Neurocomputing 483 (2022) 446–459

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A review of distributed optimization: Problems, models and algorithms ☆

Yanling Zheng [a,b], Qingshan Liu [a,b,*]

[a] School of Mathematics, Southeast University, Nanjing 210096, China
[b] Jiangsu Provincial Key Laboratory of Networked Collective Intelligence, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 31 March 2021
Revised 25 May 2021
Accepted 3 June 2021
Available online 2 November 2021

Keywords:
Distributed optimization
Distributed algorithms
Distributed models
Multi-agent systems

ABSTRACT

With the development of big data and artificial intelligence, distributed optimization has emerged as an indispensable tool for solving large-scale problems. In particular, the multi-agent system based on distributed information processing can be elaborately designed for distributed optimization, in which the agents collaboratively minimize a global objective function made up of a sum of local objective cost functions subject to some local and/or global constraints. Inspired by the applications involving resource allocation, machine learning, power systems, sensor networks and cloud computing, a variety of distributed optimization models and algorithms have been investigated and developed. The optimization models include unconstrained and constrained problems in continuous and discontinuous systems with undirected and directed communication topology graphs. The constraints include bounded constraint, separable and inseparable equality and inequality constraints. Meanwhile, in distributed algorithms, every agent executes its local computation and updating on basis of its own data information and that exchanging with its neighboring agents by means of the underlying communication networks, in order to deal with the optimization problems in a distributed way. This paper is designed to provide a comprehensive overview of extant distributed models and algorithms for distributed optimization.

1. Introduction

In recent years, mathematical optimization has developed as a crucial component in a wide range of technical fields, such as traffic engineering [1], mobile robots cooperative transportation [2], smart grids [3] and so on. Meanwhile, mathematical optimization has over the last decade emerged as a indispensable instrument in extracting information in communication which leads to the development of network systems. These network systems are composed of a great amount of interactive agents, where the agents require to cooperate with each other on behalf of achieving a desirable global target. The applications for these network dynamic systems contain power economic dispatch [4], distributed control [5] and robot systems [6]. Moreover, many applications considered in such network systems are able to be constructed as a framework of distributed optimization [7], which can be divided into unconstrained optimization and constrained optimization to fulfill some special demands such as security, privacy and kinetics. The constrained optimization problems comprise of a desirable global objective function subject to bound, equality and inequality constrains. Unlike the traditional centralized strategies such as competitive mechanism integrated whale optimization algorithms [8] and heuristic algorithms such as particle swarm optimization approaches [9,10], the distributed framework has better performance [11], such as stable robustness rather than needing malfunction detection, lower communication bandwidth requirement and computation load, and higher extendibility and elasticity, which already have made significant applications on distributed algorithms to deal with optimization problems [12–14].

Tracing back to the seminal works under circumstance of distributed and parallel computation [15,16], the distributed optimization and algorithms have been investigated widely. Various distributed optimization approaches have been proposed in the literature [17,18]. Most presently existing distributed algorithms are on the strength of consensus and able to be split into two classifications by virtue of whether the distributed method is in discrete-time systems or in continuous-time systems. In these approaches, every agent has a dynamic state as the estimate of the optimization decision variable, and updates its estimated value on account of its own local information interaction from its connected neighbors by means of the underlying communication network. Meanwhile, the communication networks are divided into undirected connected

☆ This work was supported in part by the National Natural Science Foundation of China under Grant 61876036, and in part by the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence under Grant BM2017002.

* Corresponding author at: School of Mathematics, Southeast University, Nanjing 210096, China.

E-mail addresses: zhengyl@seu.edu.cn (Y. Zheng), qsliu@seu.edu.cn (Q. Liu).

https://doi.org/10.1016/j.neucom.2021.06.097
0925-2312/

and directed connected topological graph to ensure communications among agents at different power channels which signifies owning communication capability in different direction and channels.

Recent reviews on distributed optimization [5,19] mainly focus on discrete time approaches with damping step sizes, and the review in [20] discusses consensus averaging-based algorithms for unconstrained distributed problems. However, this paper makes a comprehensive review on distributed optimization models and algorithms, which are not only discrete time approaches with both damping and fixed constant step sizes, but also continuous-time algorithms. The reviewed distributed optimization problems include decomposable objective functions, decomposable or undecomposable variables. In addition, we also review both unconstrained and constrained distributed optimization problems, where the constraints include bound constraints, separable and inseparable equality and inequality constraints, hybrid constraints comprised of local bound constraints, and global coupled equality and inequality constraints. The algorithms include the first order method based on gradient information and Newton approach based on Hessian information. Moreover, the survey in [21] is mainly concentrated in constrained optimization algorithms with first order augmented Lagrangian approach and primal-dual method.

The remainder of the paper is organized as follows. In Section 2, some preliminaries are presented on some useful basic inequalities, convex analysis, graph theory and the frameworks of distributed optimization. Section 3 reviews various algorithms for unconstrained distributed optimization in both undirected and directed communication graphs. Considering the convergence time, the infinite and finite/fixed time consensus of distributed optimization models and algorithms have also been discussed. Section 4 concentrates on the distributed optimization problems with different types of bound constraints. Section 5 and Section 6 review the various models and algorithms for distributed optimization problems with equality and inequality constraints, respectively. Finally, Section 7 reviews various models and algorithms for distributed optimization problems with hybrid constraints comprised of local bound constraints, global equality and inequality constraints. In addition, Tables 1 and 2 show the comparisons of some typical algorithms for unconstrained and constrained distributed optimization.

2. Preliminaries

In this section, some necessary notations and mathematical concepts are described briefly to assist the follow-up technical discussions in the review. First, some notations are provided. $\Re$ and $\Re_+$ are represented as real number set and nonnegative real number set respectively. Let symbols $0_N \in \Re^N$ and $\mathbf{1}_N \in \Re^N$ denote the column vectors where each element is 0 and 1 respectively. $I_N \in \Re^{N\times N}$ is the identity matrix. The superscript $(\cdot)^T$ and $\|\cdot\|$ represent the transpose and the Euclidean norm respectively.

2.1. Basic inequality

A function $f : \Re^s \to \Re$ is called convex if for any $\theta \in (0,1)$ and any $x_1, x_2 \in \Re^s$, there is

$$f(\theta x_1 + (1-\theta)x_2) \leqslant \theta f(x_1) + (1-\theta)f(x_2). \tag{1}$$

It means that a convex function is always staying below a line which concatenates any two points on the function curve. If $f : \Re^s \to \Re$ is continuously differentiable, an equivalent definition of convex function is for any $x_1, x_2 \in \Re^s$,

$$f(x_1) \geqslant f(x_2) + \nabla^T f(x_2)(x_1 - x_2), \tag{2}$$

where $\nabla f(x_2)$ represents the gradient of f at point x_2.

A function $f : \Re^s \to \Re$ is called μ- strongly convex if for any $x_1, x_2 \in \Re^s$, there is

$$f(x_1) \geqslant f(x_2) + \nabla^T f(x_2)(x_1 - x_2) + \frac{\mu}{2}\|x_1 - x_2\|^2, \tag{3}$$

where μ is a positive constant. Note that the convexity in (2) is weaker than (3) as result of a quadratic lower bound is imposed extra on f in (3).

A function $f : \Re^s \to \Re$ is called L_f- Lipschitz if for any $x_1, x_2 \in \Re^s$, there is

$$|f(x_1) - f(x_2)| \leqslant L_f\|x_1 - x_2\|, \tag{4}$$

where $L_f > 0$.

