Distributed Smoothing Projection Neurodynamic Approaches for Constrained Nonsmooth Optimization

You Zhao, Xiaofeng Liao, *Fellow, IEEE*, and Xing He

***Abstract*—This article considers constrained nonsmooth generalized convex and strongly convex optimization problems. For such problems, two novel distributed smoothing projection neurodynamic approaches (DSPNAs) are proposed to seek their optimal solutions with faster convergence rates in a distributed manner. First, we equivalently transform the original constrained optimal problem into a standard smoothing distributed problem with only local set constraints based on an exact penalty and smoothing approximation methods. Then, to deal with nonsmooth generally convex optimization, we propose a novel DSPNA based on continuous variant of Nesterov's acceleration (called DSPNA-N), which has a faster convergence rate $\mathcal{O}(1/t^2)$, and we design a novel DSPNA inspired by the continuous variant of Polyak's heavy ball method (called DSPNA-P) to address the nonsmooth strongly convex optimal problem with an explicit exponential convergent rate. In addition, the existence, uniqueness, and feasibility of the solution of our proposed DSPNAs are also provided. Finally, numerical results demonstrate the effectiveness of DSPNAs.**

***Index Terms*—Arithmetic and exponential convergence, distributed smoothing projection neurodynamic approach (DSPNA), nonsmooth.**

I. Introduction

DISTRIBUTED optimization aims to minimize the additive function with certain constraints by using the agents' and their neighbors' local information upon a multiagent network, and it plays an important role in sensor networks [1], machine learning [2], resource allocation [3]–[10], etc.

Manuscript received 24 January 2022; accepted 15 June 2022. Date of publication 15 July 2022; date of current version 16 January 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0100101; in part by the National Natural Science Foundation of China under Grant 61932006 and Grant 62176027; in part by the Chongqing Technology Innovation and Application Development Project, China, under Grant cstc2020jscx-msxmX0156; and in part by the Fundamental Research Funds for the Central Universities, China, under Project XDJK2020TY003. This article was recommended by Associate Editor J. Liang. *(Corresponding author: Xiaofeng Liao.)*

You Zhao and Xiaofeng Liao are with the Key Laboratory of Dependable Services Computing in Cyber Physical Society-Ministry of Education, College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: zhaoyou1991sdtz@163.com; xfliao@cqu.edu.cn).

Xing He is with the Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, School of Electronics and Information Engineering, Southwest University, Chongqing 400715, China (e-mail: hexingdoc@swu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2022.3186019.

Digital Object Identifier 10.1109/TSMC.2022.3186019

Multifarious distributed methods, i.e., distributed continuous-time approaches and distributed discrete-time algorithms, have been presented to solve constrained and unconstrained optimization problems. Regarding the distributed discrete-time algorithms, they include the distributed subgradient algorithm [7], the distributed dual averaging algorithm [8], the distributed discrete-time algorithm based on primal–dual method [9], distributed discrete-time algorithm [10] based on Nesterov's accelerated method and distributed discrete-time algorithm [11] inspired by the "momentum" method, ADMM [12], and so on. The continuous-time approaches are increasingly popular because they can be realized by analog circuits, and furthermore, the Lyapunov analysis method of the dynamical system, providing an effective guiding tool for their convergence analysis. Many recent interesting works on continuous-time approaches were researched for addressing distributed optimization problems. For example, these works include the distributed continuous-time approaches (i.e., differential equations and differential inclusions) based on augmented Lagrangian multiplier methods [13], [14] for unconstrained distributed optimization, distributed approaches based on primal–dual dynamical system and projection operators [15], [16], and penalty-based approaches [17], [18] for constrained distributed problems.

Convergence rate is an essential evaluation criterion for the performance of distributed continuous-time dynamical approaches. Particularly, the distributed continuous-time approaches, which have an arithmetic convergence rate for distributed generally convex optimization and an exponential convergence for distributed strongly convex optimization, have been studied. For instance, in order to solve the constrained distributed nonsmooth constrained optimization, distributed nonsmooth dynamical approaches inspired by primal–dual dynamics with an algebraic convergence rate $\mathcal{O}(1/t)$ have been investigated in [19] and [20], and a continuous-time multiagent neurodynamic approach (MNA) on the basis of projection operators and penalty method with a convergence rate $\mathcal{O}(1/t^r), 0 < r < 1$ was proposed in [21]. In the distributed strongly convex case, several distributed continuous-time approaches [22], [23] were proposed based on primal–dual dynamics, which have exponential convergence rates for constrained and unconstrained optimizations. In addition, to solve constrained distributed nonsmooth strongly convex optimization problems, Li *et al.* [24] proposed a distributed subgradient projection continuous-time approach based on the differential projection operator with an exponential convergence rate. In [25] and [26], two distributed continuous-time approaches were explored to address a constrained resource allocation problem with an exponential convergence rate. Wang *et al.* [27] investigated a distributed projection algorithm in delayed multiagent networks, which achieves the $\mathcal{O}(1/\sqrt{k})$ convergence rate for a constrained distributed convex problem. Yang *et al.* [28] investigated a distributed neurodynamic approaches for convex optimization under the presence of communication delays. Zhao and Liu [29] proposed a consensus algorithm based on the collective neurodynamic system (CNS) to solve distributed constrained nonsmooth convex optimization problems. Zhu *et al.* [30] proposed a distributed approach based on primal–dual dynamics and differential projection operator to solve distributed constrained, nonsmooth, convex optimization problem.

Recently, Su *et al.* [31] revealed that the continuous variant of Nesterov's acceleration is a second-order dynamical system with a vanishing friction, which has a faster arithmetic convergence rate $\mathcal{O}(1/t^2)$ for unconstrained smooth convex optimal problems. Combining the continuous variant of Nesterov's acceleration and primal–dual method, a primal–dual accelerated dynamical method (PDAD) was first proposed in [32], which maintains the fast convergence rate $\mathcal{O}(1/t^2)$ for solving the smooth convex optimal problem with affine constraints. In addition, the PDAD was extended to solve two unconstrained distributed smooth optimization problems. Jiang *et al.* [33] proposed a second-order accelerated neurodynamic approach to deal with constrained convex problems based on inertial systems and the time-varying penalty method, which has a superquadratic convergence rate when the time is greater than a certain moment. In [34], by using the proximal operator to deal with nonsmooth functions, Wang *et al.* proposed a distributed proximal-gradient algorithm based on second-order multiagent systems (i.e., continuous variant of Polyak's heavy ball method [35]). However, the authors only analyzed the convergence of their proposed algorithm and did not discuss its convergence rate. Wei *et al.* [36] investigated a double proximal primal–dual algorithm to deal with a class of distributed nonsmooth convex consensus optimization problems only with the analysis of the convergence properties of the proposed approach. In addition, Chai *et al.* [37] presented an approximation-based strategy to deal with chance-constrained trajectory optimization problems (CCOCPs). They converted the model of CCOCP into a parametric nonlinear programming model by using the smooth differentiable approximation function to replace probabilistic constraints, and then used the two-nested gradient-based algorithm [38] to deal with them. Chai *et al.* [39] investigated a specific multiple-shooting discretization technique with the newest NSGA-III optimization algorithm and three constraint handling algorithms to deal with multiobjective trajectory planning problems. Nevertheless, for constrained distributed nonsmooth convex and strongly convex optimization problems, the designs for distributed accelerated neurodynamic approaches with faster convergence rate $\mathcal{O}(1/t^2)$ and faster exponential convergence rate still remain challenging issues. To the best of our knowledge, there are two difficulties as follows.

1) There exist both consensus and set constraints. It is difficult to satisfy the above constraints effectively in designing the distributed smoothing projection neurodynamic approaches (DSPNAs).
2) The other is to use a new approximation technique to solve nonsmooth objective functions in optimization problems. Although the differential inclusion technique can effectively solve the nonsmooth optimization problem, there exists subgradient selection challenge in implementing neurodynamic approaches especially when designing accelerated neurodynamic approaches.

Inspired by the works [24], [31], and [35], we focus on designing neurodynamic approaches to solve the additive nonsmooth optimal problem with local set constraints, where the objective functions are nonsmooth generally convex and strongly convex. Our contributions are listed as follows.

1) By applying an exact nonsmooth penalty method for solving consensus constraints, a projection operator with an auxiliary variable is used to contend with set constraints, and a smoothing approximation technique is applied to dispose of the nonsmooth functions. Consequently, we propose two faster DSPNAs for distributed nonsmooth generally convex and strongly convex optimizations, i.e., a DSPNA-N for nonsmooth generally convex optimization with faster convergence rate $\mathcal{O}(1/t^2)$ and DSPNA-P for nonsmooth strongly convex optimization with faster explicit exponential convergent rate. In addition, our proposed DSPNAs have lower communication and computation costs than the existing primal–dual algorithms [4], [40] since auxiliary variables do not need to communicate with their neighbors.
2) Compared with the works in [15], [20], and [24] based on the differential inclusion method, our proposed DSPNAs can avoid the difficulties of calculating and implementing differential inclusion or differential projection operators. Moreover, different from the distributed approaches in [21] and [23], we relax the regularity assumptions on the objective function and guarantee the existence and uniqueness to the solution of DSPNAs.
3) Different from the works in [16] and [18], we offer two novel Lyapunov functions based on projection operators and smoothing parameters that can easily derive explicit accelerated convergence rate.

The research is organized as follows. In Section II, some preliminaries are introduced. The distributed constrained optimization problem is reformulated by the exact penalty and smoothing approximation methods are discussed in Section III. In Section IV, two DSPNAs are proposed based on continuous variant of Nesterov's acceleration and Polyak's heavy ball method to solve constrained generally convex and strongly convex optimization problems, respectively. In Section V, experimental results are provided, and the conclusion is given in Section VI.

Notations: Let R be a real number set and $R_{>0}$ be a positive real number and R^n be a n-dimensional column vector set. Superscript T represents the transpose. $|x| = \sum_{i=1}^{n} |x_i|$ denotes the 1-norm if $x \in R^n$; moreover, it denotes the absolute value function if $x \in R$. $\|x\| = (\sum_{i=1}^{n} x_i^2)^{1/2}$ denotes the Euclidean norm. Denote $\|x\|_0$ as the 0-norm that describes the number of nonzero elements of $x \in R^n$. Denote $\mathbf{1}$ and $\mathbf{0}$ as vectors with all entries being 1 and 0, respectively. $L^1(0, +\infty)$ denotes the Lebesgue integrable function in $(0, +\infty)$. If $x_i \in R^n$, $i = 1, \ldots, m$, then $x = \mathrm{col}(x_1, \ldots, x_m) \in R^{nm}$. Let $\Omega_1 \times \cdots \times \Omega_n$ be the Cartesian product of sets $\Omega_1, \ldots, \Omega_n$. $\mathrm{sgn}(x)$ is the signum function.

II. Preliminaries

A. Projection Operator

For a nonempty, closed and convex set $\Omega \in R^n$, the projection operator is defined as $\Pi_\Omega(x) = \operatorname*{argmin}_{u\in\Omega}\|u - x\|$ and it has the following properties.

Lemma 1 [41]: If $\Omega \in R^n$ is a nonempty, closed, and convex set, then it yields the following.

1) $\|u - v\| \geq \|\Pi_\Omega(u) - \Pi_\Omega(v)\| \quad \forall\, u, v \in R^n$.
2) $(\Pi_\Omega(v) - v)^T(\Pi_\Omega(v) - u) \leq 0 \quad \forall\, v \in R^n, u \in \Omega$.
3) The normal cone of set Ω is

$$\begin{aligned}\mathcal{N}_\Omega(x) &= \mathrm{cl}\left(\bigcup_{\nu\geq 0} \nu\partial d_\Omega(x)\right)\\ &= \left\{v \in R^n | v^T(y - x) \leq 0 \quad \forall y \in \Omega\right\}\end{aligned}$$

where the symbol $\mathrm{cl}(S)$ represents closure of set S, $d_\Omega(x) = \min_{u\in\Omega}\|x - u\|$.

From 2) and 3) in Lemma 1, we have $x - \Pi_\Omega(x) \in \mathcal{N}_\Omega(x)$. In addition, the ν is a nonnegative scalar and $\partial d_\Omega(x)$ means the Clarke generalized gradient of function $d_\Omega(x)$, which is given in Section II-B.

Lemma 2 [41]: Define a function $\phi(u, v) : R^n \times R^n \to R$ as follows:

$$\phi(u, v) = \frac{1}{2}\left(\|u - \Pi_\Omega(v)\|^2 - \|u - \Pi_\Omega(u)\|^2\right) \tag{1}$$

and it has two properties as follows: 1) $\phi(u, v) \geq (1/2)\|\Pi_\Omega(u) - \Pi_\Omega(v)\|^2$ and 2) $\phi(u, v)$ is a continuously differentiable function, and its gradient with respect to u is $\nabla_u\phi(u, v) = \Pi_\Omega(u) - \Pi_\Omega(v)$.

Lemma 3 [42]: For *box*, *Euclidean sphere*, and *affine* constrained sets, there exist analytical expressions of the projection operator.

