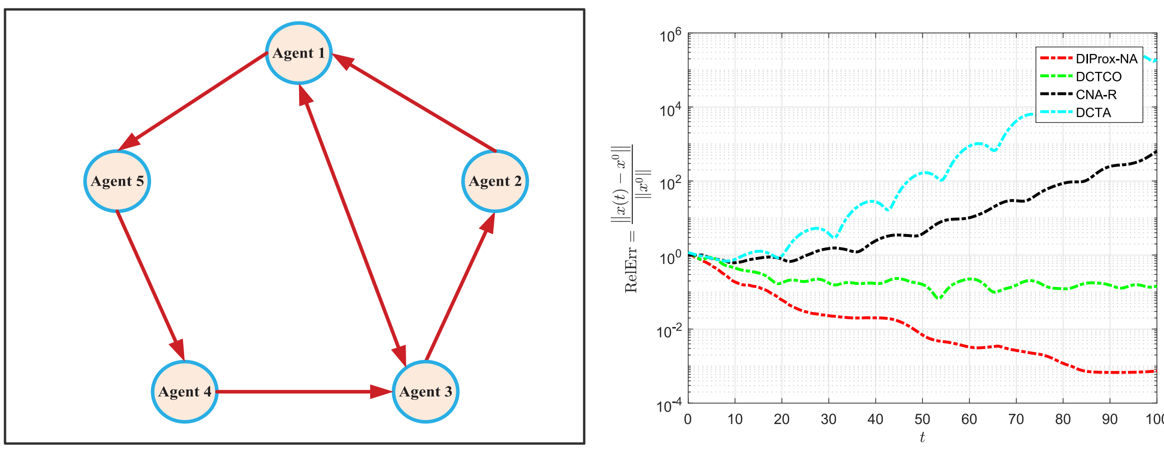

Fig. 7. (Left) Directed communication digraph of five agents, (Right) Convergence rate with various distributed neurodynamic approaches in $m = 64$, $n = 128$, and $S = 10$.

All these choices of the parameters γ_i, θ_i and α_i, β_i comply with the requirements in Theorem 3 and Remark 2. Fig. 5 displays the results of convergence properties and restoration of sparse signals by DIProx-NA (17) under a directed graph in Fig. 4. In addition, Fig. 5 (top, left) visualizes the state $x(t)$ is consensus (i.e., $L^T x = 0$) of DIProx-NA (17) and globally asymptotically stable which is consistent with the expression of Theorem 3. It is indicated from Fig. 5 (except the (top, left)) that the optimal solution of every agent in DIProx-NA (17) almost converges to the same original sparse signals. The results from Fig. 5 further confirm the effectiveness of DIProx-NA (17) and the correctness of Theorem 3.

To exhibit the effectiveness and superiority of DIProx-NA (17), we first compare it with some classical distributed neurodynamic approaches DCPNA-R [30], CNA-R [28], SMN [17], and DCTA [18] over undirected communication graph in Fig. 6. The RelErr $= [(\|x(t) - x^o\|)/(\|x^o\|)]$ (x^o stands for the true signal and $x(t)$ is the trajectory of DIProx-NA.) is used to evaluate the validity. As seen in Fig. 6 (right) that the DIProx-NA (17) has enjoyed faster convergence than DCPNA-R [30] and CNA-R [28] due to the fact that it has a inertial item (acceleration), moreover, it avoids the difficulty of choosing subgradients than SMN [17] and DCTA [18].

In addition, we compare DIProx-NA (17) with neurodynamic approaches DCTOC [23], CNA-R [28] and DCTA [18] over a directed communication graph in Fig. 7. As we can see from Fig. 7 (right) that DIProx-NA (17) converges faster due to it has a inertial item (acceleration) and has no trouble in choosing the subgradient than DCTCO [23]. From Fig. 7 (right), we also obtain that both CNA-R [28] and DCTA [18] are divergent, that is, they cannot solve the problem (15) over directed communication graphs.