A function $f : \Re^s \to \Re$ is called L_g smoothness if for any $x_1, x_2 \in \Re^s$, there is

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leqslant L_g\|x_1 - x_2\|, \tag{5}$$

where $L_g > 0, \nabla f(x_1)$ and $\nabla f(x_2)$ are the gradients of f at point x_1 and x_2 respectively. An equivalent condition of smoothness is

$$f(x_1) \leqslant f(x_2) + \nabla^T f(x_2)(x_1 - x_2) + \frac{L_g}{2}\|x_1 - x_2\|^2. \tag{6}$$

2.2. Graph theory

Denote a directed graph $\mathscr{G} = (\mathscr{V}, \xi, \mathscr{A})$ with N vertexes, which can be described by a node set $\mathscr{V} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_N\}$, a directed edge set $\xi \subseteq \mathscr{V} \times \mathscr{V}$, and an adjacent matrix $\mathscr{A} = [a_{ij}]_{N\times N}$ with $a_{ji} > 0$ when the edge $(\vartheta_i, \vartheta_j) \in \xi$ and $a_{ji} = 0$ otherwise. Assume that the digraph does not have any multiedges and self loops. A directed graph is regarded as strongly connected if for any two nodes ϑ_i, ϑ_j there is a directed path from ϑ_i to ϑ_j, i.e., there are nodes $\vartheta_{i0} = \vartheta_i, \vartheta_{i1}, \ldots, \vartheta_{ir} = \vartheta_j$ such that $(\vartheta_{i,k-1}, \vartheta_{i,k}) \in \xi$ for $1 \leqslant k \leqslant r$. Define i's out-neighbors $\mathscr{N}_i^{out} = \{\vartheta_j | (\vartheta_i, \vartheta_j) \in \xi\}$ and in-neighbors

Table 1
Comparisons of some typical algorithms for unconstrained distributed optimization.

Method	Communication graph	Conditions	References
DGD method	undirected, connected, doubly stochastic	convex, bounded subgradient	[22,23]
distributed Nesterov gradient method	connected, undirected, symmetric, doubly stochastic	convex, bounded subgradient, Lipschitz continuous gradient	[24]
dual subgradient-push averaging method	directed, strongly connected, row-stochastic	strongly-convex, differentiable, Lipschitz-continuous gradient	[25]
projected subgradient method	undirected, connected, doubly stochastic	convex, subgradient uniformly bounded	[26,27]
primal-dual augmented Lagrangian method	undirected, connected, doubly stochastic	strongly-convex, differentiable, Lipschitz continuous gradient	[28–30]
subgradient-push method	directed, B-strongly connected, column-stochastic	convex, bounded subgradient	[31]

Table 2
Comparisons of some typical algorithms for constrained distributed optimization.

Method	Communication graph	Conditions	References
distributed PI	undirected connected, doubly stochastic	convex, bounded gradient	[32,33]
distributed projected subgradient method	connected, doubly stochastic	convex, nonsmooth	[17,18,34]
distributed subgradient-push-sum method	uniformly jointly connected	convex, nonsmooth	[35]
distributed stochastic gradient method	undirected connected, random graph	convex, subgradient uniformly bounded	[36]
distributed primal-dual subgradient method	directed, strongly connected, balanced	convex, nonsmooth	[37,38]
distributed primal-dual subgradient method	undirected connected	convex, bounded subgradient	[39]

$\mathcal{N}_i^{in} = \{\vartheta_j | (\vartheta_j, \vartheta_i) \in \xi\}$ respectively. The degree of node ϑ_i is defined as the out-degree of node ϑ_i, where the degree $d_i = |\mathcal{N}_i^{out}|$. Moreover, a graph is undirected if and only if $(\vartheta_i, \vartheta_j) \in \xi$ implies $(\vartheta_j, \vartheta_i) \in \xi$ for all pairs of nodes. The Laplacian matrix $L \in \Re^{N\times N}$ of a graph $\mathcal{G}$ is defined as $L = \mathcal{D} - \mathcal{A}$, where $\mathcal{D} = \text{diag}\{d_1, d_2, \ldots, d_N\}$. For undirected graph $\mathcal{G}$, there are $L = L^T$ and $L\mathbf{1}_N = L^T\mathbf{1}_N = 0$. Furthermore, if $\mathcal{G}$ is connected, the Laplacian matrix L has a simple zero eigenvalue and all the other eigenvalues are positive.

2.3. Matrix theory

A matrix is called as column stochastic (row stochastic) if it is a nonnegative matrix and every column (row) sum is one. And it is called doubly stochastic if it is both row and column stochastic. For a row stochastic matrix M, π_M is its positive left eigenvector corresponding to the eigenvalue 1 subject to $\pi_M^T\mathbf{1}_n = 1$. Likewise, for a column stochastic matrix N, π_N is its positive right eigenvector corresponding to the eigenvalue 1 subject to $\mathbf{1}_n^T\pi_N = 1$.

2.4. Distributed optimization with decomposable objective function

The target of N nodes in a network system is to cooperatively minimize a convex objective function denoted as a sum of all agents' local convex cost functions. Based on local computation and local communication, all the agents cooperatively together with each other to deal with the following convex distributed optimization with decomposable objective function

$$\min_x \sum_{i=1}^{N} f_i(x), \tag{7}$$

where $x \in \Omega$ is the global optimization variable, local objective function $f_i : \Re^n \to \Re$ is satisfied by the agent $i (i = 1, 2, \ldots, N)$, and Ω is the feasible set defined by the constraints.

2.5. Distributed optimization with decomposable objective function and variables

Concerned with a set of N agents in a network system which interacting communication over a connected topology graph, each agent has its own local cost function $f_i(x_i)$, and all agents collaborate with each other to solve the following optimization problem (see Fig. 1)

$$\min_x \sum_{i=1}^{N} f_i(x_i), \tag{8}$$

where $x_i \in \Omega_i$ is the state of agent $i, f_i(x_i) : \Re^{n_i} \to \Re$ is the local cost function and Ω_i is the feasible set of agent i. Note that through reformulating problem (7), the problem (7) can be transformed into problem (8) under a consensus condition that all agents arrive at a consensus while minimizing the total cost function.

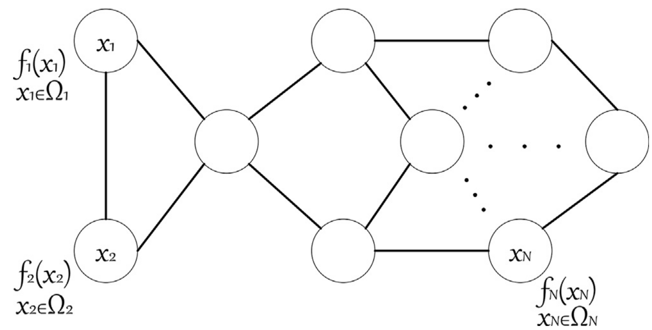

Fig. 1. The basic structure of multi-agent network for distributed optimization.

Remark 1. Generally, many applications considered in network systems are able to be constructed as a framework of distributed optimization with local private objective functions, in which the target of distributed optimization is to minimize a global objective function. Many real science and engineering problems are investigated for distributed optimization, such as minimizing the transportation cost from the production center to the multiple distribution factories in [40,41] and the moving distance for the multi-robot system optimal formation problem [42], the shortest distance rendezvous problem in [43], and minimizing total power generation cost of distributed resource allocation problem investigated in Section 5.2.

3. Unconstrained distributed optimization

Considering the following unconstrained optimization problem

$$\min_{x\in\Re^n} \quad f(x) = \sum_{i=1}^{N} f_i(x). \tag{9}$$

It is obviously noted that problem (9) is equivalent to the optimization problem with decomposable objective function which all agents achieve a consensus while minimizing the total cost function, when the communication network is connected, that is

$$\begin{aligned} &\min_x \quad f(x) = \sum_{i=1}^{N} f_i(x_i), \\ &\text{subject to} \quad x_i = x_j \in \Re^n. \end{aligned} \tag{10}$$

Let optimal value be $f(x^*) = f^*$, where $x^* \in \Re^n$ is the optimal solution of problem (10). The optimal solution set $X^* = \left\{x_i \in \Re^n | \sum_{i=1}^N f_i(x_i) = f^*\right\}$ is nonempty and compact.

The existing classical distributed algorithms for addressing problem (10) contain the gradient method [22,23], the push-sum method [25,28], the fast gradient method [24,44], the alternating direction method of multipliers [7], average consensus [45,46] and distributed proportional integral (PI) control algorithm [47–49]. Compared with traditional centralized algorithms, when facing large dimension of the big data sets and the increased availability of computational resources, distributed algorithms decompose a large scale network into simple frameworks with lower dimension but also typically have more restrictive assumptions, such as bounded subgradient and convexity. The initial algorithms are based on gradient descent (GD) [50,51], where the convergence rates are subject to convex or strongly convex functions or bound assumptions.