1) Let Ω be a *box* or *hyperrectangle* set, i.e., $\Omega = \{x \in R^n \mid x_{i,\mathrm{m}} \leq x_i \leq x_{i,M}\ ,\ i = 1, \ldots, n\}$; then

$$\Pi_\Omega(x)_i = \begin{cases} x_{i,\mathrm{m}}, & x_i < x_{i,\mathrm{m}}\\ x_i, & x_{i,\mathrm{m}} \leq x_i \leq x_{i,M}\\ x_{i,M}, & x_i > x_{i,M}.\end{cases} \tag{2}$$

2) Assume Ω is an *Euclidean sphere* set, i.e., $\Omega = \{x \in R^n \mid \|x - z\| \leq r\ ,\ z \in R^n,\ d > 0\}$, then

$$\Pi_\Omega(x) = \begin{cases} x, & \|x - z\| \leq d\\ z + \frac{r(x-z)}{\|x-z\|}, & \|x - z\| > d.\end{cases} \tag{3}$$

3) If Ω is an *affine* set, i.e., $\Omega = \{x \in R^n \mid Ax = b\}$, then

$$\Pi_\Omega(x) = x + A^\dagger(b - Ax) \tag{4}$$

where $A^\dagger$ is the *Moore–Penrose pseudoinverse* of A. For example, if $m \leq n$ and $rank(A) = m$, then this projection operator yields

$$\Pi_\Omega(x) = x + A^T\left(AA^T\right)^{-1}(b - Ax). \tag{5}$$

B. Convex Analysis

The Clarke generalized gradient of the locally Lipschitz function g is defined as follows:

$$\partial g(x) = \mathrm{co}\left\{\lim_{x_k\to x: x_k\in D_g} \nabla\partial g(x_k)\right\}$$

where D_g is a set at which g is differentiable and $\mathrm{co}(S)$ defines the convex hull of a set S.

A function $g : \Omega \to R$ is a nonsmooth generally convex function if it only satisfies

$$g(z) - g(w) \geq (z - w)^T g_f(w) \quad \forall z, w \in \Omega \tag{6}$$

where $g_f(w) \in \partial g$ and $\Omega \subset R^n$.

For any $z, w \in \Omega$, the function g is a strongly convex if it satisfies

$$g(z) - g(w) \geq (z - w)^T g_f(w) + \frac{\eta}{2}\|-w\|^2 \tag{7}$$

where $g_f(w) \in \partial g$.

C. Smoothing Approximation

Definition 1 [43]: If $\hat{g} : R^n \times (0, +\infty) \to R$ is a smoothing function of g, where $g : R^n \to R$ is a locally Lipschitz function, then we have the following.

1) For any $\mu > 0$, $\hat{g}(\cdot, \mu)$ is continuously differentiable of x in R^n, and $\hat{g}(x, \cdot)$ is differentiable in $[0, +\infty)$ with any fixed $x \in R^n$.
2) $\lim_{\mu\to 0^+} \hat{g}(x, \mu) = g(x)$ with any fixed $x \in R^n$.
3) There exists a positive constant $\kappa_{\hat{g}} > 0$ such that
$$\left|\nabla_\mu \hat{g}(x, \mu)\right| \leq \kappa_{\hat{g}} \quad \forall\mu \in [0, +\infty),\ x \in R^n.$$
4) $\{\lim_{z\to x,\mu\to 0} \nabla_z\hat{g}(z, \mu)\} \subseteq \partial g(x)$. From the above definitions 2) and 3), for any fixed $x \in R^n$, the following conditions are true.
5) $\lim_{z\to x,\mu\to 0} \hat{g}(z, \mu) = g(x)$.
6) $|\hat{g}(x, \mu) - g(x)| \leq \kappa_{\hat{g}}\mu \quad \forall\mu \in [0, +\infty), x \in R^n$.

Lemma 4 [44]: Let $\hat{g}_1, \ldots, \hat{g}_m$ be smoothing functions of $g_1, \ldots, g_m$, such that $\sum_{i=1}^{m} a_i\hat{g}_i$ is a smoothing function of $\sum_{i=1}^{m} a_i g_i$ with $\kappa_{\sum_{i=1}^{m} a_i\hat{g}_i} = \sum_{i=1}^{m} a_i\kappa_{\hat{g}_i}$ when $a_i \geq 0$ and g_i is regular for any $i = 1, 2, \ldots, m$.

In this article, a smoothing approximation function of the absolute value function $|s|$, $s \in R$ presented in [43] is used

$$\hat{g}(s, \mu) = \begin{cases} |s|, & \text{if } |s| > \mu\\ \frac{s^2}{2\mu} + \frac{\mu}{2}, & \text{if } |s| \leq \mu\end{cases} \tag{8}$$

where $\lim_{\mu\to 0^+} \hat{g}(s, \mu) = |s|$ from 2) in Definition 1, and it is shown in Fig. 1.

In addition, the derivative $\hat{g}(s, \mu)$ of s with any fixed $\mu > 0$ is given by

$$\nabla_s\hat{g}(s, \mu) = \begin{cases} \mathrm{sgn}(s), & \text{if } |s| > \mu\\ \frac{s}{\mu}, & \text{if } |s| \leq \mu.\end{cases} \tag{9}$$

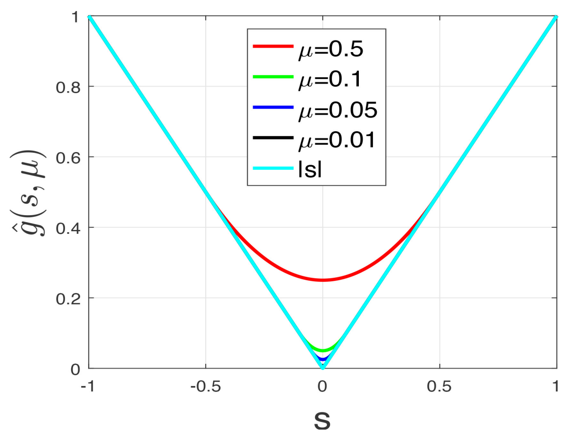


Fig. 1. $\hat{g}(s, \mu)$ with different parameters μ.

In addition, the derivative $\hat{g}(s, \mu)$ at μ with any fixed x is represented as

$$\nabla_{\mu}\hat{g}(s, \mu) = \begin{cases} 0, & \text{if } |s| > \mu \\ \frac{1}{2} - \frac{s^2}{2\mu^2}, & \text{if } |s| \leq \mu. \end{cases} \tag{10}$$

Note that $(1/2) \geq \nabla_{\mu}\hat{g}(s, \mu) \geq 0$ holds for any fixed x and $\mu \in [0, +\infty)$.

D. Graph Theory

An undirected communication topology graph is a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with node set $\mathcal{V} = \{v_1, v_2, \ldots, v_m\}$, edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and connection matrix $\mathcal{A} = \{a_{ij}\}_{m\times m}$ with nonnegative elements $a_{ij} = a_{ji} > 0$ if $(i, j) \in \mathcal{E}$, and $a_{ij} = a_{ji} = 0$ otherwise. The coupling of agents in an undirected graph is unordered, which means that there exists information exchange for both agent i and agent j. A path in an undirected graph between agent i and agent j is a sequence of edges of the form (i, i_1), (i_1, i_2), $\ldots$, (i_s, j), where i, i_1, $\ldots$, i_s, j denote different agents. Let $\mathcal{N}_i = \{j|(i, j) \in \mathcal{E}\}$ be an agent i's neighbors set. The undirected graph $\mathcal{G}$ is connected if there exists a path between any pair of distinct nodes v_i and v_j $(i, j = 1, 2, \ldots, m)$.

III. Problem

A. Distributed Optimization Problem Formulation

Consider an undirected network that consists of m agents. For every agent, there exists a local nonsmooth generally convex or a nonsmooth strongly convex objective function $f_i : R^n \to R$ with local feasible constraints $\Omega_i \in R^n$. Then, all agents work together with their neighbors to achieve a consistent solution that optimizes the global objective function $\sum_{i=1}^{m} f_i(x)$ in a constraint set $\bigcap_{i=1}^{m} \Omega_i$. Therefore, the optimization problem is reformulated as

$$\min \bar{F}(x) = \sum_{i=1}^{m} f_i(x) \;\; \text{s.t.} \;\; x \in \bigcap_{i=1}^{m} \Omega_i \tag{11}$$

where Ω_i is a closed convex set for any i, and $x \in R^n$ is the optimization variable. Problem (11) arises in many areas, such as signal processing, resource scheduling, and wireless communication network.

Moreover, the local constraints are unavoidable due to the agents' limitations in computation performance and communication capabilities.

Assumption 1: The function f_i is Lipschitz continuous on constrained set Ω_i $\forall\, i \in 1, \ldots, m$, i.e.,

$$\|f_i(u) - f_i(v)\| \leq \mathfrak{l}_i \|u - v\|, i = 1, \ldots, m \tag{12}$$

with a positive constant $\mathfrak{l}$.

Assumption 2: f_i is nonsmooth generally convex or strongly convex, i.e., it satisfies conditions (6) or (7), respectively.

Assumption 3: The graph is undirected and connected.

B. Reformulation

Note that problem (11) is a centralized structure. By utilizing Assumption 3, problem (11) is equivalent to

$$\begin{aligned} &\min \;\; F(x) = \sum_{i=1}^{m} f_i(x_i) \\ &\text{s.t.} \;\; x_i \in \Omega_i \in R^n, x_i = x_j, i \in \mathcal{V}_i, j \in \mathcal{N}_i \end{aligned} \tag{13}$$

where $\mathcal{N}_i$ represents the neighbor set of agent i. Furthermore, based on the penalty method, (13) becomes

$$\begin{aligned} &\min \;\; \sum_{i=1}^{m} f_i(x_i) + \frac{\Upsilon}{2} \sum_{i=1}^{m} \sum_{j\in\mathcal{N}_i} |x_i - x_j| \\ &\text{s.t.} \;\; x_i \in \Omega_i, i \in \mathcal{V}_i \end{aligned} \tag{14}$$

where $|\cdot|$ is the 1-norm, and $\Upsilon > 0$ is a penalty parameter.

Lemma 5 (Sufficient Condition): If Assumptions 1 and 3 hold and $\Upsilon \geq \sqrt{m} \max_{1\leq i\leq m} \{\mathfrak{l}_i\}$; then, $x^* = \text{col}(x_1^*, \ldots, x_n^*)$ is an optimal solution of problem (11) and (13) if and only if x^* is also the optimal solution of (14).

Proof: Let $\bar{x} = (1/m) \sum_{i=1}^{m} x_i$ and $D(x)^2 = \sum_{i=1}^{m} \|x_i - \bar{x}\|^2 \leq (1/m) \sum_{i=1}^{m} \sum_{j=1}^{m} \|x_i - x_j\|^2 \leq (1/m) \sum_{i=1}^{m} \sum_{j=1}^{m} |x_i - x_j|^2$. Moreover, for any $p, q \in \mathcal{V}$, there exists a path $\mathcal{P}_{pq} \subset \mathcal{E}$ because Assumption 3 holds, such that

$$\begin{aligned} H(x) &= \frac{1}{2} \sum_{i=1}^{m} \sum_{j\in\mathcal{N}_i} |x_i - x_j| = \frac{1}{2} \sum_{(p,q)\in\mathcal{E}} |x_p - x_q| \\ &\geq \frac{1}{2} \sum_{(p,q)\in\mathcal{P}_{ij}} |x_p - x_q| \geq |x_i - x_j|. \end{aligned} \tag{15}$$

Furthermore, we have $D(x)^2 \leq mh(x)^2 \Rightarrow D(x) \leq \sqrt{m}H(x)$.

Let $\Upsilon \geq \sqrt{m} \max_{1\leq i\leq m} \{\mathfrak{l}_i\}$, thus

$$\begin{aligned} F(x) + \Upsilon H(x) &\geq F(x) + \max_{1\leq i\leq m} \{l_i\} D(x) \\ &= F(\bar{x}) + F(x) - F(\bar{x}) + \max_{1\leq i\leq m} \{l_i\} D(x) \geq F(\bar{x}). \end{aligned} \tag{16}$$

The first and second inequalities are derived from the condition (15) and the Lipschitz continuous property in (12), respectively. Equation (16) implies that $F(x) + \Upsilon H(x) \geq \min_{x_i=x_j} F(\bar{x})$, i.e., $F(x) + \Upsilon H(x) = \min_{x_i=x_j} F(\bar{x})$ holds if $x_i = \bar{x}, i \in \mathcal{V}$. ■

C. Smoothing Approximation Reformulation

It follows from Lemma 5 that solving problem (14) without consensus constraints is equivalent to solving problem (13). However, problem (14) is nonsmooth, even though the function $F(x)$ is smooth, due to the existence of a nonsmooth penalty function. Smoothing approximation is a useful tool to address nonsmooth problems in optimization and has a lower computational cost than the approximation operator. By utilizing the smoothing approximation method in (8), the smoothing problem of (14) is given as follows:

$$\begin{aligned} \min \ & \hat{\mathcal{F}}(x,\mu) \\ & = \sum_{i=1}^{m}\hat{f}_i(x_i,\mu_i) + \frac{\Upsilon}{2}\sum_{i=1}^{m}\sum_{j\in\mathcal{N}_i}\hat{h}(x_i-x_j,\mu_i) \\ \text{s.t.} \ & x_i\in\Omega_i, i\in\mathcal{V}_i \end{aligned} \tag{17}$$

where $\hat{f}_i(x_i,\mu_i)$ is a smoothing approximation function of $f_i(x_i)$, and

$$\begin{aligned} \hat{h}(x_i-x_j,\mu_i) &= \sum_{k=1}^{n}\hat{h}_k(x_i-x_j,\mu_i) \\ &= \begin{cases} |x_{i,k}-x_{j,k}|, & \text{if } |x_{i,k}-x_{j,k}|>\mu_i \\ \frac{(x_{i,k}-x_{j,k})^2}{2\mu_{i,k}}+\frac{\mu_{i,k}}{2}, & \text{if } |x_{i,k}-x_{j,k}|\le\mu_i \end{cases} \\ i &= 1,\ldots,m, \qquad k=1,\ldots,n. \end{aligned}$$

By 2) in Definition 1, one has $\lim_{\mu_i\to 0^+}\hat{f}_i(x_i,\mu_i)=f_i(x_i)$, $\lim_{\mu_i\to 0^+}\hat{h}(x_i-x_j,\mu_i)=|x_i-x_j|$.