Image Restoration: To further demonstrate the effectiveness of the DIProx-NA (17) over a directed graph that is shown in Fig. 4, we conduct an image reconstruction experiment with a "Cameraman" image in Fig. 8 (top,left), where the size of it is 128×128. We set the measurement number of measurement matrix (Hadamard matrix) to be $m = 90$ and utilize the peak-signal noise ratio (PSNR) to assess the performance of the image reconstruction, which is given by

$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}}$$

in which $\mathrm{MSE} = 1 \Big/ (n \times n) \sum_{i,j} (\bar{x}(i,j) - x(i,j))^2$ and $\bar{x}(i,j)$, $x(i,j)$ represents the pixel values in the original and recovered

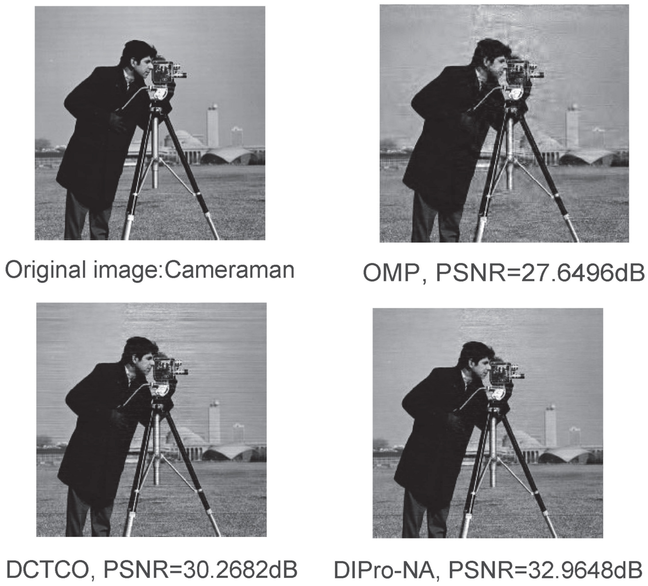

Fig. 8. Top: (left) Original "Cameraman" image; (right) Recovered image of OMP [10] with PSNR=27.6496dB; Bottom: (left) Recovered image of DCTCO [23] with PSNR=30.2682dB; (right) Recovered image of DIProx-NA (17) with PSNR=32.9648dB.

images, respectively, and n represents the size (width or height) of the image.

As depicted in Fig. 8, it is evident that DIProx-NA (17) surpasses other methods in performance since it employs proximal operator of L_1-norm, avoiding the dilemma of subgradients selection at a nondifferential point of L_1-norm than DCTCO [23], moreover, the DIProx-NA (17) is designed based on convex optimization theory, so it has a greater probability of obtaining an optimal solution of the image reconstruction problem compared than OMP [10].

VI. Conclusion

We have already researched the distributed L_1-minimization problem depending on the row partition of measurement matrix A, then proposed a novel fully DIProx-NA in this article. The proximal operator of L_1-norm is carried out to effectively address the nonsmooth objective function, and the primal-dual frame with inertia is also applied to deal with the linear observation and consensus constraints. Furthermore, we have theoretically proved the optimality and weak convergence of the DIProx-NA by employing the KKT conditions, Lipschitz condition, theory of maximal monotone operator and Opial's lemma, respectively. The reconstruction results on sparse signals and "Lena" image have shown the feasibility and effectiveness of DIProx-NA over directed graphs. In our future work, we plan to investigate the distributed L_1-norm sparse signal reconstruction problem in weight unbalanced directed graph scenarios using distributed accelerated neurodynamic approaches. Additionally, we aim to expand the L_1-norm distributed optimization problem by employing the nonconvex $L_p(1 \geqslant p > 0)$-norm instead of L_1-norm under undirected graph scenario, explore the correspondingly distributed neurodynamic approaches. However, there are very few studies exploring the convergence rates of distributed convex optimization problems with only nonsmooth, convex functions. This is primarily due to their weaker properties compared to strongly convex functions. The latest literatures [42] and [43] focuss on exploring the convergence properties of the algorithms, which is an important direction for further study in the future.

References

[1] F. Bach, J. Mairal, J. Ponce, and G. Sapiro, "Sparse coding and dictionary learning for image analysis," in *Proc. Int. Conf. Adv. Comput. Intell.*, 2010, pp. 1–52.

[2] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[3] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 41, no. 1, pp. 38–52, Feb. 2011.

[4] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.* vol. 49, no. 1, pp. 136–144, Jan. 2019.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[6] S. S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[7] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[8] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[9] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[10] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[11] Q. Liu and J. Wang, "L_1–minimization algorithms for sparse signal reconstruction based on a projection neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 698–707, Mar. 2016.

[12] R. Feng, C. Leung, A. G. Constantinides, and W. Zeng, "Lagrange programming neural network for nondifferentiable optimization problems in sparse approximation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2395–2407, Oct. 2017.

[13] H. Wang, R. Feng, C.-S. Leung, J. Sum, and A. G. Constantinides, "A globally stable LPNN model for sparse approximation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 5218–5226, Aug. 2023.