3.1. Unconstrained distributed optimization with undirected communication graph

Now we introduce the classical distributed gradient descent (GD) approach to minimize a differentiable function f in (10) under undirected graphs. At node i, the distributed algorithm in [52] can be described as

$$x_{k+1}^{i}=\sum_{j=1}^{N}a_{ij}x_{k}^{j}-\eta^{k}\nabla f_{i}(x_{k}^{i}), \tag{11}$$

where x_k^i is a local state estimate at iteration k, a_{ij} is the connection weight between nodes i and j that comes its neighborhood, η^k is the step size at the kth iteration. The weight matrix A is doubly stochastic, that is $\mathbf{1}_n^T A=\mathbf{1}_n^T, A\mathbf{1}_n=\mathbf{1}_n$. It is worthy noted that GD method degenerates into average consensus when $\eta^k=0$, and the state x_k^i converges to $1/N\sum_{j=1}^{N}x_0^i$ for each agent, where x_0^i is the initial point of agent i. In other words, all agents reach the average of their initial states.

For further study, (11) is rewritten into a compact form for all agents together as follows

$$x_{k+1}=\widetilde{A}x_k-\eta^k\nabla\mathbf{f}_i(x_k), \tag{12}$$

where $x_k=\mathrm{col}\left((x^1)^T,\ldots,(x^N)^T\right)^T\in\Re^{Nn}, \widetilde{A}=A\otimes I_n\in\Re^{Nn\times Nn}$, and $\nabla\mathbf{f}_i(x_k)\in\Re^{Nn}$. Similarly, assume that $\eta^k=0$, the GD method degenerates into average consensus. When picking the diminishing step size in [17] where the step satisfies $\lim_{k\to\infty}\eta^k=0$ and $\sum_{k=1}^{\infty}(\eta^k)^2<\infty$, the agent state estimates generated by the GD algorithm can converge to the optimal solution of problem (10). With the conditions of the bounded subgradient and Lipschitz continuous objective function, [44] indicates that setting damping step sizes $\eta^k=1/\sqrt{k}$ results in $O\left(\ln k/\sqrt{k}\right)$ convergence rate. The algorithm in [24] has a convergence rate of $O(\ln k/k)$ with the damping step sizes $\eta^k=1/k^{\frac{1}{3}}$ when the communication graph has no inner loop. Note that the algorithms with damping step sizes are intuitive and computationally simple, but they are slow near the optimal solution. However, if the fixed step size $\eta^k=\eta$ is set, the GD method can not converge to the exact optimal solution of problem (10) but its approximate solution. In order to acquire the exact optimal solution with constant step-size, [26] proposes the EXTRA method by correcting the gradients iteration of DGD method, that is

$$\begin{cases} x_1^i = \sum_{j=1}^{N}a_{ij}x_0^j-\eta\nabla f_i(x_0^i),\\ x_{k+2}^i = x_{k+1}^i+\sum_{j=1}^{N}a_{ij}x_{k+1}^j-\sum_{j=1}^{N}\tilde{a}_{ij}x_k^j\\ \qquad -\eta\left[\nabla f_i(x_{k+1}^i)-\nabla f_i(x_k^i)\right], \end{cases} \tag{13}$$

where $\widetilde{A}=[\tilde{a}_{ij}]\in\Re^{N\times N}$ is a doubly stochastic mixing matrix, defined as $\widetilde{A}=(I+A)/2$. For the kth iteration, agent i computes the gradient $\nabla f_i(x_k^i)$ once and uses it twice for x_{k+1}^i and x_{k+2}^i. With the conditions of strongly convex and Lipschiz differentiable objective functions, the EXTRA method (13) can converge to the global optimal solution of problem (10) with a linear convergence rate $O\left(1/C^k\right)$, and has a convergence rate $O(1/k)$ when f_i is convex. Furthermore, [29] shows that algorithm (13) is a type of primal-dual gradient-like method [53,54], by denoting $\mathscr{L}=I-A$, then problem (10) can be rewritten as $\min_x f(x)=\sum_{i=1}^{N}f_i(x_i)$, subject to $\frac{1}{\sigma}\mathscr{L}^{1/2}x=0$, where σ is the penalty parameter. [30] proposes a generalized unifying primal-dual method based on the augmented Lagrangian function and primal-dual gradient [55,56], which has global R-linear convergence rate under strongly convex cost function f_i with Lipschitz continuous gradient. All these methods above assume that the networks have balanced (or undirected) graphs. However, in practice, it may not always be possible to assume undirected communication due to agents may transmit information at different power channels which means owning communication capability in one direction but has not in the opposite direction. Therefore, the distributed optimization algorithms are developing to be applicable to directed graphs.

3.2. Unconstrained distributed optimization with directed communication graph

For strongly connected digraph, the connection weight matrix is either column stochastic or row stochastic or doubly stochastic, which make up the deficiency of only doubly-stochastic weighted matrix in undirected graph. Some works in [31,57,58] have proposed the distributed optimization methods over digraphs. Therefore, when the weighted matrix in (12) is row stochastic, the agents/nodes do not reach the optimal solution [57]. Likewise, when the weighted matrix in (12) is column stochastic, the algorithm can not reach consensus since the right eigenvector of A corresponding to the eigenvalue of 1 is not $\mathbf{1}_n$ which is crucial for agreement [45,46,59]. Next, we formally introduce these issues.

The relative classical algorithm with column-stochastic weights is well known for the subgradient push method [31], which performs the following rules for each node i only with its local out-degree

$$\begin{cases} w_{t+1}^i=\sum_j\in\mathcal{N}_i^{in}\frac{x_t^j}{d_j},\\ y_{t+1}^i=\sum_j\in\mathcal{N}_i^{in}\frac{y_t^j}{d_j},\\ x_{t+1}^i=\sum_j\in\mathcal{N}_i^{in}\frac{w_{t+1}^j}{y_{t+1}^i},\\ z_{t+1}^i=w_{t+1}^i-\eta^{t+1}\mathbf{g}_{t+1}^i, \end{cases} \tag{14}$$

where y_t^i is a scalar variable, w_t^i is the auxiliary vector, $\mathbf{g}_{t+1}^i$ is the subgradient of function $f_i(x)$ at point x_{t+1}^i. The element a_{ij} in weighted matrix A is $1/d_j$ if $j\in\mathcal{N}_i^{in}$, and $a_{ij}=0$ otherwise. Under the assumptions that the graph sequence $\{G(k)\}$ is uniformly strongly connected and each function f_i is convex with uniformly bounded subgradients $\mathbf{g}^i$, the approach is initiated with an arbitrary vector $z_0^i\in\Re^n$ and a scalar $y_0^i=1$ at node i. The subgradient terms in the updates of z_{t+1}^i are designed to guide the consensus point towards the optimal solution x^*, while the push-sum updates terms to lead the vectors x_{t+1}^i to each other, then the objective function in (9) or (10) converges to its optimal value f^*. [58] designs a push-sum approach using the left eigenvector to eliminate the imbalance. By combining the push-sum consensus strategy with dual methods, [60] presents a push-sum distributed dual averaging method which converges to the true average consensus without knowing the update matrix in advance.

As other type of directed connected graph with row-stochastic weighted matrix, the corresponding distributed algorithm comes into being. In [57], the agent/node i adopts the following updates at the kth iteration

$$\begin{cases} x_{k+1}^i = \sum_{j=1}^{N} a_{ij} x_k^j - \eta z_k^i, \\ y_{k+1}^i = \sum_{j=1}^{N} a_{ij} y_k^j, \\ z_{k+1}^i = \sum_{j=1}^{N} a_{ij} z_k^j + \frac{\nabla f_i(x_{k+1}^i)}{[y_{k+1}^i]_i} - \frac{\nabla f_i(x_k^i)}{[y_k^i]_i}, \end{cases} \tag{15}$$

where each agent i has three vectors, x_k^i, y_k^i and z_k^i for the ith iteration, $[y_k^i]_i$ represents the ith element of y_k^i. The algorithm is initiated with an arbitrary vector $x_0^i, y_0^i = (0, \ldots, 1, \ldots, 0)^T$, where the ith element is 1, and $z_0^i = \nabla f(x_0^i)$ at node i. Each agent has an additional variable that asymptotically converges to the left eigenvector of the row-stochastic weight matrix $\mathbf{1}_n \pi^T x_k$ while achieves the best known convergence rate $O(\mu^k)$. In other words, there is a constant $M > 0$ such that $\|x_k - \tilde{x}^*\| \leqslant M(\gamma + \epsilon)^k$, where $\tilde{x}^* = x^* \otimes \mathbf{1}_n$ is the optimal solution to problem (9), $0 < \gamma < 1, \gamma < \mu < 1$ and ϵ is an arbitrarily small positive constant. Unlike column stochastic matrix, row stochastic matrix is easier to receive on account of each agent autonomously determines the weights of the in-neighbors information. According to the unique character, [61] designs a distributed projection subgradient algorithm with row-stochastic matrix. Furthermore, to suit for the directed networks with row-stochastic weights, [62] proposes a distributed Nesterov-like gradient tracking algorithm with momentum terms and nonuniform step-sizes, in which the algorithm linearly converges to the optimal solution in (9) or (10) owning smooth and strongly convex objective function f_i.