IV. Main Results[1]

In this section, two DSPNAs are proposed to address the problem (14). For problem (14) with nonsmooth generally convex objective functions, DSPNA-N based on continuous variant of Nesterov's acceleration is proposed with a fast arithmetic convergence rate $\mathcal{O}(1/t^2)$. Then, we present DSPNA-P on account of continuous variant of Polyak's heavy ball method for (14) with nonsmooth, strongly convex objective functions, which can reach an explicit exponential convergence rate.

A. DSPNA-N for Problem (14) With Nonsmooth Generally Convex Objective Functions

For problem (14) with nonsmooth generally convex objective functions, we present the following DSPNA-N:

$$\begin{cases} \dot{x}_i = \frac{\alpha}{t}\left(\Pi_{\Omega_i}(y_i)-x_i\right) \\ \dot{y}_i = -\frac{t}{\alpha}\nabla_x\hat{f}_i(x_i,\mu_i) \\ \qquad -\frac{t}{\alpha}\Upsilon\sum_{j\in\mathcal{N}_i}\frac{1}{2}\nabla_x\hat{h}(x_i-x_j,\mu_i) \\ x_{i,0}\in\Omega_i, \quad i=1,\ldots,m \end{cases} \tag{18}$$

where

$$\nabla_x\hat{f}_i(x_i,\mu_i)=\sum_{k=1}^{n}\nabla_x\hat{f}_i(x_{i,k},\mu_{i,k})$$

$$\nabla_x\hat{h}(x_i-x_j,\mu_i)=\sum_{k=1}^{n}\nabla_x\hat{h}_k(x_{i,k}-x_{j,k},\mu_{i,k})$$

$$\begin{aligned} &\nabla_x\hat{h}_k(x_{i,k}-x_{j,k},\mu_i) \\ &= \begin{cases} \operatorname{sgn}(x_{i,k}-x_{j,k}), & \text{if } |x_{i,k}-x_{j,k}|>\mu_{i,k} \\ \frac{x_{i,k}-x_{j,k}}{\mu_{i,k}}, & \text{if } |x_{i,k}-x_{j,k}|\le\mu_{i,k}. \end{cases} \end{aligned}$$

The compact form of (18) is given by

$$\begin{cases} \dot{x} = \frac{\alpha}{t}\left(\Pi_{\bar{\Omega}}(y)-x\right) \\ \dot{y} = -\frac{t}{\alpha}\left(\nabla_x\hat{F}(x,\mu)+\Upsilon\nabla_x\hat{H}(x,\mu)\right) \\ x_0\in\bar{\Omega} \end{cases} \tag{19}$$

where $\nabla_x\hat{F}(x,\mu)=\sum_{i=1}^{m}\hat{f}(x_i,\mu_i)$, $\bar{\Omega}=\Omega_1\times\cdots\times\Omega_m$ and $\nabla_x\hat{H}(x,\mu)=(1/2)\sum_{i=1}^{m}\sum_{j\in\mathcal{N}_i}\nabla_x\hat{h}(x_i-x_j,\mu_i)$.

Theorem 1: For any given initial values $x_0\in\bar{\Omega}$, $y_0\in R^{mn}$, DSPNA-N (19) has a unique strong global solution if the smoothing gradient $\nabla_x\hat{F}(x,\mu)$ is Lipschitz continuous. In addition, the solution x is always in the closed convex set $\bar{\Omega}$ if $x_0\in\bar{\Omega}$.

Proof: Existence and Uniqueness: In phase space $R^{mn}\times R^{mn}$, the DSPNA-N (19) can be equivalently described as

$$\dot{Z}=\mathcal{F}(t,Z) \tag{20}$$

where $Z=(x,y)$

$$\begin{aligned} \mathcal{F}(t,Z) = \Big(&\frac{\alpha}{t}\left(\Pi_{\bar{\Omega}}(y)-x\right) \\ &-\frac{t}{\alpha}\left(\nabla_x\hat{F}(x,\mu)+\nabla_x\hat{H}(x,\mu)\right)\Big). \end{aligned}$$

By the Cauchy–Lipschitz–Picard theorem, we derive the existence and uniqueness of the solution to (19) as follows.

1) Let $t\in[t_0,+\infty)$ be fixed and consider the pair (x,y) and $(\bar{x},\bar{y})$ in $R^{mn}\times R^{mn}$. Then, by the Lipschitz continuity property of function $\nabla_x\hat{F}(x,\mu)$ and inequality $\|\mathfrak{a}+\mathfrak{b}\|^2\le 2\mathfrak{a}^2+2\mathfrak{b}^2 \ \ \forall\mathfrak{a},\mathfrak{b}\in R^{mn}$, one has

$$\begin{aligned} &\|\mathcal{F}(t,x,y)-\mathcal{F}(t,\bar{x},\bar{y})\| \\ &\le \Bigg(2\left(\frac{\alpha}{t}\right)^2\|y-\bar{y}\|^2+2\left(\frac{\alpha}{t}\right)^2\|x-\bar{x}\|^2+2\left(\frac{t}{\alpha}\right)^2 \\ &\quad\times\Bigg(l^2+\Bigg(\frac{\Upsilon m^2n^2}{2\min\limits_{1\le i\le m,1\le k\le n}\{\mu_{i,k}\}}\Bigg)^2\Bigg)\|x-\bar{x}\|^2\Bigg)^{\frac{1}{2}} \\ &\le \Bigg(4\left(\frac{\alpha}{t}\right)^2+2\left(\frac{t}{\alpha}\right)^2\Bigg(l^2+\Bigg(\frac{\Upsilon m^2n^2}{2\min\limits_{1\le i\le m,1\le k\le n}\{\mu_{i,k}\}}\Bigg)^2\Bigg)\Bigg)^{\frac{1}{2}} \\ &\quad\times\|x-\bar{x},y-\bar{y}\|. \end{aligned} \tag{21}$$

Denote

$$L(t)=\Bigg[4\left(\frac{\alpha}{t}\right)^2+2\left(\frac{t}{\alpha}\right)^2\Bigg(l^2+\Bigg(\frac{\Upsilon n^2}{2\min\limits_{1\le i\le m,1\le k\le n}\{\mu_{i,k}\}}\Bigg)^2\Bigg)\Bigg]^{\frac{1}{2}}$$

where l is the Lipschitz constant of $\nabla_x\hat{F}(x,\mu)$. We have

$$\|\mathcal{F}(t,x,y)-\mathcal{F}(t,\bar{x},\bar{y})\|\le L(t)\|x-\bar{x},y-\bar{y}\|.$$

Note that $L(t)$ is continuous on $[t_0,+\infty)$; hence, $L(t)$ is integrable on $(0,T)$ for all $0<T<+\infty$.

[1]For convenience, the variables $x(t)$, $y(t)$, and $\mu(t)$ are simply marked as x, y, and μ in the neurodynamic approaches and the proof.

2) Next, we will show that $\mathcal{F}(\cdot, x, y) \in L^1((0,+\infty), R^{mn} \times R^{mn})$ for any $x, y \in R^{mn}$. For any $T > 0$, we obtain the following estimate:

$$\begin{aligned}
&\int_0^{T}\|\mathcal{F}(t,x,y)\|dt \\
&\leq \int_0^T \Bigg(2\Big(\frac{\alpha}{t}\Big)^2\|\Pi_{\bar\Omega}(y)\|^2 + 2\Big(\frac{\alpha}{t}\Big)^2\|x\|^2 + 2\Big(\frac{t}{\alpha}\Big)^2 \\
&\qquad \times \|\nabla_x \hat F(x,\mu)\|^2 + 2\Big(\frac{t}{\alpha}\Big)^2 \frac{m^3n^3\|x\|^2}{\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}^2}\Bigg)^{\frac12} dt \\
&\leq \int_0^T \Bigg(2\Big(\frac{\alpha}{t}\Big)^2\|\Pi_{\bar\Omega}(y)\|^2 + 2\Big(\frac{t}{\alpha}\Big)^2\|\nabla_x \hat F(x,\mu)\|^2 \\
&\qquad + 2\Bigg(\Big(\frac{\alpha}{t}\Big)^2 + \Big(\frac{t}{\alpha}\Big)^2\frac{m^3n^3\|x\|^2}{\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}^2}\Bigg)\|x\|^2\Bigg)^{\frac12} dt \\
&\leq \Big(\|x\|^2 + \|\Pi_{\bar\Omega}(y)\|^2 + \|\nabla_x \hat F(x,\mu)\|^2\Big)^{\frac12} \\
&\qquad \times \int_0^T \Bigg(4\Big(\frac{\alpha}{t}\Big)^2 + 2\Big(\frac{t}{\alpha}\Big)^2\Bigg(\frac{m^3n^3\|x\|^2}{\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}^2} + 1\Bigg)\Bigg)^{\frac12} dt
\end{aligned}$$

where the first inequality is derived from $\nabla_x \hat H(x,\mu) \leq ([n^{3/2}m^{3/2}\|x\|]/[\min\limits_{1\le i\le m,1\le k\le n}\{\mu_{i,k}\}])$ and $\|\mathfrak{a} + \mathfrak{b}\|^2 \leq 2\mathfrak{a}^2 + 2\mathfrak{b}^2 \quad \forall \mathfrak{a}, \mathfrak{b} \in R^{mn}$. The conclusion is satisfied since

$$t \to \Bigg(4\Big(\frac{\alpha}{t}\Big)^2 + 2\Big(\frac{t}{\alpha}\Big)^2\Bigg(\frac{m^3n^3\|x\|^2}{\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}^2} + 1\Bigg)\Bigg)^{\frac12}$$

is continuous.

Thus, the existence and uniqueness of the solutions of DSPNA-N (19) hold from the Cauchy–Lipschitz–Picard theorem.

Feasibility: Next, the solution $x(t)$ is always in the closed convex set $\bar\Omega$ for any $t > 0$. From the first equation in (19) $\dot x(t) = (\alpha/t)(\Pi_{\bar\Omega}(y(t)) - x(t))$, we have $(d/dt)(t^\alpha x(t)) = \alpha t^{\alpha-1}\Pi_{\bar\Omega}(y(t))$. Integrating the inequality above from 0 to t, one has $t^\alpha x(t) = \int_0^t \alpha s^{\alpha-1}\Pi_{\bar\Omega}(y(s))ds$ and it can be equivalently written as $x(t) = \int_0^t (\alpha s^{\alpha-1}/t^\alpha)\Pi_{\bar\Omega}(y(s))ds$. Since $(\alpha s^{\alpha-1}/t^\alpha) > 0$, for any $t > 0$, $\Pi_{\bar\Omega}(y(t)) \in \bar\Omega$, $0 < s \leq t$ and $\int_0^t (\alpha s^{\alpha-1}/t^\alpha)ds = 1$ and $\bar\Omega$ is a nonempty, closed, and convex set, we have $x(t) = \int_0^t ([\alpha s^{\alpha-1}\Pi_{\bar\Omega}(y(t))]/t^\alpha)ds \in \bar\Omega$, since $t \in (0,+\infty)$ is arbitrary, which concludes that $x(t) \in \bar\Omega$, $t \in (0, t)$. Combining the above results with the condition $x_0 \in \bar\Omega$, the conclusion can be drawn. ■

Theorem 2: Let $x^* = \mathrm{col}(x_1^*, \ldots, x_n^*)$ be an optimal strong global solution of NSPNA-N (19) and $\alpha \geq 2$, $\mu_{i,k} \geq 0$, $\dot\mu_{i,k} < 0$, $\int_0^{+\infty} t\mu_{i,k}dt < +\infty$, $i = 1,\ldots,m$, $k = 1,\ldots,n$; there exists a constant $\mathcal{C}$, such that:

1) the trajectory of $(t\dot x, x) \quad \forall\, t > 0$ is bounded;
2) DSPNA-N (19) has a faster arithmetic convergence rate

$$\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - \big(F(x^*) + \Upsilon H(x^*)\big) \leq \frac{\mathcal{C}}{t^2} \quad \forall x \in \bar\Omega$$

and $\|\dot x\| = \mathcal{O}(1/t)$.