[14] H. Wen, H. Wang, and X. He, "A neurodynamic algorithm for sparse signal reconstruction with finite-time convergence," *Circuits. Syst. Signal Process.*, vol. 39, pp. 6058–6072, May 2020.

[15] X. He, H. Wen, and T. Huang, "A fixed-time projection neural network for solving l_1-minimization problem," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7818–7828, Dec. 2022.

[16] H. Wen, X. He, T. Huang, and J. Yu, "Neurodynamic algorithms with finite/fixed-time convergence for sparse optimization via ℓ_1-regularization," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 1, pp. 131–142, Jan. 2024, doi: 10.1109/TSMC.2023.3304850.

[17] Q. Liu and J. Wang, "A second-order multi-agent network for bound-constrained distributed optimization," *IEEE Trans. Autom. Control.*, vol. 60, no. 12, pp. 3310–3315, Dec. 2015.

[18] X. Zeng, P. Yi, Y. Hong, and L. Xie, "Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems," *SIAM J. Control Optim.*, vol. 56, no. 6, pp. 3973–3993, 2018.

[19] X. Le, S. Chen, F. Li, Z. Yan, and J. Xi, "Distributed neurodynamic optimization for energy Internet management," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 8, pp. 1624–1633, Aug. 2019.

[20] X. Yi, L. Yao, T. Yang, J. George, and K. H. Johansson, "Distributed optimization for second-order multi-agent systems with dynamic event-triggered communication," in *Proc. IEEE Conf. Decis. Control (CDC)*, 2018, pp. 3397–3402.

[21] Y. Wei, H. Fang, X. Zeng, J. Chen, and P. M. Pardalos, "A smooth double proximal primal-dual algorithm for a class of distributed nonsmooth optimization problem," *IEEE Trans. Autom. Control*, vol. 65, no. 4, pp. 1800–1806, Apr. 2020.

[22] Q. Wang, J. Chen, X. Zeng, and B. Xin, "Distributed proximal-gradient algorithms for nonsmooth convex optimization of second-order multiagent systems," *Int J. Robust Nonlin.*, vol. 30, no. 17, pp. 7574–7592, 2020.

[23] B. Gharesifard and J. Cortès, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control.*, vol. 59, no. 3, pp. 781–786, Mar. 2014.

[24] X. Jiang, S. Qin, and X. Xue, "Continuous-time algorithm for approximate distributed optimization with affine equality and convex inequality constraints," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 9, pp. 5809–5818, Sep. 2021.

[25] X. Wang, S. Yang, Z. Guo, and T. Huang, "A second-order projected primal-dual dynamical system for distributed optimization and learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6568–6577, Sep. 2023.

[26] J. Wang, G. Wen, and Z. Duan, "Distributed antiwindup consensus control of heterogeneous multiagent systems over Markovian randomly switching topologies," *IEEE Trans. Autom. Control.*, vol. 67, no. 11, pp. 6310–6317, Nov. 2022.

[27] J. Wang, X. Deng, J. Guo, Y. Luo, and K. Li, "A fully distributed antiwindup control protocol for intelligent-connected electric vehicles platooning with switching topologies and input saturation," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 3, pp. 1683–1694, Jun. 2023.

[28] Y. Zhao, X. Liao, X. He, and R. Tang, "Centralized and collective neurodynamic optimization approaches for sparse signal reconstruction via l_1-minimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7488–7501, Dec. 2022.

[29] J. Xu, X. He, X. Han, and H. Wen, "A two-layer distributed algorithm using neurodynamic system for solving L_1-minimization," *IEEE Trans. Circuits Syst. II, Exp. Briefs.*, vol. 69, no. 8, pp. 3490–3494, Aug. 2022, doi: 10.1109/TCSII.2022.3159814.

[30] Y. Zhao, X. Liao, and X. He, "Distributed continuous and discrete time projection neurodynamic approaches for sparse recovery," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 6, pp. 1411–1426, 2022.

[31] X. He, T. Huang, J. Yu, C. Li, and C. Li, "An inertial projection neural network for solving variational inequalities," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 809–814, Mar. 2017.

[32] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2014.

[33] B. Abbas, H. Attouch, and B. F. Svaiter, "Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces," *J. Optim. Theory Appl.*, vol. 161, pp. 331–360, 2014.

[34] R. E. Bruck, "Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces," *J. Funct. Anal.* vol. 18, no. 1, pp. 15–26, 1975.

[35] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (CMS Books in Mathematics). New York, NY, USA: Springer, 2011.

[36] E. Candès and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, pp. 2313–2351, Dec. 2007.