3.3. Unconstrained distributed optimization with finite/fixed-time consensus

All the algorithms mentioned above can only arrive at consensus over an infinite time horizon, that is, the number of iterations $k \to \infty$ in discrete-time system and $t \to \infty$ in continuous-time system. Some researchers design a suitable distributed control laws [47,27] to drive the agents cooperatively reach consensus in finite-time or fixed-time to solve the convex optimization problem (9). Generally, the distributed protocol includes two parts that one is for finite-time or fixed-time consensus and the other is for the optimality. A widely used distributed control protocol is PI control, which depends on the exact relative state $x_i - x_j$ with neighbored agents and steers each agent moving towards its neighbors. Consider the following PI controller

$$\begin{aligned} u^i(t) &= \kappa_P \sum_{j=1}^{N} a_{ij}\left(x^j(t) - x^i(t)\right) \\ &\quad + \kappa_I \int_0^t \sum_{j=1}^{N} a_{ij}\left(x^j(s) - x^i(s)\right) ds, \end{aligned} \tag{16}$$

where proportional control gain κ_P and integral control gain κ_I are positive constants. [63] proposes a dynamics equation $\dot{x}^i = -\nabla f_i(x^i(t)) + u^i(t)$ for the ith agent. Based on the dynamic system, the designed $u^i(t)$ employs local interaction and local information to drive all agents cooperatively reaching the optimal solution x^* while solving problem (9) or (10). Combined the dynamic system and distributed control protocol (16), there is

$$\begin{cases} \dot{\mathbf{x}} = -\nabla \mathbf{f}(\mathbf{x}) - \kappa_P \mathbb{L}\mathbf{x} - \mathbf{y}, \\ \dot{\mathbf{y}} = \kappa_I \mathbb{L}\mathbf{x}, \end{cases} \tag{17}$$

Eqn. (17) where $y^i(t) = \kappa_I \int_0^t \sum_{j=1}^{N} a_{ij}(x^j(s) - x^i(s))ds$ with $\mathbf{y}(0) = 0, \mathbf{y} = \mathrm{col}\left((y^1)^T, \ldots, (y^N)^T\right)^T \in \Re^{Nn}, \nabla \mathbf{f}(\mathbf{x}) = \mathrm{col}\left(\nabla f_1(x^1)^T, \ldots, \nabla f_N(x^N)^T\right)^T \in \Re^{Nn}, \mathbb{L} = L \otimes I_n$. It can be deduced that the equilibrium point of system (17) satisfies $\mathbf{x}^* \in \mathrm{span}\{\mathbf{1}_N \otimes \upsilon\}$, where $\upsilon \in \Re^n$ is an arbitrary vector, and $\mathbf{1}_N^T \nabla \mathbf{f}(\mathbf{x}^*) = \sum_{i=1}^{N} \nabla f_i(x^*) = \mathbf{0}$. Thus the equilibrium of system (17) is an optimal solution of problem (10). [47] designs a continuous-time dynamic model to solve problem (9) by controlling the sum of subgradients of the global optimal function to force the state $x_i(t)$ to the optimal solution point. In addition, the control protocol in [47] is still appropriate for general linear systems [64], second-order system [65], and multi-agent system with time-varying communication delays [32]. To deal with a distributed rendezvous problem with shortest distance, [43] presents a distributed finite-time control protocol based on sign function, which shows that the agents can rendezvous in finite time and move towards to an optimal point to minimize the objective function as $t \to \infty$. Furthermore, [27] designs a distributed edge-based fixed-time consensus protocol to address time-invariant cost function in problem (10) and time-varying case.

4. Distributed optimization with constraint of intersection by convex sets

This section considers the following constrained optimization problem described as

$$\begin{aligned} \min_x \quad & f(x) = \sum_{i=1}^{N} f_i(x), \\ \text{subject to} \quad & x \in \bigcap_{i=1}^{N} \Omega_i, \end{aligned} \tag{18}$$

where $x = (x_1, x_2, \ldots, x_n)^T \in \Re^n, f_i : \Re^n \to \Re$ is a nonsmooth convex function, $\Omega_i \subseteq \Re^n$ is a convex compact set, which is only known to agent i. This problem is generally regarded as distributed optimization with constraint of intersection by convex sets and has been widely investigated [34,66]. Note that problem (18) is made up of the distributed decomposable objective functions (7) and bound constraints Ω_i. With the previous analysis, for an undirected connected graph, problem (18) is equivalent to the following optimization problem with decomposable objective function and variables

$$\begin{aligned} \min_x \quad & \tilde{f}(\tilde{x}) = \sum_{i=1}^{N} f_i(x_i), \\ \text{subject to} \quad & L\tilde{x} = 0, \ \tilde{x} \in \Omega, \end{aligned} \tag{19}$$

where $\tilde{x} = \mathrm{col}(x_1, x_2, \ldots, x_N) \in \Re^{Nn}, L = L_N \otimes \mathbf{I}_n \in \Re^{Nn \times Nn}$, and $\Omega = \prod_{i=1}^{N} \Omega_i$ represented by the Cartesian product.

According to the Karush-Kuhn-Tucker (KKT) conditions [67], $\tilde{x}^* \in \Re^{Nn}$ is an optimal solution to problem (19) if and only if there exists vector $y^* \in \Re^{Nn}$ such that $(\tilde{x}^*, y^*)$ satisfies the following equations

$$\begin{cases} \tilde{x}^* - g\left(\tilde{x}^* - \partial \tilde{f}\left(\tilde{x}^*\right) - Ly^*\right) = 0, \\ Ly^* = 0, \end{cases} \tag{20}$$

where $\partial \tilde{f}(\tilde{x}^*)$ is the subgradient of $\tilde{f}$ at $\tilde{x}^*$, and map $g(\cdot)$ is a projection operator from $\Re^{Nn}$ to Ω.

The earlier works focus on a special case where the bound constraint sets are identical, i.e., $\Omega_1 = \Omega_2 = \cdots = \Omega_N$. For this case, [35] designs a distributed projected subgradient method where the damping step size satisfies $\sum_k \eta_k = \infty$ and $\sum_k \eta_k^2 < \infty$. Each agent executes the distributed algorithm (11) and then projects the $k+1$ iteration x^i_{k+1} into the identical constraint set under the assumption of bounded subgradient set of each f_i. In continuous-time dynamic system, [33] proposes a distributed protocol containing local information averaging, local subgradient and local projection, which guarantee that all state estimates asymptotically converge to the optimal point of (18) within the common set constraint. By relaxing the continuity and differentiability assumptions to nonsmooth local objective functions, [68] proposes an exact penalty algorithm to ensure that all the agents reach an global optimal solution to (18). The aforesaid approaches are applied for connected undirected communicating interaction graph. As for a directed uniformly jointly (or B-strongly) connected communication graph, [69] develops a distributed dual averaging push-sum algorithm with a decoupled structure with push-sum weight step and subgradient step in different update flows, and the two terms are mixed together to update the primal variable and recover the primal solution. It indicates that all the agents reach a consensus and seek for a solution to problem (18).

The general case is that the constraint sets Ω_i are not identical. For solving the distributed optimization problem (19) with different hyper-box or hyper-sphere constraint sets, [70] designs a second-order multi-agent network in form of the following differential inclusions

$$\begin{cases} \frac{d\tilde{x}}{dt} \in 2\Big[-\tilde{x} + g\Big(x - \partial\tilde{f}(\tilde{x}) - L(\tilde{x} - y)\Big)\Big], \\ \frac{d\tilde{y}}{dt} = \tilde{x}. \end{cases} \tag{21}$$

Based on the KKT conditions (20) and Lyapunov method [71], it shows that the state vector $\tilde{x}$ of the second-order multi-agent network (21) is globally convergent with consensus to an optimal solution.