Proof:

1) Designing a Lyapunov function on $\bar\Omega$ as

$$\begin{aligned}
V(t,x,y) &= t^2\Big(\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - F(x^*) - \Upsilon H(x^*)\Big) \\
&\quad + t^2\kappa_{\hat\Gamma}\mathbf{1}^T\mu + \alpha\Big(\|y - \Pi_{\bar\Omega}(y^*)\|^2 \\
&\qquad - \|y - \Pi_{\bar\Omega}(y)\|^2\Big), x \in \bar\Omega
\end{aligned}$$

where $\hat F + \Upsilon\hat H = \hat\Gamma$, $x^* = \Pi_{\bar\Omega}(y^*)$ and $V(t,x,y) \geq \alpha\|\Pi_{\bar\Omega}(y) - \Pi_{\bar\Omega}(y^*)\|^2$, which can be obtained from the first equality in (19) and property 6) in Definition 1 and Lemma 2. Taking the derivative of $V(t,x,y)$ in time and along the trajectory of DSPNA-N (19), we have

$$\begin{aligned}
&\dot V(t,x,y) \\
&= \frac{2t}{\alpha}\Big(\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - \big(F(x^*) + \Upsilon H(x^*)\big)\Big) \\
&\quad + t\Big(\nabla_x\hat F(x,\mu) + \Upsilon\nabla_x\hat H(x,\mu)\Big)^T\big(\Pi_{\bar\Omega}(y) - x\big) \\
&\quad - t\big(\Pi_{\bar\Omega}(y) - \Pi_{\bar\Omega}(y^*)\big)^T\Big(\nabla_x\hat F(x,\mu) + \Upsilon\nabla_x\hat H(x,\mu)\Big) \\
&\quad + t^2\Big(\nabla_\mu\hat F(x,\mu) + \Upsilon\nabla_\mu\hat H(x,\mu) + \kappa_{\hat\Gamma}\mathbf{1}\Big)^T\dot\mu + 2t\kappa_{\hat\Gamma}\mathbf{1}^T\mu \\
&\leq \frac{2t}{\alpha}\Big(\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - \big(F(x^*) + \Upsilon H(x^*)\big)\Big) \\
&\quad + t\Big(\nabla_x\hat F(x,\mu) + \Upsilon\nabla_x\hat H(x,\mu)\Big)^T\big(x^* - x\big) + 2t\kappa_{\hat\Gamma}\mathbf{1}^T\mu \\
&\leq \frac{2t}{\alpha}\Big(\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - \big(F(x^*) + \Upsilon H(x^*)\big)\Big) \\
&\quad + 2t\kappa_{\hat\Gamma}\mathbf{1}^T\mu + \frac{\alpha t}{\alpha}\Big(\hat F(x^*,\mu) + \Upsilon\hat H(x^*,\mu)\Big) \\
&- \frac{\alpha t}{\alpha}\Big(\hat F(x,\mu) + \Upsilon\hat H(x,\mu)\Big) \\
&= \Big(\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - F(x^*) - \Upsilon H(x^*) + \kappa_{\hat\Gamma}\mathbf{1}^T\mu\Big) \\
&\quad \times \frac{(2-\alpha)t}{\alpha} - \frac{(2-\alpha)t}{\alpha}\kappa_{\hat\Gamma}\mathbf{1}^T\mu + 3t\kappa_{\hat\Gamma}\mathbf{1}^T\mu + \frac{\alpha t}{\alpha} \\
&\quad \times \Big(\hat F(x^*,\mu) + \Upsilon\hat H(x^*,\mu) - \kappa_{\hat\Gamma}\mathbf{1}^T\mu \\
&\qquad - F(x^*) - \Upsilon H(x^*)\Big) \\
&\leq \Big(4 - \frac{2}{\alpha}\Big)t\kappa_{\hat\Gamma}\mathbf{1}^T\mu
\end{aligned} \tag{22}$$

where the first inequality holds due to $\dot\mu_{i,k} < 0$, $i = 1,\ldots,m$, $k = 1,\ldots,n$ and $|\nabla_\mu\hat F(x,\mu) + \Upsilon\nabla_\mu\hat H(x,\mu)| \leq \kappa_{\hat\Gamma}$, which comes from 3) in Definition 1. The second inequality holds from the convexity of $\hat F(x,\mu) + \Upsilon\hat H(x,\mu)$. The last inequality holds from $\alpha \geq 2$ and the following inequalities:

$$\begin{aligned}
&1)\ |\hat F(x,\mu) + \Upsilon\hat H(x,\mu) - (F(x) + \Upsilon H(x))| \leq \kappa_{\hat\Gamma}\mathbf{1}^T\mu; \\
&2)\ \hat F(x,\mu) + \Upsilon\hat H(x,\mu) + \kappa_{\hat\Gamma}\mathbf{1}^T\mu - F(x^*) - \Upsilon H(x^*); \\
&\qquad \geq F(x) + \Upsilon H(x) - \big(F(x^*) + \Upsilon H(x^*)\big) \geq 0 \\
&3)\ |\hat F(x^*,\mu) + \Upsilon\hat H(x^*,\mu) - \big(F(x^*) + \Upsilon H(x^*)\big)| \leq \kappa_{\hat\Gamma}\mathbf{1}^T\mu.
\end{aligned} \tag{23}$$

Therefore, $\dot V(t,x,y) \leq (4 - (2/\alpha))t\kappa_{\hat\Gamma}\mathbf{1}^T\mu$. Then, integrating both sides of it from 0 to $+\infty$ and combining the condition $\mu_{i,k} \geq 0$, $\dot\mu_{i,k} < 0$ and $\int_0^{+\infty} t\mu_{i,k}dt < +\infty$,

$i = 1, \ldots, m, k = 1, \ldots, n$, we deduce that the positive part $[\dot{V}(t, x, y)]^+$ of $\dot{V}(t, x, y)$ belongs to $L^1 \in (0, +\infty)$. Since $V(t, x, y)$ is bounded from below, it follows that $V(t, x, y)$ has a limit $\mathcal{C}$ as $t \to +\infty$, and hence, $\mathcal{C}$ is bounded, i.e., $V(t, x, y) \le \mathcal{C} < +\infty$. Note that $\bar{\Omega}$ is a closed convex set, $x \in \bar{\Omega}\ \ \forall t$ in *Feasibility* implies that x is bounded. Moreover, since $\alpha\|\Pi_{\bar{\Omega}}(y) - x^*\|^2 = \alpha\|(t/\alpha)\dot{x} + x - x^*\|^2 \le V(t, x, y) \le \mathcal{C} < +\infty$, which combines with the boundedness of $x(t)$ to yield the boundedness of $t\dot{x}$.

2) Since $t^2(\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F(x^*) - \Upsilon H(x^*)) \le V(t, x, y) \le \mathcal{C}\ \ \forall x \in \bar{\Omega}$, which implies that $\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F(x^*) - \Upsilon H(x^*) \le (\mathcal{C}/t^2)\ \ \forall x \in \bar{\Omega}$. In addition, $\alpha\|(t/\alpha)\dot{x} + x - x^*\|^2 \le \mathcal{C} < +\infty$ in 1) implies $(t^2/\alpha)\|\dot{x}\|^2 \le \mathcal{C} + \alpha\|x - x^*\|^2$. Since $x(t)$ is bounded, there exists $\mathcal{A} = \mathcal{C} + \alpha\sup_x\|x - x^*\|^2 < +\infty$, such that $\|\dot{x}\|^2 \le ([\alpha(\mathcal{C} + \alpha\|x - x^*\|^2)]/t^2) \le (\alpha\mathcal{A}/t^2)$, i.e., $\|\dot{x}\| = \mathcal{O}(1/t)$. Thus, the proof is completed. ∎

Remark 1: The Lyapunov function $V(t, x, y)$ in Theorem 2 is constructed on feasible set $\bar{\Omega}$, which is feasible due to $x \in \bar{\Omega}$ introduced in *Feasibility* of Theorem 1. In addition, the conditions $\mu_{i,k} \ge 0$, $\dot{\mu}_{i,k} < 0$, $\int_0^{+\infty} t\kappa_{\hat{\Gamma}}\mu_{i,k}dt < +\infty, i = 1, \ldots, m, k = 1, \ldots, n$ imply $\mu_{i,k} = t^{-\gamma}, \gamma > 2$. Furthermore, from inequality (22), we have $V(t, x, y) \le V(0, x_0, y_0) + \int_0^t (4 - [2/\alpha])s\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu ds = C < +\infty$; thus, the result in Theorem 2 holds and $t^2(\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F(x^*) - \Upsilon H(x^*) + \kappa_{\hat{\Gamma}}\mathbf{1}^T\mu) \le \mathcal{C}$ is also satisfied. By using inequality 2) in (23), we have $F(x) + \Upsilon H(x) - F(x^*) - \Upsilon H(x^*) \le (\mathcal{C}/t^2)$.

Note that according to $\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - f(x^*) - \Upsilon H(x^*) \le (\mathcal{C}/t^2)\ \ \forall\ x \in \bar{\Omega}$ in Theorem 3, we have $\lim_{t\to+\infty}\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F(x^*) - \Upsilon H(x^*) = \mathbf{0}, \lim_{t\to+\infty}\mu = \mathbf{0}\ \ \forall\ x \in \bar{\Omega}$. Since $\hat{F}(x, \mu)$ with respect to x is convex with any fixed $\mu > 0$, NSPNA-N (19) converges to the optimal solution as the time goes to infinity.

B. DSPNA-P for Problem (14) With Nonsmooth Strongly Convex Objective Functions

In this section, to address problem (13) with nonsmooth strongly convex objective functions with an explicit exponential convergent rate, we propose a DSPNA-P based on continuous variant of Polyak's heavy ball method as follows:

$$\begin{cases} \dot{x}_i = \beta\big(\Pi_{\Omega_i}(y_i) - x_i\big) \\ \dot{y}_i = \beta(x_i - y_i) \\ \qquad -\frac{\beta}{\eta}\Big(\nabla_x\hat{f}_i(x_i, \mu_i) + \Upsilon\sum_{j\in\mathcal{N}_i}\frac{1}{2}\nabla_x\hat{h}\big(x_i - x_j, \mu_i\big)\Big) \\ x_{i,0} \in \Omega_i, i = 1, \ldots, m \end{cases} \tag{24}$$

where $\beta > 0$, and the compact form of DSPNA-P (24) is given by

$$\begin{cases} \dot{x} = \beta\big(\Pi_{\bar{\Omega}}(y) - x\big) \\ \dot{y} = \beta\Big(x - y - \frac{1}{\eta}\Big(\nabla_x\hat{F}(x, \mu) + \Upsilon\nabla_x\hat{H}(x, \mu)\Big)\Big) \\ x_0 \in \bar{\Omega}. \end{cases} \tag{25}$$

Theorem 3: Under Assumptions 1 and 3, if the conditions $\lim\limits_{t\to+\infty}\mu = \mathbf{0}$, $x \in \bar{\Omega}$ hold, and $x^* = \mathrm{col}(x_1^*, \ldots, x_m^*), y^* = \mathrm{col}(y_1^*, \ldots, y_m^*) \in R^{mn}$ is an equilibrium point of DSPNA-P (25), then $x^* = \mathrm{col}(x_1^*, \ldots, x_m^*) \in R^{mn}$ of DSPNA-P (25) is an optimal solution of problem (14).

Proof: Based on Karush–Kuhn–Tucker (KKT) conditions, $x^* = \mathrm{col}(x_1^*, \ldots, x_m^*)$ is an optimal solution of problem (14) if and only if it fulfills the following condition:

$$\partial f_i\big(x_i^*\big) + \Upsilon\sum_{j\in\mathcal{N}_i}\mathrm{sgn}\Big(x_i^* - x_j^*\Big) + \mathcal{N}_{\Omega_i}\big(x_i^*\big) \ni \mathbf{0}$$
$$i = 1, \ldots, m \tag{26}$$

where $\mathrm{sgn}(x_i^* - x_j^*) = \sum_k^n \mathrm{sgn}(x_{i,k}^* - x_{j,k}^*) \in \partial|x_i^* - x_j^*|$ and $\mathcal{N}_{\Omega_i}$ is a normal cone of set Ω_i defined in Lemma 1.

Let (x^*, y^*) be an equilibrium point of DSPNA-P (25). In what follows, we will show that x^* satisfies the KKT conditions (26). From the first equation in DSPNA-P (25), we have $\Pi_{\bar{\Omega}}(y^*) = \Pi_{\bar{\Omega}}(x^*) = x^* \in \bar{\Omega}$. By using 2) in Lemma 1, we have $\eta(y^* - \Pi_{\bar{\Omega}}(y^*))^T(\mathcal{Y} - \Pi_{\bar{\Omega}}(y^*)) = \eta(y^* - x^*)^T(\mathcal{Y} - x^*) \le 0\ \ \forall\mathcal{Y} \in \Omega$ with $\eta > 0$, which implies $\eta(y^* - x^*) \in \mathcal{N}_{\bar{\Omega}}(x^*)$. Moreover, from the second equation in DSPNA-P (25), we also obtained that $\mathbf{0} = \eta(x^* - y^*) - (\nabla_x\hat{F}(x^*, \mu^*) + \Upsilon\nabla_x\hat{H}(x^*, \mu^*)) \in \eta(x^* - y^*) - (\partial F(x^*) + \Upsilon\partial H(x^*))$ from 4) in Definition 1. It combines with $\eta(y^* - x^*) \in \mathcal{N}_{\bar{\Omega}}(x^*)$, we get $\mathbf{0} \in \partial F(x^*) + \Upsilon\partial H(x^*) + \mathcal{N}_{\bar{\Omega}}(x^*)$. Therefore, the proof is completed. ∎

Theorem 4: For any given initial values $x_0 \in \bar{\Omega}, y_0 \in R^{mn}$, the DSPNA-P (25) has an unique solution if the smoothing function $\nabla_x\hat{F}(x, \mu) + \Upsilon\nabla_x\hat{H}(x, \mu)$ is Lipschitz continuous in regard to x over $\bar{\Omega}$. Moreover, the solution x is always in the closed convex set $\bar{\Omega}$.