[37] B. Guenin, J. Könemann, and L. Tuncel, *A Gentle Introduction to Optimization*. Cambridge, U. K.: Cambridge Univ. Press, 2014.

[38] H. Zhang, D. Yue, C. Dou, K. Li, and X. Xie, "Event-triggered multiagent optimization for two-layered model of hybrid energy system with price bidding-based demand response," *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 2068–2079, Apr. 2021.

[39] H. Zhang, D. Yue, C. Dou, Y. Xue, and G. P. Hancke, "Event-trigger-based distributed optimization approach for two-level optimal model of isolated power system with switching topology," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 4, pp. 2339–2349, Apr. 2023.

[40] Y. Zhu, G. Wen, W. Yu, and X. Yu, "Continuous-time distributed proximal gradient algorithms for nonsmooth resource allocation over general digraphs," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1733–1744, Apr.–Jun. 2021.

[41] Y. Zhu, G. Wen, W. Yu, and X. Yu, "Nonsmooth resource allocation of multiagent systems with disturbances: A proximal approach," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 3, pp. 1454–1464, Sep. 2021.

[42] Y. Huang, Z. Meng, J. Sun, and W. Ren, "Distributed multiproximal algorithm for nonsmooth convex optimization with coupled inequality constraints," *IEEE Trans. Autom. Control.*, vol. 68, no. 12, pp. 8126–8133, Dec. 2023.

[43] Y. Huang, Z. Meng, J. Sun, and G. Wang, "Distributed continuous-time proximal algorithm for nonsmooth resource allocation problem with coupled constraints," *Automatica*, vol. 159, Jan. 2024, Art. no. 111309.

You Zhao received the M.S. degree in signal and information processing from the College of Electronics and Information Engineering, Southwest University, Chongqing, China, in 2018, and the Ph.D. degree in computer science and technology from Chongqing University, Chongqing, in 2023.

He is currently a Postdoctoral Researcher with the College of Electronic Information Engineering, Southwest University. His current research interests include neurodynamic optimization, distributed optimization, compressed sensing, minimax optimization, and smart grid.

Xing He (Member, IEEE) received the B.S. degree in mathematics and applied mathematics from the Department of Mathematics, Guizhou University, Guiyang, China, in 2009, and the Ph.D. degree in computer science and technology from Chongqing University, Chongqing, China, in 2013.

From November 2012 to October 2013, he was a Research Assistant with Texas A&M University at Qatar, Doha, Qatar. From December 2015 to February 2016, he was a Senior Research Associate with The City University of Hong Kong, Hong Kong. He is currently a Professor with the College of Electronics and Information Engineering, Southwest University, Chongqing. His research interests include neural networks, bifurcation theory, optimization method, smart grid, and nonlinear dynamical system.

Mingliang Zhou received the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2017. He was a Postdoctoral Fellow with the Department of Computer Science, The City University of Hong Kong, Hong Kong, from September 2017 to September 2019. He was a Postdoctoral Fellow with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau, China, from October 2019 to October 2021. He is currently an Associate Professor with the School of Computer Science, Chongqing University, Chongqing, China. His research interests include image and video coding, perceptual image processing, multimedia signal processing, rate control, multimedia communication, machine learning, and optimization.

Junzhi Yu (Fellow, IEEE) received the B.E. degree in safety engineering and the M.E. degree in precision instruments and mechanology from North University of China, Taiyuan, China, in 1998 and 2001, respectively, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2003.

From 2004 to 2006, he was a Postdoctoral Research Fellow with the Center for Systems and Control, Peking University, Beijing. He was an Associate Professor with the Institute of Automation, Chinese Academy of Sciences in 2006, where he was a Full Professor in 2012. In 2018, he was a Tenured Full Professor with the College of Engineering, Peking University. His current research interests include intelligent robots, motion control, and intelligent mechatronic systems.

Tingwen Huang (Fellow, IEEE) received the B.S. degree in mathematics from Southwest Normal University, Chongqing, China, in 1990, the M.S. degree in applied mathematics from Sichuan University, Chengdu, China, in 1993, and the Ph.D. degree in mathematics from Texas A&M University, College Station, TX, USA, in 2002.

He was a Lecturer with Jiangsu University, Zhenjiang, China, from 1994 to 1998, and a Visiting Assistant Professor with Texas A&M University at Qatar, Doha, Qatar, in 2003. He was an Assistant Professor from 2003 to 2009 and an Associate Professor from 2009 to 2013 with Texas A&M University at Qatar, Doha, Qatar, where he has been a Professor since 2013. His current research interests include neural networks, complex networks, chaos and dynamics of systems, and operator semi-groups and their applications.