Unlike [70], for bounded constraints, which makes the projection operation on both state vector $\tilde{x}$ and gradient information $\partial\tilde{f}$, as well as the coupling connection with agent i and agent j, [72] proposes a continuous-time multi-agent system for bound constrained distributed optimization with only primary state x_i being projected from $\Re^n$ to Ω_i. Meanwhile, a proportional-integral (PI) distributed controller is utilized in the system for converging to an optimal solution of problem (19). In [73–75], the collective neurodynamic approach is proposed for solving the bounded constrained problem (18). However, different from neurodynamic approaches proposed in [76] for distributed optimization, these neurodynamic models are not coupled and consensus is not necessary. By relaxing the hyper-box or hyper-sphere bounded constraints into the assumption that the set $\bigcap_{i=1}^N \Omega_i$ only at least has one interior point, [77] presents a distributed primal-dual approach with a constant fixed step size, which drives that the agents asymptotically reach a consensus at an optimal solution under connected graph. Considering a random communication graph, [36] develops a distributed asynchronous constrained stochastic consensus algorithm with diminishing step sizes for solving problem (18). For global bound constraint set, [78] designs a continuous-time distributed hierarchical algorithm on account of projected gradient method to ensure all the estimates reach consensus of the optimal solution as iterations increase. Note that these existing aforementioned optimization algorithms acquire each node to compute the projection on its local constraint set, but it is hard to implement. To improve the convergence rate, [37] proposes a distributed primal-dual algorithm with a constant step size on basis of the augmented Lagrange method. By virtue of the equivalent formulas in (20), a distributed continuous-time algorithm based on nonsmooth analysis is presented in [79] for nonsmooth convex local cost functions, in which all the agents can look for the uniform optimal solution, and meanwhile keep the states stable.

In particular, if the objective function $f_i(x)$ is the quadratic form $f_i(x) = \sum_{i=1}^N |x|^2_{\Omega_i}$, where $|x|_{\Omega_i} = \inf_{y\in\Omega_i}|x-y| = |x - P_{\Omega_i}(y)|$, and P_{Ω_i} is the projection operator onto the closed convex set Ω_i, then (18) is considered as the shortest distance optimization problem for seeking a point with the shortest quadratic distance to the sets Ω_i, such as minimizing the transportation cost from the production center to the multiple distribution factories [40,41] and convex intersection problems [43,80,81]. For connected graphs, [43] proposes two distributed continuous time algorithms to solve the shortest distance optimization problem for the convex empty intersection case, i.e., $\bigcap_{i=1}^N \Omega_i = \phi$: the first one is designed to achieve optimal consensus by virtue of sign functions, and the second one is to reach the approximate optimal solution. Different from the exact projection in [43,80] presents a discrete-time approximate projected consensus method by setting the approximate projection points onto Ω_i located in a triangle region. Moreover, [81] proposes a distributed continuous-time approach for time-varying communication graph, where the approximate projection is related to the boundary surfaces of Ω_i other than triangle region, which is suitable for the cases of $\bigcap_{i=1}^N \Omega_i = \phi$ and $\bigcap_{i=1}^N \Omega_i \neq \phi$.

5. Distributed optimization with equality constraint

In Section 4, the distributed optimization with local bounded constraint set has been reviewed. In this section, we consider the distributed optimization problem with equality constraint in addition to the bound constraint.

5.1. A general framework of distributed optimization with equality constraint

Considering the distributed optimization problem with global coupled equality constraint as follow

$$\begin{aligned} \min_x \quad & f(x) = \sum_{i=1}^N f_i(x), \\ \text{subject to} \quad & B_i x = b_i, x \in \bigcap_{i=1}^N \Omega_i, \end{aligned} \tag{22}$$

where $x = (x_1, x_2, \ldots, x_n)^T \in \Re^n, f_i : \Re^n \to \Re \ (i = 1, 2, \ldots, N)$ is the objective function, matrix $B_i \in \Re^{p_i \times n}$, vector $b_i \in \Re^{p_i}$, and $\Omega_i \subseteq \Re^n$ is a convex compact set. Assume $\bigcap_{i=1}^N \Omega_i$ is not empty, i.e., there exists feasible solution to problem (22). According to the equivalent conditions in [48], problem (22) is equivalent to the following optimization problem

$$\begin{aligned} \min_{\tilde{x}} \quad & \tilde{f}(\tilde{x}) = \sum_{i=1}^N f_i(x^i), \\ \text{subject to} \quad & B\tilde{x} = b, L\tilde{x} = 0, \tilde{x} \in \Omega, \end{aligned} \tag{23}$$

where $\tilde{x} = \mathrm{col}(x^1, x^2, \ldots, x^N) \in \Re^{Nn}, x^i \in \Re^n$ is the estimated solution of (22) by agent i, the block diagonal matrix $B = \mathrm{diag}\{B_1, B_2, \ldots, B_N\} \in \Re^{p \times Nn}, b = \mathrm{col}(b_1, b_2, \ldots, b_N) \in \Re^p$ and $p = \sum_{i=1}^N p_i, L = L_N \otimes I_n$ with L_N being the Laplacian matrix of the undirected connected communication graph of the network, and $\Omega = \prod_{i=1}^N \Omega_i$ defined by the Cartesian product. In fact, assume $\tilde{x}^* = \mathbf{1}_N \otimes x^*$ to be an optimal solution of (23). On one side, since $L\tilde{x}^* = (L_N \otimes I_n)(\mathbf{1}_N \otimes x^*) = (L_N \mathbf{1}_N) \otimes x^* = 0$, then x^* is an optimal solution of (22). On the other side, if x^* is an optimal solution of (22), $\mathbf{1}_N \otimes x^* = \tilde{x}^*$ is an optimal solution of (23) under the condition that the communication graph is connected.

According to the KKT conditions applied in [34], $\tilde{x}^* \in \Re^{Nn}$ is an optimal solution to problem (23) if and only if there exist $y^* \in \Re^p$ and $z^* \in \Re^{Nn}$ such that $(\tilde{x}^*, y^*, z^*)$ satisfies the following equations

$$\begin{cases} \tilde{x}^* - g\left[\tilde{x}^* - \eta\left(\nabla \tilde{f}\left(\tilde{x}^*\right) + By^* + Lz^*\right)\right] = 0, \\ B\tilde{x}^* = b, L\tilde{x}^* = 0, \end{cases} \tag{24}$$

where $g(\cdot)$ is a projection map from $\Re^{Nn}$ to Ω.

To solve the optimization problem in (23), [82] presents a distributed constrained consensus algorithm with fixed step size over multi-agent networks on account of (24) as follows

$$\begin{cases} x_i(k+1) = g_i\left[x_i(k) - \eta\left(\nabla_i f(x_i(k)) + B_i^T(y_i(k) + B_i x_i(k) - b_i) + \sum_{j=1, j\neq i}^{N} a_{ij}(x_i(k) - x_j(k))\right)\right], \\ y_i(k+1) = y_i(k) + B_i x_i(k+1) - b_i, \\ z_i(k+1) = z_i(k) + \sum_{j=1, j\neq i}^{N} a_{ij}(x_i(k) - x_j(k)), \end{cases} \tag{25}$$

where a_{ij} is the connection weight between agents i and j in the communication graph, and (25) gives the new state estimate of agent i at the $k+1$ iteration, which makes an estimate solution to (22).

If the function $f_i(x)$ is convex and differentiable, its gradient $\nabla f_i(x)$ is Lipschitz continuous, and the communication graph is undirected and connected, the state vector $x_i(k)$ in (25) globally reach consensus at an optimal solution of problem (22) under the situation of the step size $\eta < 1/\left[2L_f + \lambda_{\max}\left(B^T B + L\right)\right]$. In addition, if $f_i(x)$ is twice continuously differentiable and $\nabla^2 f_i(x)$ is bounded on Ω_i, $x_i(k)$ in (25) globally reach consensus at an optimal solution of problem (22) under the situation of the step size $\eta < 1/\left[2\max_{\tilde{x}\in\Omega} + \lambda_{\max}\left(B^T B + L\right)\right]$.

Meanwhile, problem (22) also can be treated as nonlinear programming with equality and bound constraints, where the classical interior and penalty methods [83] are applicable. By eliminating equality constraint and penalty function, [18] designs a projection neural network for problem (22) which guarantees the stability and global convergence to the optimal solution. However, for large scale optimization, compared with the centralized algorithms proposed in [18,83] which require global knowledge of the whole network and lack of robustness due to time-varying nature of the network, it is meaningful to devise distributed algorithms for problem (22) to simplify communication with lower dimension and make sure individual privacy.