Proof Existence and Uniqueness: The DSPNA-P (25) can be equivalently described as $\dot{Y} = G(Y)$, where $Y = (x^T, y^T)^T$

$$G(Y) = \begin{pmatrix} \beta\big(\Pi_{\bar{\Omega}}(y) - x\big) \\ \beta(x - y) - \frac{\beta}{\eta}\Big(\nabla_x\hat{F}(x, \mu) + \nabla_x\hat{H}(x, \mu)\Big) \end{pmatrix}.$$

Let $\overline{Y}$, $\underline{Y} \in R^{nm}$ be two different solutions to the DSPNA-P (25) with the same initial state $Y_0 \in R^{nm}$. Then, there exist $\hat{t} > 0$ and $\delta > 0$, such that $\overline{Y}(t) \ne \underline{Y}(t)$ for any $t \in [\hat{t}, \hat{t} + \delta]$. Since $\overline{Y}(t)$, $\underline{Y}(t)$, and $\mu(t) > 0$ are continuous and bounded on $t \in [0, \hat{t} + \delta]$, there exists $\mathsf{l} = 4\beta + (\beta/\eta)l + ([m^2n^2\Upsilon\beta]/[2\min_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}\eta]) > 0$, such that

$$\begin{aligned}
&\|G\big(\overline{Y}\big) - G\big(\underline{Y}\big)\| \\
&\le \beta\|\Pi_{\bar{\Omega}}(\bar{y}) - \Pi_{\bar{\Omega}}\Big(\underline{y}\Big)\| + 2\beta\|\bar{x} - \underline{x}\| + \beta\|\bar{y} - \underline{y}\| \\
&\quad + \frac{\beta}{\eta}l\|\bar{x} - \underline{x}\| + \frac{\beta}{\eta}\frac{\frac{1}{2}m^2n^2\Upsilon}{\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}}\|\bar{x} - \underline{x}\| \\
&\le 2\beta\|\bar{y} - \underline{y}\| + \|\bar{x} - \underline{x}\|\left(2\beta + \frac{\beta l}{\eta} + \frac{\beta}{\eta}\frac{\frac{1}{2}m^2n^2\Upsilon}{\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}}\right) \\
&\le \left(4\beta + \frac{\beta}{\eta}l + \frac{m^2n^2\Upsilon\beta}{2\min\limits_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}\eta}\right)\|\overline{Y} - \underline{Y}\| \\
&= \mathsf{l}\|\overline{Y} - \underline{Y}\|\quad \forall t \in \big[0, \hat{t} + \delta\big]
\end{aligned}$$

where $\mathsf{l} = 4\beta + (\beta/\eta)l + ([m^2n^2\Upsilon\beta]/[2\min_{1\le i\le m, 1\le k\le n}\{\mu_{i,k}\}\eta])$. It follows that:

$$\frac{d}{dt}\frac{1}{2}\|\overline{Y} - \underline{Y}\|^2 = \big(\overline{Y} - \underline{Y}\big)^T\left(\frac{d}{dt}\overline{Y} - \frac{d}{dt}\underline{Y}\right)$$

$$= \left(\overline{Y} - \underline{Y}\right)^T \left(G(\tilde{Y}) - G(Y)\right)$$
$$\leq l\|\overline{Y} - \underline{Y}\|^2, t \in \left[0, \hat{t} + \delta\right].$$

Integrating the above inequality from 0 to $t \in (0, \hat{t}+\delta]$ and then applying Gronwall's inequality, one has $\overline{Y}(t) = \underline{Y}(t) \quad \forall t \in [0, \hat{t} + \delta]$, which leads to a contradiction.

Feasibility: Next, the solution x of approach (25) that is always in the nonempty, closed, and convex set $\bar{\Omega}$ for any $t \geq 0$ will be discussed. From the first equation in (25), we have $\dot{x}(t) = \beta(\Pi_{\bar{\Omega}}(y(t)) - x(t))$, which is equivalent to $(d/dt)(e^{\beta t}x(t)) = \beta e^{\beta t}\Pi_{\bar{\Omega}}(y(t))$. Integrating the both sides of the inequality mentioned above from 0 to t and then dividing it by $e^{\beta t}$, we derive that $x(t) = (1/e^{\beta t})x(t_0) + \int_0^t (\beta e^{\beta s}/e^{\beta t})\Pi_{\bar{\Omega}}(y(s))ds$ and it can be equivalently written as $(1/e^{\beta t})x_0 + (1 - (1/e^{\beta t})) \int_0^t (\beta e^{\beta s}/e^{\beta t} - 1)\Pi_{\bar{\Omega}}(y(s))ds$. Since $(\beta e^{\beta s}/e^{\beta t} - 1) > 0$, $\Pi_{\bar{\Omega}}(y(s)) \in \bar{\Omega} \quad \forall\, 0 \leq s \leq t$, $\int_0^t (\beta e^{\beta s}/e^{\beta t} - 1)ds = 1$ and $\bar{\Omega}$ is a nonempty, closed, and convex set, we have $\int_0^t (\beta e^{\beta s}/e^{\beta t} - 1)\Pi_{\bar{\Omega}}(y(s))ds \in \bar{\Omega} \quad \forall t \in [0, +\infty)$. Since $x_0 \in \bar{\Omega}$, we further obtain $x(t) \in \bar{\Omega} \quad \forall t \in [0, +\infty)$. Therefore, the proof is completed. ■

Theorem 5: Suppose that the smoothing function $\hat{F}(x, \mu)$ of $F(x)$ is a strongly convex function and that $\mu_{i,k} \geq 0$, $\dot{\mu}_{i,k} < 0$, $\int_0^{+\infty} e^{\beta t}\kappa_{\hat{\Gamma}}\mu_{i,k}dt < +\infty, i = 1, \ldots, m, k = 1, \ldots, n$ and $\Pi_{\bar{\Omega}}(y^*) = x^*$, $y^* \in R^{mn}$ is the optimal solution to problem (13) with strongly convex objective functions. Then, under the initial values $x_0 \in \bar{\Omega}$, $y_0 \in R^{mn}$, one has:

1) The trajectory of $(\dot{x}, x) \quad \forall\, t > 0$ of DSPNA-P (25) is bounded;
2) For any $t \geq 0$ and $\forall x \in \bar{\Omega}$, DSPNA-P (25) has the following exponential convergence rate:

$$\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F\left(x^*\right) - \Upsilon H\left(x^*\right) \leq \frac{\mathcal{B}}{e^{\beta t}},$$
$$\|x - x^*\| \leq \frac{2\sqrt{\mathcal{B}}}{\eta e^{\frac{\beta}{2}t}} \quad (27)$$

where $\mathcal{B}$ is a positive constant.

Proof:

1) A novel Lyapunov function on set $\bar{\Omega}$ is established as follows:

$$E(t, x, y) = e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu + e^{\beta t}\left(\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F\left(x^*\right) - \Upsilon H\left(x^*\right)\right) + \frac{\eta}{2}\left(\|y - \Pi_{\bar{\Omega}}\left(y^*\right)\|^2 - \|y - \Pi_{\bar{\Omega}}(y)\|^2\right). \quad (28)$$

The derivative of $E(t, x, y)$ along (25) satisfies

$$\begin{aligned}\dot{E}(t, x, y) &= \beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu + \eta\beta e^{\beta t}\left(\Pi_{\bar{\Omega}}(y) - \Pi_{\bar{\Omega}}\left(y^*\right)\right)^T \\ &\quad \times \left(x - y - \frac{1}{\eta}\left(\nabla_x\hat{F}(x, \mu) + \Upsilon\nabla_x\hat{H}(x, \mu)\right)\right) \\ &\quad + \left(\kappa_{\hat{\Gamma}}\mathbf{1} + \left(\nabla_\mu\hat{F}(x, \mu) + \Upsilon\nabla_\mu\hat{H}(x, \mu)\right)\right)^T\dot{\mu} \\ &\quad + \beta e^{\beta t}\left(\nabla_x\hat{F}(x, \mu) + \Upsilon\nabla_x\hat{H}(x, \mu)\right)^T\left(\Pi_{\bar{\Omega}}(y) - x\right) \\ &\quad + \beta e^{\beta t}\left(\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F\left(x^*\right) - \Upsilon H\left(x^*\right)\right) \\ &\quad + \beta e^{\beta t}\left(\frac{\eta}{2}\|y - \Pi_{\bar{\Omega}}\left(y^*\right)\|^2 - \frac{\eta}{2}\|y - \Pi_{\bar{\Omega}}(y)\|^2\right). \quad (29)\end{aligned}$$

According to 3) in Definition 1, one has that every element of $\kappa_{\hat{\Gamma}}\mathbf{1} + \nabla_\mu\hat{F}(x, \mu) + \Upsilon\nabla_\mu\hat{H}(x, \mu)$ is greater than or equal to 0 and combining with $\dot{\mu}_{i,k} < 0, i = 1, \ldots, m, k = 1, \ldots, n$, we have $(\kappa_{\hat{\Gamma}}\mathbf{1} + (\nabla_\mu\hat{F}(x, \mu) + \Upsilon\nabla_\mu\hat{H}(x, \mu)))^T\dot{\mu} \leq 0$. Since $\Pi_{\bar{\Omega}}(y^*) = x^*$ [which can be obtained from the first equation in (25)], we have

$$\begin{aligned}\dot{E}(t, x, y) &\leq \beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu + \eta\beta e^{\beta t}\left(\Pi_{\bar{\Omega}}(y) - \Pi_{\bar{\Omega}}\left(y^*\right)\right)^T(x - y) \\ &\quad + \beta e^{\beta t}\left(\nabla_x\hat{F}(x, \mu) + \Upsilon\nabla_x\hat{H}(x, \mu)\right)^T\left(\Pi_{\bar{\Omega}}\left(y^*\right) - x\right) \\ &\quad + \beta e^{\beta t}\left(\hat{F}(x, \mu) + \Upsilon\hat{H}(x, \mu) - F\left(x^*\right) - \Upsilon H\left(x^*\right)\right) \\ &\quad + \beta e^{\beta t}\left(\frac{\eta}{2}\|y - \Pi_{\bar{\Omega}}\left(y^*\right)\|^2 - \frac{\eta}{2}\|y - \Pi_{\bar{\Omega}}(y)\|^2\right). \quad (30)\end{aligned}$$

Using the strongly convex property of smoothing function $\hat{F}(x, \mu)$

$$\begin{aligned}&\left(\nabla_x\hat{F}(x, \mu) + \Upsilon\nabla_x\hat{H}(x, \mu)\right)^T\left(x^* - x\right) \\ &\quad \leq \hat{F}\left(x^*, \mu\right) + \Upsilon\hat{H}\left(x^*, \mu\right) - \hat{F}(x, \mu) \\ &\quad -\Upsilon\hat{H}(x, \mu) - \frac{\eta}{2}\|x - x^*\|^2\end{aligned}$$

and $\hat{F}(x^*, \mu) + \Upsilon\hat{H}(x^*, \mu) - F(x^*) - \Upsilon H(x^*) \leq \kappa_{\hat{\Gamma}}\mathbf{1}^T\mu$ from 6) in Definition 1, we derive

$$\begin{aligned}\dot{E}(t, x, y) &\leq 2\beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu + \beta e^{\beta t}\eta \\ &\quad \times \left(-\frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|y - \Pi_{\bar{\Omega}}\left(y^*\right)\|^2\right. \\ &\quad - \frac{1}{2}\|y - \Pi_{\bar{\Omega}}(y)\|^2 \\ &\quad \left. + \left(\Pi_{\bar{\Omega}}(y) - \Pi_{\bar{\Omega}}\left(y^*\right)\right)^T(x - y)\right).\end{aligned}$$

By the condition $\Pi_\Omega(y^*) = x^*$, one has

$$\begin{aligned}&\left(\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\right)^T(x - y) \\ &\quad = -\|\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\|^2 + \left(\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\right)^T \\ &\quad \times \left(x - x^*\right) + \left(\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\right)^T(\Pi_\Omega(y) - y)\end{aligned}$$

such that $\dot{E}(t, x, y)$ satisfies

$$\begin{aligned}\dot{E}(t, x, y) &\leq \beta e^{\beta t}\eta \\ &\quad \times \left(-\frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|y - P_\Omega\left(y^*\right)\|^2\right. \\ &\quad - \frac{1}{2}\|y - P_\Omega(y)\|^2 - \|\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\|^2 \\ &\quad \left. + \left(\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\right)^T\left(x - x^*\right)\right) + 2\beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu \\ &\quad + \eta e^{\beta t}\eta\left(\Pi_\Omega(y) - \Pi_\Omega\left(y^*\right)\right)^T(\Pi_\Omega(y) - y).\end{aligned}$$

By adopting

$$\begin{aligned}&\frac{1}{2}\|y - \Pi_{\bar{\Omega}}\left(y^*\right)\|^2 \\ &\quad = \frac{1}{2}\|y - \Pi_{\bar{\Omega}}(y)\|^2 + \frac{1}{2}\|\Pi_{\bar{\Omega}}(y) - \Pi_{\bar{\Omega}}\left(y^*\right)\|^2 \\ &\quad + \left(y - \Pi_{\bar{\Omega}}(y)\right)^T\left(\Pi_{\bar{\Omega}}(y) - \Pi_{\bar{\Omega}}\left(y^*\right)\right)\end{aligned}$$

and

$$\left(\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*)\right)^T\left(\Pi_{\bar{\Omega}}(y^*)-x\right)$$
$$=-\frac{1}{2}\|x-\Pi_{\bar{\Omega}}(y)\|^2+\frac{1}{2}\|x-\Pi_{\bar{\Omega}}(y^*)\|^2$$
$$+\frac{1}{2}\|\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*)\|^2$$

we obtain that

$$\begin{aligned}\dot{E}(t,x,y) &\le \beta e^{\beta t}\eta\\ &\times\Big(-\frac{1}{2}\|x-x^*\|^2+\frac{1}{2}\|y-\Pi_{\bar{\Omega}}(y)\|^2\\ &-\frac{1}{2}\|y-\Pi_{\bar{\Omega}}(y)\|^2+\frac{1}{2}\|\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*)\|^2\\ &+(y-\Pi_{\bar{\Omega}}(y))^T(\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*))\\ &-\|\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*)\|^2+(\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*))^T(x-x^*)\\ &+(\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*))^T(\Pi_{\bar{\Omega}}(y)-y)\Big)+2\beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu\\ &=\beta e^{\beta t}\eta\Big(-\frac{1}{2}\|x-x^*\|^2-\frac{1}{2}\|\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*)\|^2\\ &\quad+(\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*))^T(x-x^*)\Big)+2\beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu\\ &=-\frac{\beta e^{\beta t}\eta}{2}\left(\|x-\Pi_{\bar{\Omega}}(y)\|^2\right)+2\beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu.\end{aligned}\tag{31}$$

Thus, we derive $\dot{E}(t,x,y)\le 2\beta e^{\beta t}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu$. By integrating both sides of the above inequality from 0 to t, one has $E(t,x,y)\le E(0,x_0,y_0)+\int_0^t 2\beta e^{\beta s}\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu ds$. Consequently, combining with $\mu_{i,k}\ge 0$, $\dot{\mu}_{i,k}<0$ and $\int_0^{+\infty}e^{\beta t}\kappa_{\hat{\Gamma}}\mu_{i,k}dt<+\infty, i=1,\dots,m, k=1,\dots,n$, we deduce that the positive part $[\dot{E}(t,x,y)]^+$ of $\dot{E}(t,x,y)$ belongs to $L^1\in(0,+\infty)$. Since $E(t,x,y)$ is bounded from below, it implies that $E(t,x,y)$ has a limit as $t\to+\infty$, and hence it is bounded, and the bound is defined as $\mathcal{B}<+\infty$. According to Lemma 2, one has

$$\frac{\eta}{2}\left(\|y-\Pi_{\bar{\Omega}}(y^*)\|^2-\|y-\Pi_{\bar{\Omega}}(y)\|^2\right)$$
$$\ge\frac{\eta}{2}\|\Pi_{\bar{\Omega}}(y)-\Pi_{\bar{\Omega}}(y^*)\|^2=\frac{\eta}{2}\left\|\frac{\dot{x}}{\beta}+x-x^*\right\|^2.$$

It implies $(\eta/2)\|(\dot{x}/\beta)+x-x^*\|^2\le E(t,x,y)<+\infty$. From the *Feasibility* in Theorem 4, we obtain that $x(t)\in\bar{\Omega}\ \ \forall t\in[0,+\infty)$. Since $\bar{\Omega}$ is closed a convex set that x is bounded can be obtained. Using again $(\eta/2)\|(\dot{x}/\beta)+x-x^*\|^2\le E(t,x,y)<+\infty$, we have $\dot{x}$ is bounded.