Actually, for undirected connected graphs, the consensus or projection consensus approach has motivated distributed algorithms for solving linear algebraic equations [84–86], which decompose a large scale linear equalities system into smaller ones that can be cooperatively addressed by multi-agent networks.

Inspired by the discussion on solving linear equations $B_i x = b_i$ in [87,88] designs a distributed algorithm for multi-agent networks to acquire the optimal solutions of least square function $1/2\sum_{i=1}^{N}\|B_i x - b_i\|^2$ with consensus equality constraint $Lx = 0$, in which the distributed optimization problem is a special form of (23). By defining $h_i(x) = B_i x - b_i = 0$, [38] proposes a distributed penalty primal-dual subgradient algorithm under the assumption that f_i is continuous and convex. [89] proposes a penalty-like neurodynamic method to relax the requirement of the gradients of local objective functions, which extends the feasibility of these gradient-based approaches to a certain extent. Meanwhile, for problem (22) with a directed communication graph, [90] develops a nonnegative surplus-based distributed optimization algorithm for the unidirectional information exchanging as a result of nonuniform communication powers in practice, which shows that the algorithm converges to the optimal solution. Recently, [91] presents a distributed projected approach with time-varying piecewise step size which is not damping to zero. It indicates that the agents converge to the optimal solution of problem (22) under the condition that the objective function is convex rather than strongly convex.

Unlike the global coupled equality constraint $B_i x = b_i$ in problem (22), there is another type of optimization problem with decoupled equality constraints. Besides each agent owns local constraint variable $x_i \in \Omega_i$, these variables x_i are coupled together through the global decoupled linear equality constraint $\sum_{i=1}^{N} B_i x_i = \sum_{i=1}^{N} b_i$ [92–95] other than $B_i x = b$, or directly setting as $\sum_{i=1}^{N} B_i x_i = b/N$ [96–98]. An alternating direction method of multipliers (ADMM) distributed optimization algorithm is investigated in [96]. The algorithm has faster convergence rate with inexact step size under strongly convex and Lipschitz continuous gradient assumptions for undirected connected graph. By relaxing the conditions to require only convex functions, [97] shows a global linear convergence ADMM for large scale multi-block optimization models. Different from [96,97,94] designs a distributed neurodynamic optimization approach for solving large-scale problems based on projected recurrent neural network. Recently, [99] investigates problem (22) with linear separable equality constraint $\sum_{i=1}^{N}(B_i x_i - b_i) = 0$, where a regularized dual gradient based distributed algorithm is designed. Based on the push-sum protocol and dual decomposition, it indicates that the proposed algorithm has a convergence rate $O(\ln T/T)$ under the assumption of strongly convex local objective functions for unbalanced directed graphs. However, the aforementioned decomposition methods rely heavily on the decomposable structure of the dual function, which has proverbial disadvantages, such as slow convergence rate, nonuniqueness of solutions and inexact optimal solution. To alleviate these drawbacks, the augmented Lagrangian method (ALM) framework [100] is applied, which is analogous to the regularization of the dual function [15]. The problem of the form (22) with separability structure of equality constraints is called extended monotropic optimization problem in [101]. Enlightened by the discussion, [102] proposes an accelerated distributed augmented Lagrangian method (ADALM) with fixed step size by establishing successive separable approximations of the quadratic term $\|\sum_{i=1}^{N} B_i x_i - b\|^2$, which addresses the major drawback of ALM that the quadratic penalty term is not amenable to decomposition [100]. It shows that any sequence of x_i^k generated by the ADALM has an accumulation point and any such point is an optimal solution of problem (22).

5.2. A special framework of distributed optimization with equality constraint

Especially, the distributed resource allocation problem is an application case of (22) where a group of power generators attempts to satisfy power generation supply-demand balance equality constraint $\sum_{i=1}^{N} x_i = D$, while minimizing the total power generation cost $f(x)$ and achieving individual generation capacity constraints Ω_i, which can be mathematically formulated as

$$\begin{aligned} \min_x \quad & f(x) = \sum_{i=1}^{N} f_i(x_i), \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{N} x_i = D, \\ & x_i \in \Omega_i = \left[x_i^{\min}, x_i^{\max}\right], i = 1, 2, \ldots, N, \end{aligned} \tag{26}$$

where the scalar quantity x_i is the power generation of the ith power generator, positive constant N is the number of power generators, $D\left(D = \sum_{i=1}^{N} d_i\right)$ is the total power demand (d_i is the power demand of generator i), $f_i(\cdot) : \Re \to \Re$ is the cost of power generation, and $x_i^{\min}$ and $x_i^{\max}$ represent the lower and the upper bounds of power generation i, respectively. Assume the local cost function $f_i(x_i)$ be strictly convex and twice continuously differentiable on Ω_i for all $i = 1, 2, \ldots, N$.

The problem constructed in (26) is a representative in energy resource, and many literatures have researched various applications of resource allocation, such as energy internet management [94], transportation systems [103], supply chain network [104], network utility maximization systems [105] and economic dispatch for microgrids [106]. On account of the wide representation of problem constructed in (26), the distributed consensus and optimization for distributed resource allocation problems have been briefly discussed in a few survey papers [5,107]. Thus, the main target here is to retrospect distributed resource allocation algorithms.

To solve problem (26), we consider the following Lagrangian function

$$\mathscr{L}(x, \nu) = \sum_{i=1}^{N} f_i(x_i) + \nu\left(\sum_{i=1}^{N} d_i - \sum_{i=1}^{N} x_i\right),$$

where $x = (x_1, x_2, \ldots, x_N)^T$ and $\nu \in \Re$ is Lagrangian multiplier. Then the corresponding Lagrangian dual problem is obtained as

$$\max_{\nu} \; g(\nu) = \sum_{i=1}^{N} g_i(\nu), \tag{27}$$

where $g(\nu) = \inf_{x_i \in \Omega_i} \sum_{i=1}^{N} (f_i(x_i) + \nu(d_i - x_i))$. Therefore, the constrained problem (26) can be solved by virtue of its dual problem (27), which has the form of unconstrained optimization problem (7). Compared with problem (7), the only distinction is that problem (27) is a maximization dual problem. Thus, the distributed algorithms reviewed in Section 3 for the optimization problem (7) can be utilized.

The earlier work focused on problem (26) is that the power cost function f_i is quadratic and the topology graph of communication network is undirected. [3] proposes an incremental cost leader-follower consensus algorithm in a distributed manner in which the leader detects the mismatch between the total and the demand power generations, and then drives the renewals of the incremental cost. It illustrates that the distributed approach converges to the optimal power generations while subject to the power balance constraint under different undirected communication topologies with minimal spanning tree. Inspired by [26,108] reformulates the resource allocation problem (26) and develops a distributed primal-dual method with a fixed constant step size which converges to the optimal solution and has a fast convergence rate. Different from the discrete time method proposed in [108], by importing an adaptive parameter switching strategy, [109] designs a continuous-time distributed primal-dual approach to seek for the optimal solution. Taking more practical application scenarios into consideration in information communication, [4] proposes a distributed consensus algorithm on basis of the push-sum method in [31] with diminishing step sizes for time-varying jointly connected directed communication graphs. Considering communication time-delay, [106] presents a delay-free-based distributed algorithm and [92] studies a continuous-time distributed algorithm over multi-agent system with weight-balanced and strongly connected directed graph to reduce communication cost. Furthermore, for both uniformly jointly strongly connected time-varying communication digraph and strongly connected fixed unchange communication digraph, [110] designs a distributed push-pull method for solving problem (26) with a fixed constant step size. Moreover, the distributed push-pull algorithm can be regarded as a generalization of the distributed optimization algorithm presented in [111], which is suitable for general strictly convex non-quadratic cost functions with time-varying communication graphs.

6. Distributed Optimization with Inequality Constraint

In this section, we consider the distributed constrained optimization problem with inequality constraint in addition to the bound or equality constraint.

6.1. The case of only inequality constraint

Firstly, this section considers the case which only contains global inequality constraints without local constraints. The constrained optimization problem is given by

$$\begin{aligned} \min_{x} \quad & f(x) = \sum_{i=1}^{N} f_i(x), \\ \text{subject to} \quad & g_i(x) \leqslant 0, \end{aligned} \tag{28}$$

where $x \in \Re^n$ is a global state decision vector, $f_i : \Re^n \to \Re$ $(i = 1, 2, \ldots, N)$ is the objective function only known by node i, and convex function $g_i(x)$ is a linear or nonlinear mapping from $\Re^n$ to $\Re^{q_i}$. To deal with the inequality constraint $g_i(x)$, a regularized Lagrangian function $L(x, \nu)$ is introduced as follows

$$\begin{aligned} L(x, \nu) &= \sum_{i=1}^{N} f_i(x) + \nu_i^T N g_i(x) \\ &= \sum_{i=1}^{N} (f_i(x) + \nu_i^T g_i(x)) \\ &= \sum_{i=1}^{N} L_i(x, \nu_i), \end{aligned} \tag{29}$$

where $\gamma > 0$ is a parameter, the Lagrangian multiplier $\nu_i \in \Re_+^{q_i}$, and $L_i(x, \nu_i) = f_i(x) + \nu_i^T g_i(x)$. Then the Lagrange dual function is $q(\nu_i) = \inf_x \sum_{i=1}^{N} L_i(x, \nu_i)$, and the Lagrangian dual problem associated with the primal problem (28) is defined as

$$\begin{aligned} \max_{\nu} \quad & q(\nu), \\ \text{subject to} \quad & \nu \in \Re_+^p, \end{aligned} \tag{30}$$

where $\nu = \mathrm{col}(\nu_1, \nu_2, \ldots, \nu_N)$ and $p = \sum_{i=1}^{N} p_i$. Denote d^* as the dual optimal value of the Lagrangian dual problem (30) and f^* is the primary optimal value of problem (28), under the Slater's condition [12], there is $d^* = f^*$. Let x^* and ν^* are the primary optimal solution and the dual optimal solution, respectively. Based on (29), (x^*, ν^*) is a saddle point of $L(x, \nu)$ if $L(x^*, \nu) \leqslant L(x^*, \nu^*) \leqslant L(x, \nu^*)$. To characterize the saddle point of the Lagrangian function (29), the Lagrangian saddle-point theorem [112] displays that (x^*, ν^*) is a saddle point if and only if it is a pair of primal and Lagrangian dual optimal solutions and the following strong minimax property holds

$$\sup_{\nu \in \Re_+^{p_i}} \inf_{x} L(x, \nu) = \inf_{x} \sup_{\nu \in \Re_+^{p_i}} L(x, \nu), \tag{31}$$

On account of (31), [113] shows that if (x^*, ν^*) is a saddle point of the Lagrangian function (29), then $L(x^*, \nu^*) = f^*$.

Based on the Lagrangian saddle point theorem, [39] designs a distributed consensus-based regularized primal-dual subgradient algorithm with a constant step size to solve problem (28). By regarding the global inequality constraint as a bound constraint set, the methods reviewed in Section 4 can be used to address this problem. However, most of the methods require to project the estimate states onto the constraint set at every iteration, while the distributed regularized primal-dual subgradient approach only requires one projection at the last iteration. In addition, an explicit convergence rate of $O\left(k^{-1/4}\right)$ for error $\|f(x)-f(x^*)\|$ is built for time-varying weighted-balanced directed communication graph which is jointly strongly connected. Instead of discrete-time optimization method, [114] develops a distributed continuous-time model with differential inclusion auxiliary function, which relaxes the strong convexity and reduces the number of state variables. It shows that the system converges with consensus and finally reaches the optimal solution of problem (28) for both directed and undirected strongly connected graphs. Inspired by [39], a distributed online primal-dual push-sum algorithm is developed in [115] to solve problem (28) for the case with coupled inequality constraint $g(x)=\sum_{i=1}^{N}g_i(x_i)$, which does not count on the boundedness of Lagrange multipliers and is suitable for unbalanced communication digraphs. Considering the globally coupled inequality constraint $\sum_{i=1}^{N}g_i(x_i)\leqslant 0$, based on (31), a distributed primal-dual perturbation algorithm is proposed in [116] to construct approximate saddle points of the Lagrangian function of (28) under undirected and connected communication graph. For directed interaction communication graph, under the assumption that local objective function f_i is convex, [117] designs a distributed random-fixed projected algorithm, which transforms the optimization problem (28) with convex function to linear function within a convex set inspired by [12]. It indicates that the algorithm guarantees each node asymptotically converges to the optimal solution for strongly connected communication digraphs.

6.2. The case of inequality and bound constraints

Next, we consider the case with globally coupled inequality constraint and local bound constraint as the following problem

$$\begin{aligned} \min_{x} \quad & f(x)=\sum_{i=1}^{N}f_i(x), \\ \text{subject to} \quad & g_i(x)\le 0, i=1,2,\ldots,N, \\ & x\in\Omega=\bigcap_{i=1}^{N}\Omega_i, \end{aligned} \tag{32}$$

where $g_i(x):\Re^n\to\Re^{q_i}$ is convex functions known by agent i in the network. Compared with problem (28), the local bound constraint is added in problem (32).

The earlier researches are mainly centered on the special case where the local constraints sets Ω_i are identical. For periodical strongly connected time-varying balanced communication directed graphs, [113] devises a distributed penalty primal-dual subgradient algorithm with damping step size by virtue of saddle-point theorem and penalty function to guarantee all the agents to asymptotically agree on optimal solutions. However, [118] proposes a distributed primal-dual subgradient algorithm with constant step size which provides an approximate solution with a given accuracy in a finite number of iterations based on the Lagrangian function (29). For undirected connected communication graph, it reveals that the optimal value can be found within the error level relying on the number of consensus updates. Moreover, the algorithm has better convergence rate in comparison with the primal-dual subgradient approach in [113], which performs more than one consensus update at each consensus step.

For the case with different local constraints sets Ω_i, [119] considers a convex optimization problem with randomly occurring local convex constraint sets, and a distributed primal-dual random projection subgradient algorithm with diminishing step size is proposed to cope with the problem for time-varying strongly connected double stochastic communication graphs. In view of the effects of stochastic errors during computing the subgradient, [120] designs a distributed primal-dual stochastic subgradient algorithm for two models, which leverages the distributed average consensus algorithm to evolute the algorithm and acquires convergence bounds with diminishing step size for the synchronous time model. As for the asynchronous time model, the gossip algorithm is applied as a mechanism to establish the optimal error bound with a fixed step size based on saddle-point theorem for undirected connected graph.

In addition, the case in which globally coupled inequality constraints in problem (32) is separable with respect to $x_1,x_2,\ldots,x_N$, which can be expressed in the form of

$$\begin{aligned} \min_{x_i} \quad & f(x)=\sum_{i=1}^{N}f_i(x_i), \\ \text{subject to} \quad & g(x)=\sum_{i=1}^{N}g_i(x_i)\leqslant 0, \\ & x_i\in\Omega_i, i=1,2,\ldots,N, \end{aligned} \tag{33}$$

where $x_i\in\Re^{n_i}$ is the local vector of agent $i, x=\left(x_1^T,x_2^T,\ldots,x_N^T\right)^T\in\Re^n\left(\sum_{i=1}^{N}n_i=n\right), g_i:\Re^{n_i}\to\Re^q$ is the local constraint and $f_i:\Re^{n_i}\to\Re$ is the local objective function.

To solve problem (33) with separable inequality constraints, based on the Lagrangian function (29), [121] establishes a distributed continuous-time projected primal-dual subgradient method by virtue of a modified Lagrangian function. It shows that the algorithm is stable and the vector (x,v) converges to saddle point for connected and undirected graph under the assumption with strictly convex function f_i and convex function g_i. By relaxing the strictly convex assumption to be convex and nonsmooth, [122] provides a modified continuous-time primal-dual algorithm, which guarantee the existence of Caratheodory solutions and the convergence of the algorithm to an optimal solution for the problem. Based on the Lagrange duality framework (29) and ADMM aforementioned, [123] constructs a distributed relaxed hybrid ADMM algorithm with an adaptive penalty parameter selection scheme, which drives the state converges to the optimal solution of dual problem (30). According to averaging consensus in the dual decomposition method and saddle-point theorem, the primal optimal solution is obtained. However, without any averaging mechanism, [124] proposes a fully distributed duality approach with a diminishing step size. On basis of a relaxation of the primal problem and an exploration of duality theory, the algorithm drives each node to seek for a pair of primal-dual optimal solution and then to update auxiliary local estimates on several duality steps. It displays that the all agents asymptotically arrive at an optimal solution of the original problem (33).