2) From 1), we have $E(t,x,y)\le\mathcal{B}<+\infty$. Note that $E(t,x,y)\ge e^{\beta t}(\hat{F}(x,\mu)+\Upsilon\hat{H}(x,\mu)-F(x^*)-\Upsilon H(x^*))$, therefore, we have $\hat{F}(x,\mu)+\Upsilon\hat{H}(x,\mu)-F(x^*)-\Upsilon H(x^*)\le(\mathcal{B}/e^{\beta t})\ \ \forall x\in\bar{\Omega}, t>0$. In addition, since $(\eta/2)\|x-x^*\|^2\le F(x)+\Upsilon H(x)-F(x^*)-\Upsilon H(x^*)\le\hat{F}(x,\mu)+\Upsilon\hat{H}(x,\mu)-F(x^*)-\Upsilon H(x^*)+\kappa_{\hat{\Gamma}}\mathbf{1}^T\mu\le(\mathcal{B}/e^{\beta t}), x\in\bar{\Omega}\ \ \forall t\ge 0$, such that $\|x-x^*\|\le(2\sqrt{\mathcal{B}}/\eta e^{\beta/2t}), x\in\bar{\Omega}\ \ \forall t\ge 0$ holds. ∎

Remark 2: 1) The Lyapunov function $E(t,x,y)$ in Theorem 5, which is designed on set $\bar{\Omega}$, is valid from the conclusion in the *Feasibility* premise in Theorem 4 and 2) the conditions $\mu_{i,k}\ge 0, \dot{\mu}_{i,k}<0$, $\int_0^{+\infty}e^{\beta t}\kappa_{\hat{\Gamma}}\mu_{i,k}dt<+\infty$, $i=1,\dots,m, k=1,\dots,n$ are satisfied if we take $\mu_{i,k}=e^{-\beta s}s^{-\gamma}$, $\gamma>2, i=1,\dots,m, k=1,\dots,n$.

Remark 3: Note that the same lemma (i.e., Lemma 4) and smoothing approximation function $\hat{g}(s,\mu)$ are used in this article and in [45], but they have the following different purposes.

1) Smoothing function $\hat{g}(s,\mu)$ is used to approximate different functions. In this article, the smoothing approximation function $\hat{g}(s,\mu)$ of the absolute value function $|s|$ is used to approximate nonsmooth convex functions (1-norm) to get $\sum_{i=1}^m\hat{f}_i(x_i,\mu_i)$. In addition, we used

$$\hat{h}(x_i-x_j,\mu_i)=\sum_{k=1}^n\hat{h}_k(x_i-x_j,\mu_i)$$
$$=\begin{cases}|x_{i,k}-x_{j,k}|, & \text{if } |x_{i,k}-x_{j,k}|>\mu_i\\ \frac{(x_{i,k}-x_{j,k})^2}{2\mu_{i,k}}+\frac{\mu_{i,k}}{2}, & \text{if } |x_{i,k}-x_{j,k}|\le\mu_i\end{cases}$$
$$i=1,\dots,m,\qquad k=1,\dots,n$$

to approximate the nonsmooth, convex penalty functions $(\Upsilon/2)\sum_{i=1}^m\sum_{j\in\mathcal{N}_i}|x_i-x_j|$ based on smoothing approximation function $\hat{g}(s,\mu)$. In [45], the smoothing approximation function $\hat{g}(s,\mu)$ of the absolute value function $|s|$ is used to approximate nonsmooth, nonconvex and non-Lipschitz function $\|x\|_p^p=\sum_{i=1}^n|x|^p$ to get $\hat{f}(x,\mu)=\sum_{i=1}^n g_i(x_i,\mu)^p$.

2) Smoothing function $\hat{g}(s,\mu)$ with different smoothing approximation parameter μ is used to obtain different convergence rates. In this article, in order to obtain convergence rate $\mathcal{O}(1/t^2)$ of DSPNA-N (19), the smoothing approximation parameter μ needs to satisfy the following conditions in Remark 1: the conditions $\mu_{i,k}\ge 0$, $\dot{\mu}_{i,k}<0$, $\int_0^{+\infty}t\kappa_{\hat{\Gamma}}\mu_{i,k}dt<+\infty, i=1,\dots,m, k=1,\dots,n$ imply $\mu_{i,k}=t^{-\gamma},\gamma>2$. To get convergence rate $\mathcal{O}(e^{-\beta t})$ of DSPNA-P, the smoothing approximation parameters μ should satisfy the following conditions 2) in Remark 2. The conditions $\mu_{i,k}\ge 0, \dot{\mu}_{i,k}<0$, $\int_0^{+\infty}e^{\beta t}\kappa_{\hat{\Gamma}}\mu_{i,k}dt<+\infty$, $i=1,\dots,m, k=1,\dots,n$ are satisfied if we take $\mu_{i,k}=e^{-\beta t}t^{-\theta}$, $\theta>1, i=1,\dots,m, k=1,\dots,n$.

Zhao *et al.* [45] only obtained the convergence property (the convergence rate is not included) of SINA for solving nonconvex, nonsmooth, and non-Lipschitz problem [i.e., L_p-minimization problem, $p\in(0,1)$], and the approximation parameter μ needs to meet $\mu_i\ge 0, \dot{\mu}_i(t)<0, i=1,\dots,n$.

V. Numerical Simulations

In the following, the proposed DSPNA-N (19) and DSPNA-P (25) are used to address distributed optimization problems with convex constrained sets to demonstrate their feasibility and effectiveness.

Example 1: Consider a nonsmooth generally convex optimization problem with a box constrained set over four agents as follows:

$$\min\ \bar{F}(x)=\sum_{i=1}^4|x-a_i|$$

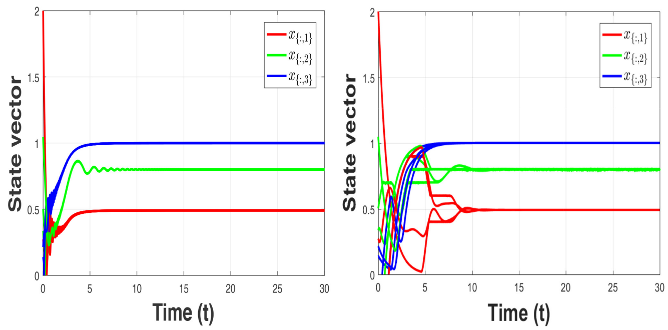


Fig. 2. Left: transient behaviors x of DSPNA-N (19); Right: transient behaviors x of PPDD [30].

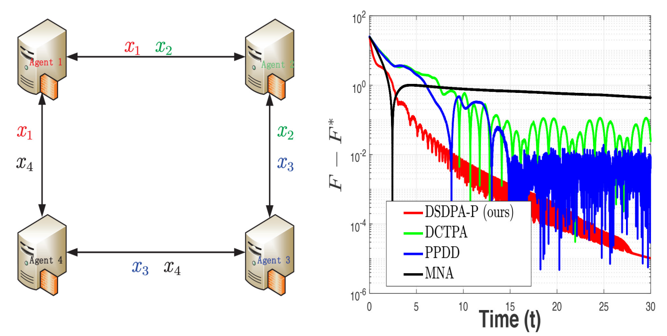


Fig. 3. Left: communication topology of four agents; Right: convergence rates of $F(x)-F^*$ of DSPNA-N (19), MNA [21], DCTPA [15], and PPDD [30].

$$\text{s.t.}\ \ x \in \bigcap_{i=1}^{4} \Omega_i \tag{32}$$

where $\Omega_i = \{x \in R^3 | d_{i,j} \le x_j \le c_{i,j}, j = 1, 2, 3.\}$, $a_1 = (0.4, 0.7, 1)^T$, $a_2 = (0.1, 0.9, 2)^T$, $a_3 = (0.9, 0.8, 1)^T$, $a_4 = (0.6, 0.8, 1)^T$, $d_i = (0, 0, 0)^T$, $c_i = (1, 1, 1)^T$, $i = 1, \ldots, 4$.

Fig. 2 displays the state trajectories x of DSPNA-N (19) and projected primal–dual dynamics (PPDDs) [30]. As can be seen from the results, both the DSPNA-N (19) with $\alpha = 4$, $\Upsilon = 4$, $\mu_{i,k} = (5/t^3), i = 1, \ldots, 3, k = 1, \ldots, 4$ and PPDD [30] converge to the same optimal solutions $x^* = (0.5, 0.8, 1)^T$. The left figure in Fig. 3 displays the communication topology of four agents. To further demonstrate the superiority of DSPNA-N (19), we compare it with the distributed continuous-time projection algorithm (DCTPA) [15], MNA [21], and PPDD [30], which are based on the differential inclusion method. As can be seen from Fig. 3(right), the proposed DSPNA-N (19) has the fastest convergence rate because it is designed based on Nesterov's accelerated methods (see Remark 4). Moreover, the PPDD outperforms the MNA and DCTPA since an augmented Lagrangian term is added in PPDD. It is meaningful to note that PPDD and DCTPA exhibit strong oscillation in the later running period because the subgradient in PPDD and DCTPA is a signal value used in simulation experiment rather than a set. The DSPNA-N can effectively avoid oscillation because the gradient of the approximation function (single-valued function) is used in DSPNA-N. Moreover, MNA does not produce oscillation because there exists a damping factor of the subgradient in MNA, which can suppress the oscillation in the experimental simulation.

Example 2: In this experiment, we consider a sparse signal reconstruction problem

$$\min_{x\in R^n} |x| \ \ \text{s.t.}\ \ Ax = b \tag{33}$$

where $A \in R^{m\times n} (m \ll n)$ is a sensing matrix, and $b \in R^m$ is an observed value. By splitting the matrix A and observed value b into K parts sequentially, problem (33) is equivalent to

$$\min_{x\in R^n} \frac{1}{K}\sum_{k=1}^{K} |x|, \ \ \text{s.t.}\ \ x \in \bigcap_{k=1}^{K} \Omega_k$$

where $\Omega_k = \{x \in R^n | A_{m_k\times n} x = b_{m_k} \in R^{\|m_k\|_0}\}$, $A = \begin{bmatrix} A_{m_1\times n} \\ A_{m_2\times n} \\ \vdots \\ A_{m_K\times n}\end{bmatrix}$, $b = \begin{bmatrix} b_{m_1\times n} \\ A_{m_2\times n} \\ \vdots \\ A_{m_K\times n}\end{bmatrix}$, $\sum_{k=1}^{K} \|m_k\|_0 = m$.
Note that problem (33) and the transformed problem mentioned above are not typical distributed optimization problems. Nevertheless, under the condition of Assumption 3, it is equivalent to solving the following distributed optimization problem:

$$\begin{aligned} &\min_{X\in R^{Kn}} \ \frac{1}{K}\sum_{k=1}^{K} |X_k| \\ &\quad \text{s.t.}\ \ A_{m_k\times n} X_k = b_{m_k} \in R^{\|m_k\|_0}, k = 1, \ldots, K \\ &\qquad\quad x_i = x_j \in R^n, i \in \mathcal{V}, j \in \mathcal{N}_i \end{aligned} \tag{34}$$

where $X = (X_1^T, \ldots, X_K^T)^T \in R^{Kn}$. DSPNA-N (19) is applied to solve problem (34) with $n = 32$, $m = 16$ and the sparsity of the original signal is 4 and $K = 5$. Fig. 4 (Top left) displays the communication topology of agents. From the top middle and top right graphs in Fig. 4, we can see that DSPNA-N (19) is globally asymptotically stable ($X_1^* = \cdots = X_5^* = x^*$) and DSPNA-N (19) has an arithmetic convergence rate, i.e., $\mathcal{O}(1/t^2)$ in which matches the conclusion in Theorem 2. Furthermore, the graphs in the bottom left in Fig. 4 illustrate that the sparse signals can be effectively reconstructed by the stable solutions of DSPNA-N (19) in a distributed way. Compared with the recovered sparse signals by PPDD [30] (bottom middle) and MNA [21] (bottom left), our DSPNA-N (19) has a better performance behavior, i.e., the solution more accurately approximates original sparse signals, because the smoothing approximation methods are used in the DSPNA-N (19).

Next, we will illustrate the effectiveness and convergence rate of DSPNA-P (25) in two examples.

Example 3: Consider a numerical example with a communication network consisting of five agents connected by a ring diagram as follows:

$$\begin{aligned} &\min\ \ \bar{F}(x) = \sum_{i=1}^{4} f_i(x) = \sum_{i=1}^{4} \frac{1}{2} x^T Q_i x + p_i^T x + |x| \\ &\text{s.t.}\ \ x \in \bigcap_{i=1}^{4} \Omega_i \end{aligned} \tag{35}$$

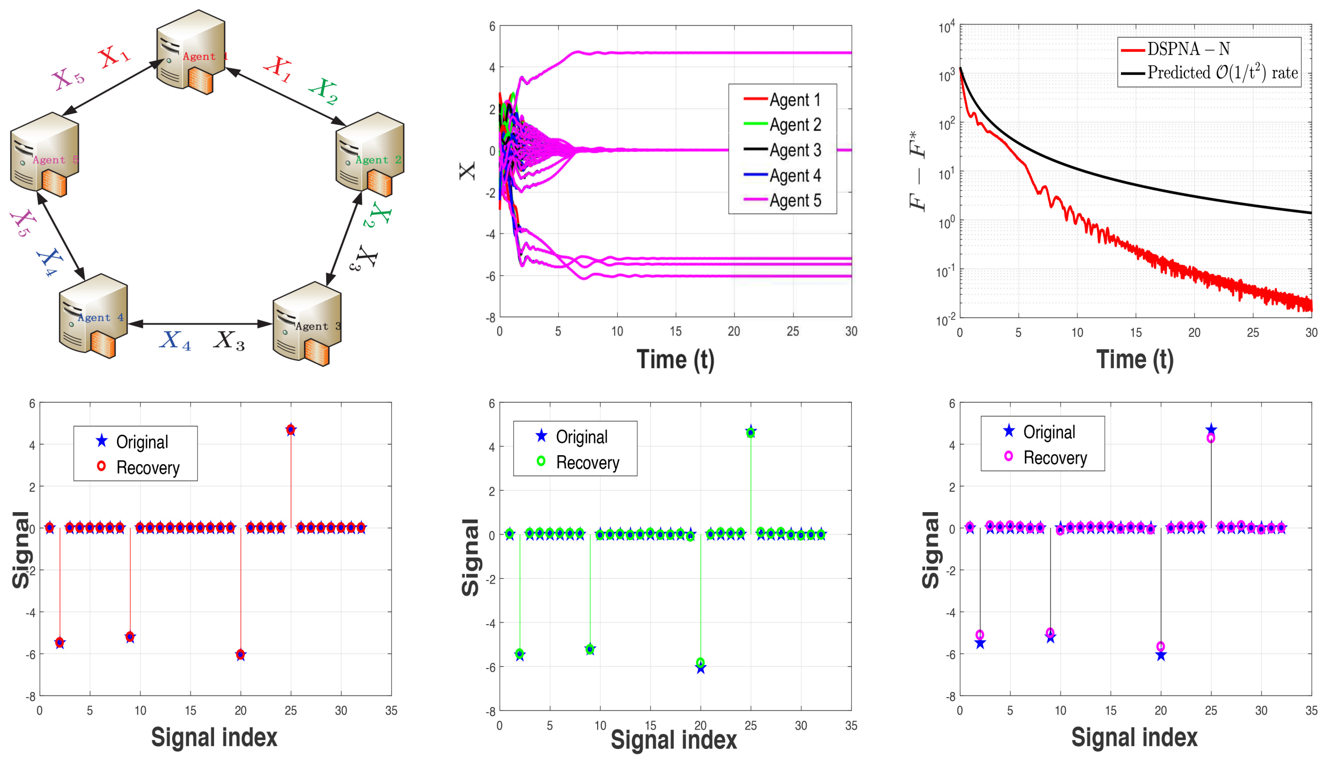


Fig. 4. Top left: communication topology of the five agents; Top middle: transient behaviors x of DSPNA-N (19); Top right: convergence rates of DSPNA-N (19); Bottom left: reconstructed signals of DSPNA-N (19); Bottom middle: reconstructed signals of PPDD [30]; Bottom right: reconstructed signals of MNA [21].

where $Q_i \in R^{4\times4}$ is a positive definite matrix and $p_i \in R^4$. The box constraints are considered, i.e., $\Omega_i = \{x \in R^4 | x_{j,\mathrm{m}} \leq x_j \leq x_{j,M}, j = 1,..,4\}$. The matrix $Q_i \in R^{4\times4}$ is positive definite and randomly generated and the elements of p_i, $-x_{j,\mathrm{m}}$, $x_{j,M}$ are randomly generated on the interval $[0, 1]$. Note that the objective function of problem (35) is nonsmooth ($|x|$) and strongly convex ($x^T Q_i x + p_i^T x$ is with a positive definite matrix $Q_i \in R^{4\times4}$).

Fig. 5 (top, bottom left) displays that the trajectories x of PPDD [30], CNS [29], and DSPNA-P (25) are globally asymptotically stable and converge to the same optimal solution in a distributed way. Fig. 5 (bottom right) displays the convergence rates of PPDD, CNS, and DSPNA-P (25). As can be seen from Fig. 5 (bottom right), DSPNA-P (25) has a predicted exponential convergence rate, i.e., $\mathcal{O}(e^{-\beta t})$ when $\beta = \sqrt{\eta}$, which is consistent with the conclusion in Theorem 5, and PPDD and CNS solve nonsmooth strongly convex optimization problems, which seemingly do not have exponential convergence rates. It is easy to see that DSPNA-P has a faster convergence rate than PPDD, CNS after a certain time.

Example 4: In this example, we apply DSPNA-P (25) to solve the sensor network localization problem. Consider a scenario in which there are five sensors and ten anchors in the plane R^2. The locations of ten anchors are marked as $b_\iota \in R^2(\iota \in \{1, 2, \ldots, 10\})$ and the positions of five sensors are tagged as $x_i = (x_i^1, x_i^2) \in R^2(i \in \{1, 2, \ldots, 5\})$. The links of all sensors and anchors are shown in Fig. 6 (top middle and bottom middle). In Fig. 6 (top middle), the anchors' locations $b_\iota(\iota \in \{1, 2, \ldots, 10\})$ are labeled by

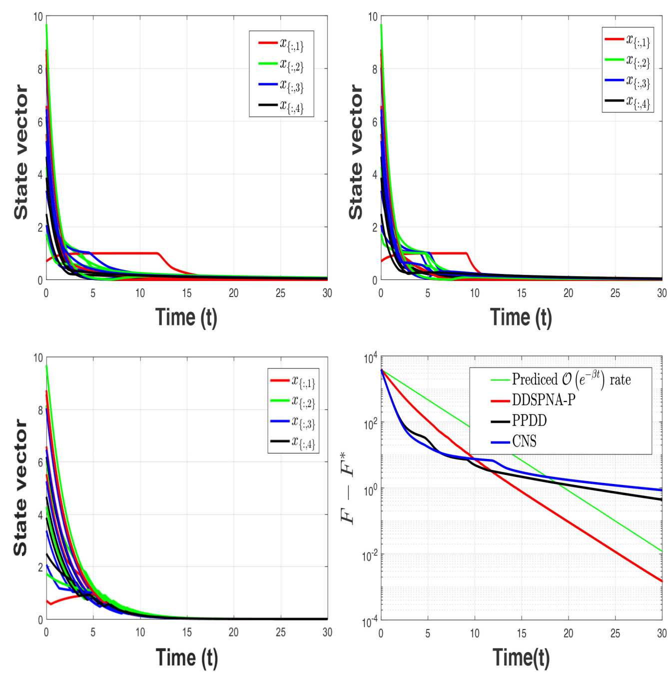


Fig. 5. Top left: transient behaviors x of PPDD [30]; Top right: transient behaviors x of CNS [29]; Bottom left: transient behaviors x of DSPNA-P (25); Bottom right: convergence rates of $F(x) - F(x^*)$ with $\beta = \sqrt{\eta}$.

yellow solid squares, while the initial locations of sensors are labeled by black solid pentagrams $x_i(\ i \in \{1, 2, \ldots, 5\})$ generated randomly. In Fig. 6 (bottom middle), the optimal sensors' locations $x_i^*(\ i \in \{1, 2, \ldots, 5\})$ are labeled by red

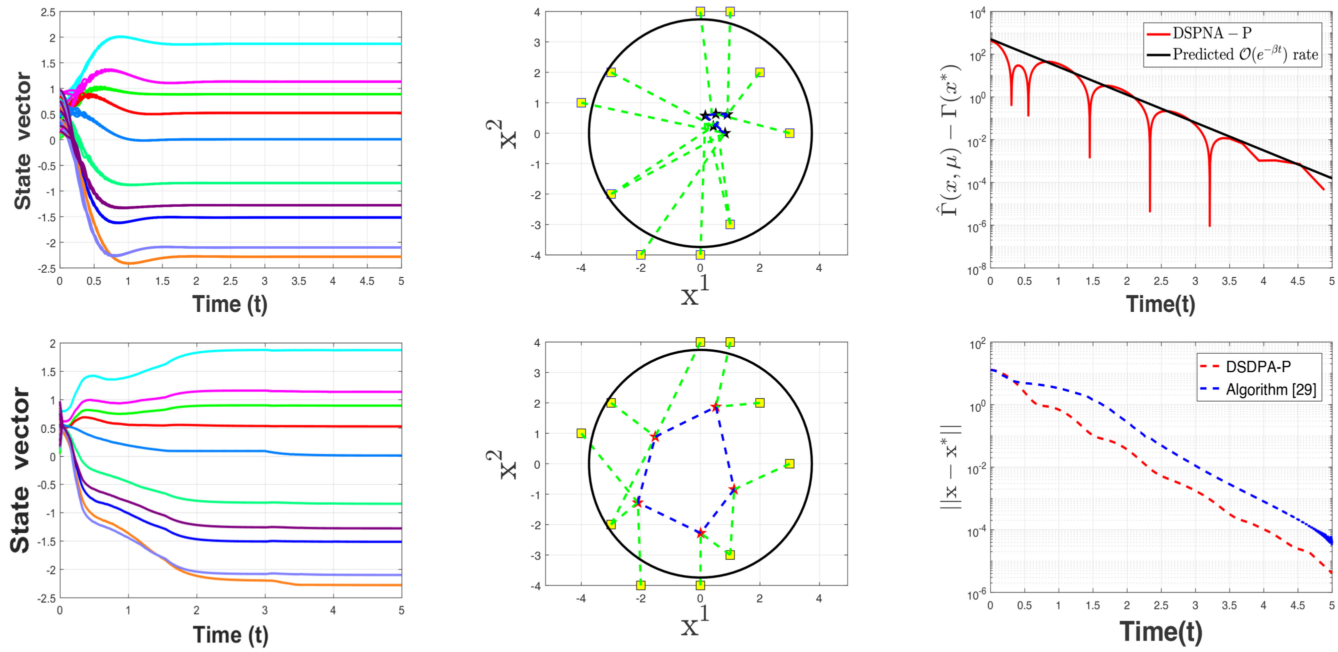


Fig. 6. Top left: consensus of the output vectors x of DSPNA-P (25); Top middle: topology and initial location of sensors and anchors; Top right: convergence rates of $\hat{\Gamma}(x,\mu)-\Gamma(x^*)$; Bottom left: consensus of the output vectors x of the DCTPA [24]; Bottom middle: topology and optimal location of sensors and anchors; Bottom right: convergence rate of $\|x-x^*\|$.

solid pentagrams, and it is not difficult to see that the constraint ($\|x_i\|^2 \le 14,\ i \in \{1,2,\ldots,5\}$) yields the optimal values $x_i^*(i=1,2,3,4,5)$ in the feasible region. Moreover, the black-dotted lines refer to links between sensors, while the green-dotted lines stand for the links between sensors and anchors and the black circle refers to the feasible region. For each sensor, there exists a bounded constraint to restrict the location of the sensors. The associated target is to minimize the connection length of the sensors and the anchors. Thus, the optimization problem is given by

$$\min \sum_{i=1}^{5} \frac{1}{2}\left(\sum_{j\in\mathcal{N}_i}\|x_i-x_j\|^2+\sum_{\iota\in\mathcal{N}_i}\|x_i-b_\iota\|^2\right)$$
$$\text{s.t.}\ \ \|x_i\|^2 \le 14,\ \ i\in\{1,2,\ldots,5\}.$$

Applying DSPNA-P (25) to deal with this problem, where its communication topology is similar to Fig. 4 (top left). For agent i, the objective function is

$$f_i(x)=\frac{1}{2}\left(\sum_{j\in\mathcal{N}_i}\|x_i-x_j\|^2+\sum_{\iota\in\mathcal{N}_i}\|x_i-b_\iota\|^2\right)$$

where $x=(x_1^T,x_2^T,x_3^T,x_4^T,x_5^T)^T$. From Fig. 6 (top left and bottom left), the output values of sensors converge to consensus optimal solutions of DSPNA-P (25) and the algorithm in [24]. In addition, Fig. 6 (top right) displays the convergence rate of $\hat{\Gamma}(x,\mu)-\Gamma(x^*)$ in DSPNA-P (25). It is easy to see that the a DSPNA-P (25) has an exponential convergence rate that is faster than the predicted exponential convergence rate that matches the conclusion in Theorem 5. Compared with the algorithm in [24], the DSPNA-P (25) has faster an exponential convergence rate since it is designed based on heavy-ball dynamical method, which has a faster exponential convergence rate for solving strongly convex optimization problems.

Remark 4: The continuous variant of Nesterov's acceleration for centralized unconstrained generally convex optimization problem $\min f(x)$ has a faster convergence rate $\mathcal{O}(1/t^2)$ than the classical gradient dynamic method $\dot{x}=-\nabla f(x)$ in [31], which is $\ddot{x}+(3/t)\dot{x}+\nabla f(x)=0$. It can be rewritten equivalently as: $\begin{cases}\dot{x}=\frac{2}{t}(y-x)\\ \dot{y}=-\frac{t}{2}\nabla f(x)\end{cases}$. Note that the NSPNA-N (19) can be obtained by replacing y with $P_\Omega(y)$ in $\dot{x}=(2/t)(y-x)$ and by taking the use of $\nabla_x\hat{f}_i(x_i,\mu_i)+\Upsilon\sum_{j\in\mathcal{N}_i}\nabla_x\hat{h}(x_i-x_j,\mu_i)$ instead of $\nabla f(x)$ in equation $\dot{y}=-(t/2)\nabla f(x)$.

The continuous variant of Polyak's heavy ball method $\ddot{x}+\lambda\dot{x}+\nabla f(x)=0, \lambda=2\sqrt{\eta}$ for centralized unconstrained strongly convex optimization problem $\min f(x)$ has a faster exponential convergence than the classical gradient dynamic method $\dot{x}=-\nabla f(x)$ in [35]. $\ddot{x}+\lambda\dot{x}+\nabla f(x)=0, \lambda=2\sqrt{\eta}$ can be rewritten equivalently as: $\begin{cases}\dot{x}=\sqrt{\eta}(y-x)\\ \dot{y}=\sqrt{\eta}(x-y-\frac{1}{\eta}\nabla f(x))\end{cases}$. Note that the NSPNA-P (25) can be obtained by replacing y and $\sqrt{\eta}$ with $P_\Omega(y)$ and β in $\dot{x}=\sqrt{\eta}(y-x)$ and by replacing $\nabla f(x)$ and $\sqrt{\eta}$ in equation $\dot{y}=\sqrt{\eta}(x-y-(1/\eta)\nabla f(x))$ with $\nabla_x\hat{f}_i(x_i,\mu_i)+\Upsilon\sum_{j\in\mathcal{N}_i}\nabla_x\hat{h}(x_i-x_j,\mu_i)$ and β. However, DCTPA [15], MNA [21], the algorithm in [24], CNS [29], and PPDD [30] are designed based on $\dot{x}=-\nabla f(x)$ and $\dot{x}=-x+\Pi_\Omega(x-\nabla f(x))$ (both of them have no acceleration characteristics). It is easy to understand that our NSPNAs converge faster than DCTPA [15], MNA [21], the algorithm in [24], CNS [29], and PPDD [30] for solving related optimization problems.

VI. Conclusion

In this article, two fast convergent DSPNAs (DSPNA-N and DSPNA-P) based on Nesterov's and polyak's accelerated method have been proposed to solve distributed, nonsmooth, constrained optimization problems (DNSCOP) in two cases (more specifically, solving DNSCOP with generally convex objective functions has an $\mathcal{O}(1/t^2)$ convergence rate by DSPNA-N and solving DNSCOP with strongly convex objective functions has faster exponential convergence rates). An exact penalty method has been used to transform the original optimization problem into an optimization problem with local set constraints and without consensus constraints. The smoothing approximation method has been introduced in our DSPNAs, which not only guarantee the existence and uniqueness of solutions of DSPNAs but also avoids the difficulties of selecting a suitable subgradient and computational difficulties of proximal operators. A novel projection approach, i.e., by introducing a projection operator for the auxiliary variable to restrict the optimal variable to satisfy the local constraint sets, is proposed and used to guarantee the feasibility of the solution of DSPNAs. A numerical example and practical application example (sparse signal reconstruction) have been done to illustrate the effectiveness and superiority of DSPNA-N. Subsequently, the proposed DSPNA-P is applied to deal with a numerical example (nonsmooth, constrained, quadratic optimization problem) and practical application example (sensor network localization problem). The experimental results and comparative studies show that the DSPNAs proposed in this article can solve DNSCOP efficiently, and the proposed DSPNAs have better performance in terms of convergence rates compared to other typical approaches. Future work may focus on designing accelerated primal–dual mirror dynamical approaches to solve centralized and distributed constrained smooth and nonsmooth convex optimization problems. In addition, considering that many engineering optimization problems in reality are nonconvex, nonsmooth, or even non-Lipschitz, we intend to investigate distributed accelerated primal–dual neurodynamic approaches for solving nonconvex, nonsmooth, or even non-Lipschitz optimization problems and their related theories in the future.

References

[1] M. Rabba and R. Nowak, "Distributed optimization in sensor networks," in *Proc. Int. Symp. Inf. Process. Sens. Netw. (Part CPS Week)*, 2004, pp. 20–27.

[2] M. Li, D. G. Andersen, and D. G. Park, "Scaling distributed machine learning with the parameter server," in *Proc. USENIX Symp. Oper. Syst. Design Implement.*, 2014, pp. 583–598.

[3] X. He, D. W. C. Ho, T. Huang, J. Yu, H. Abu-Rub, and C. Li, "Second-order continuous-time algorithms for economic power dispatch in smart grids," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 9, pp. 1482–1492, Sep. 2018.

[4] 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.

[5] Z. Deng, "Distributed algorithm design for resource allocation problems of second-order multiagent systems over weight-balanced digraphs," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3512–3521, Jun. 2021.

[6] K. Li, Q. Liu, S. Yang, J. Cao, and G. Lu, "Cooperative optimization of dual multiagent system for optimal resource allocation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4676–4687, Nov. 2020.

[7] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control.*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[8] J. Duchi, A. Agarwal, and M. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Autom. Control.*, vol. 57, no. 3, pp. 592–606, Mar. 2012.

[9] T. Chang, A. Nedić, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Trans. Autom. Control.*, vol. 59, no. 6, pp. 1524–1538, Jun. 2014.

[10] H. Li, Q. Lü, X. Liao, and T. Huang, "Accelerated convergence algorithm for distributed constrained optimization under time-varying general directed graphs," *IEEE Trans. Syst., Man, Cybern., Syst.* vol. 50, no. 7, pp. 2612–2622, Jul. 2020.

[11] F. Guo, G. Li, C. Wen, L. Wang, and Z. Meng, "An accelerated distributed gradient-based algorithm for constrained optimization with application to economic dispatch in a large-scale power system," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 4, pp. 2041–2053, Apr. 2021.

[12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[13] J. Wang and N. Elia, "Control approach to distributed optimization," in *Proc. 48th Annu. Allerton Conf. Commun. Control Comput.*, Allerton, IL, USA, 2010, pp. 557–561.

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

[15] X. Zeng, P. Yi, and Y. Hong, "Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5227–5233, Oct. 2017.

[16] 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.

[17] X. He, T. Huang, J. Yu, C. Li, and Y. Zhang, "A continuous-time algorithm for distributed optimization based on multiagent networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 12, pp. 2700–2709, Dec. 2019.

[18] 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.

[19] S. Liang, X. Zeng, and Y. Hong, "Distributed nonsmooth optimization with coupled inequality constraints via modified Lagrangian function," *IEEE Trans. Autom. Control*, vol. 63, no. 6, pp. 1753–1759, Jun. 2018.

[20] 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.

[21] L. Ma and W. Bian, "A novel multiagent neurodynamic approach to constrained distributed convex optimization," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1322–1333, Mar. 2021.

[22] S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, May 2015.

[23] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlinear Sci.*, vol. 29, pp. 1247–1272, Aug. 2019.

[24] W. Li, X. Zeng, S. Liang, and Y. Hong, "Exponentially convergent algorithm design for constrained distributed optimization via non-smooth approach," *IEEE Trans. Autom. Control.*, vol. 67, no. 2, pp. 934–940, Feb. 2022, doi: 10.1109/TAC.2021.3075666.

[25] P. Yi, Y. Hong, and F. Liu, "Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems," *Automatica*, vol. 74, pp. 259–269, Dec. 2016.

[26] Y. Zhu, W. Ren, W. Yu, and G. Wen, "Distributed resource allocation over directed graphs via continuous-time algorithms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 2, pp. 1097–1106, Feb. 2021.

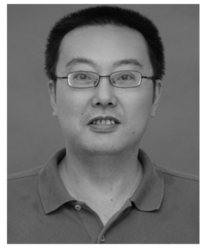
[27] H. Wang, X. Liao, T. Huang, and C. Li, "Cooperative distributed optimization in multiagent networks with delays," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 2, pp. 363–369, Feb. 2015.

[28] S. Yang, Q. Liu, and J. Wang, "Distributed optimization based on a multiagent system in the presence of communication delays," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 5, pp. 717–728, May 2017.

[29] Y. Zhao and Q. Liu, "A consensus algorithm based on collective neurodynamic system for distributed optimization with linear and bound constraints," *Neural. Netw.*, vol. 122, pp. 144–151, Feb. 2020.
[30] Y. Zhu, W. Yu, G. Wen, and G. Chen, "Projected primal-dual dynamics for distributed constrained nonsmooth convex optimization," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1776–1782, Apr. 2020.
[31] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Assoc., 2014, pp. 2510–2518.
[32] X. Zeng, J. Lei, and J. Chen, "Dynamical primal-dual accelerated method with applications to network optimization," *IEEE Trans. Autom. Control*, early access, Feb. 22, 2022, doi: 10.1109/TAC.2022.3152720.
[33] X. Jiang, S. Qin, X. Xue, and X. Liu, "A second-order accelerated neurodynamic approach for distributed convex optimization," *Neural. Netw.*, vol. 146, pp. 161–173, Feb. 2022.
[34] 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 Nonlinear Control*, vol. 30, no. 17, pp. 7574–7592, Nov. 2020.
[35] J. W. Siegel, "Accelerated first-order methods: Differential equations and Lyapunov functions," 2019, *arXiv:1903.05671*.
[36] Y. Wei, H. Fang, X. Zeng, J. Chen, and P. Pardalos, "A smooth double proximal primal-dual algorithm for a class of distributed nonsmooth optimization problems," *IEEE Trans. Autom. Control.*, vol. 65, no. 4, pp. 1800–1806, Apr. 2020.
[37] R. Chai, A. Savvaris, A. Tsourdos, S. Chai, Y. Xia, and S. Wang, "Solving trajectory optimization problems in the presence of probabilistic constraints," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4332–4345, Oct. 2020.
[38] R. Chai, A. Savvaris, A. Tsourdos, S. Chai, and Y. Xia, "Improved gradient-based algorithm for solving aeroassisted vehicle trajectory optimization problems," *J. Guid. Control Dyn.*, vol. 40, no. 8, pp. 2093–2101, 2017.
[39] R. Chai, A. Savvaris, A. Tsourdos, Y. Xia, and S. Chai, "Solving multiobjective constrained trajectory optimization problem by an extended evolutionary algorithm," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1630–1643, Apr. 2020.
[40] P. Yi, Y. Hong, and F. Liu, "Distributed gradient algorithm for constrained optimization with application to load sharing in power systems," *Syst. Control Lett.*, vol. 83, pp. 45–52, Sep. 2015.
[41] Q. Liu and J. Wang, "A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 5, pp. 812–824, May 2013.
[42] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
[43] X. Chen, "Smoothing methods for nonsmooth, nonconvex minimization," *Math. Program.*, vol. 134, no. 1, pp. 71–99, 2012.
[44] W. Bian and X. Chen, "Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 545–556, Mar. 2014.
[45] Y. Zhao, X. Liao, X. He, R. Tang, and W. Deng, "Smoothing inertial neurodynamic approach for sparse signal reconstruction via L_p-norm minimization," *Neural Netw.*, vol. 140, pp. 100–112, Aug. 2021.

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

His current research interests include neurodynamic optimization, distributed optimization, compressed sensing, and smart grid.



Xiaofeng Liao (Fellow, IEEE) received the B.S. and M.S. degrees in mathematics from Sichuan University, Chengdu, China, in 1986 and 1992, respectively, and the Ph.D. degree in circuits and systems from the University of Electronic Science and Technology of China, Chengdu, in 1997.

From 1999 to 2012, he was a Professor with Chongqing University, Chongqing, China. From July 2012 to July 2018, he was a Professor and the Dean of the College of Electronic and Information Engineering, Southwest University, Chongqing. He is currently a Professor and the Dean of the College of Computer Science, Chongqing University. He is also a Yangtze River Scholar of the Ministry of Education of China, Beijing, China. From November 1997 to April 1998, he was a Research Associate with the Chinese University of Hong Kong, Hong Kong. From October 1999 to October 2000, he was a Research Associate with the City University of Hong Kong, Hong Kong. From March 2001 to June 2001 and March 2002 to June 2002, he was a Senior Research Associate with the City University of Hong Kong. From March 2006 to April 2007, he was a Research Fellow with the City University of Hong Kong. He holds five patents, and published four books and over 300 international journal and conference papers. His current research interests include optimization and control, machine learning, neural networks, bifurcation and chaos, and cryptography.

Prof. Liao currently serves as an Associate Editor of the IEEE TRANSACTIONS ON CYBERNETICS and IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Xing He 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 the 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 School of Electronics and Information Engineering, Southwest University, Chongqing. His research interests include neural networks, bifurcation theory, optimization method, smart grid, and nonlinear dynamical system.