6.3. The case of inequality and equality constraints

In this subsection, we consider the optimization problem with globally coupled inequality and equality constraints as follows

$$\begin{aligned} \min_{x} \quad & f(x)=\sum_{i=1}^{N}f_i(x), \\ \text{subject to} \quad & g_i(x)\le 0, \\ & h_i(x)=0, i=1,2,\ldots,N, \end{aligned} \tag{34}$$

where $x\in\Re^n$, for agent $i, f_i:\Re^n\to\Re$ is the local convex objective function, $g_i(x):\Re^n\to\Re^{q_i}$ is the globally coupled convex inequality constraint, and the affine function $h_i(x):\Re^n\to\Re^{p_i}$ is the global couple equality constraint defined as $h_i(x)=B_ix-b_i$ with $B_i\in\Re^{p_i\times n}$ and $b_i\in\Re^{p_i}$ $(p_i\leqslant n)$.

For an undirected connected communication graph, the problem in (34) can be transformed into an equivalent distributed optimization problem with only adding inequality constraint $\mathbf{g}(\tilde{x}) \leqslant 0$ compared with (23), in which $\mathbf{g}(\tilde{x}) = \left(g_1(x_1)^T, g_2(x_2)^T, \dots, g_N(x_N)^T\right)^T \in \Re^q\left(\sum_{i=1}^N q_i = q\right)$. Based on the KKT conditions, $\tilde{x}^* \in \Re^{Nn}$ is an optimal solution to problem (34) if and only if there exist $\omega^* \in \Re^q$ and $z^* \in \Re^{Nn}$ such that $(\tilde{x}^*, \omega^*, z^*)$ satisfies the following equations

$$\begin{cases} \mathbf{P}\tilde{x}^* - \mathbf{q} + (\mathbf{I} - \mathbf{P})\left(\partial \tilde{f}\left(\tilde{x}^*\right) + \left(\partial \mathbf{g}\left(\tilde{x}^*\right)\right)^T \omega^* + \mathbf{L}z^*\right) = 0, \\ \omega^* = \left(\omega^* + \mathbf{g}\left(\tilde{x}^*\right)\right)^+, \\ \mathbf{L}\tilde{x}^* = 0, \end{cases} \tag{35}$$

where project matrix $\mathbf{P} = B^T\left(BB^T\right)^{-1}B, \mathbf{q} = B^T\left(BB^T\right)^{-1}b$, the block diagonal Jacobian matrix $\partial\mathbf{g}(\tilde{x}) = \mathrm{diag}\{\partial g_1(x_1), \partial g_2(x_2), \dots, \partial g_N(x_N)\}$, and $\varpi^+ = \left(\varpi_1^+, \varpi_2^+, \dots, \varpi_q^+\right)^T$ with $\varpi_i^+ = \max\{\varpi_i, 0\}$.

In order to solve problem (34), a collective neurodynamic algorithm with multiple interconnected recurrent neural networks is designed in [125] on account of formulas (35). It reveals that the outputs of the collective neurodynamic networks are convergent to consensus at the optimal solution of problem (34). Different from the collective neurodynamic algorithm, [126–128] investigate multi-agent systems for distributed optimization problem (34). Based on the Lagrangian function $L(x, \lambda, v) = \sum_{i=1}^N f_i(x) + \sum_{i=1}^N \sum_{j=1}^{q_i} \lambda_{ij} g_{ij}(x) + \sum_{i=1}^N \sum_{j=1}^{p_i} v_{ij} h_{ij}(x)$ and saddle point theorem, a distributed primal-dual mirror descent algorithm is introduced in [126], and it shows that for any initial condition $\lambda_{ij}(0) \geqslant 0$, the algorithm converges to the optimal solution. By transforming the global inequality and equality constraints into convex constraint set to every node, [127] constructs a distributed asynchronous gossip algorithm with strictly convex and twice differential local objective functions, which forces the estimates to satisfy the KKT conditions in (35) with asymptotical convergence to the optimal solution. Furthermore, by relaxing the requested assumption from strictly convex to uniformly strictly convex, [128] introduces a continuous-time distributed zero-gradient-sum method based on a Lyapunov-like function contained the Bregman divergence with individual Lagrangian function. For connected undirected graph, the state $x(t)$ asymptotically converges to the optimal solution of problem (34). Moreover, instead of convex assumption for local cost functions and constraints, [129] presents an asynchronous distributed optimization algorithm based on a local augmented Lagrangian function for solving problem (34) with nonconvex local cost functions and constraints.

For the case of distributed optimization problem with separable global constraints, the problem (34) can be reformulated as

$$\begin{aligned} \min_{x} \quad & f(x) = \sum_{i=1}^N f_i(x_i), \\ \text{subject to} \quad & \textstyle\sum_{i=1}^N g_i(x_i) \le 0, \\ & \textstyle\sum_{i=1}^N h_i(x_i) = 0, \end{aligned} \tag{36}$$

where $x = \mathrm{col}(x_1, x_2, \dots, x_N) \in \Re^n, f_i : \Re^{n_i} \to \Re$ is the convex and differentiable objective function, $g_i : \Re^{n_i} \to \Re^q$ is the convex and Lipschitz continuous constraint function, equality constraint function $h_i : \Re^{n_i} \to \Re^p$ is an affine map and $h_i(x_i) = B_i x_i + b_i$ for $B_i \in \Re^{p \times n_i}, b_i \in \Re^p$. Due to the inequality or equality coupling each g_i or h_i and requiring full information of x, the constraints $g(x) \le 0$ and $h(x) = 0$ are termed as coupled constraints. To solve problem (36), under undirected and connected communication graph, [130] develops a dual subgradient algorithm with iterate-averaging feedback algorithm with fixed constant step sizes to acquire the exact optimal solution. It presents two types of algorithms for resolving problem (36), one is centralized method with a center node in network while another one is in a decentralized way. By defining $\theta_i(x_i) = \mathrm{col}(g_i(x_i), h_i(x_i))$, the coupled constraints can be represented as a compact vector $\theta(x) \triangleq \theta_1(x_1) + \theta_2(x_2) + \cdots + \theta_N(x_N)$, then the subgradient of dual function can be expressed on bias of $\theta(x)$. Based on a saddle point theorem, the proposed distributed algorithm converges to an optimal primal-dual solution.

7. Distributed optimization with general hybrid constraints

In this section, we consider the distributed optimization problem with hybrid constraints comprised of local bound constraints, globally coupled inequality and equality constraints.

For the case of distributed optimization problem with hybrid separable inequality and equality constraints, it is generally described as follows

$$\begin{aligned} \min_{x \in \Omega} \quad & f(x) = \sum_{i=1}^N f_i(x_i), \\ \text{subject to} \quad & g(x) = \sum_{i=1}^N g_i(x_i) \le 0, \\ & h(x) = \sum_{i=1}^N h_i(x_i) = 0, \end{aligned} \tag{37}$$

in which $x = \mathrm{col}(x_1, x_2, \dots, x_N)$ is a collection of decision variables, $\Omega = \Pi_{i=1}^N \Omega_i$ is the Cartesian product of local bound constrained sets, where $x_i \in \Omega_i, f_i : \Re^{n_i} \to \Re, g_i : \Re^{n_i} \to \Re^q$ and $h_i : \Re^{n_i} \to \Re^p$ for $i = 1, 2, \dots, N$.

According to the KKT conditions, x^* is an optimal solution to problem (37) if and only if there exists ω^* such that (x^*, ω^*) satisfies the following equation

$$\begin{cases} 0 = \nabla f(x^*) + \nabla\theta(x^*)\omega^* + \mathscr{N}_\Omega(x^*), \\ 0 = (\omega^*)^T\theta(x^*), \end{cases} \tag{38}$$

where $\theta_i(x_i) = \mathrm{col}(g_i(x_i), h_i(x_i))$. Then the coupled constraints $g(x)$ and $h(x)$ can be represented as $\theta(x) = \sum_{i=1}^N \theta_i(x_i)$. $\mathscr{N}_\Omega(x^*)$ is the normal cone to Ω at x^* defined as $\mathscr{N}_\Omega(x^*) = \{\vartheta \in \Re^n | \vartheta^T(y - x^*) \le 0, \forall y \in \Omega\}$. And $\omega = \mathrm{col}(\omega_1, \omega_2, \dots, \omega_N)$ with $\omega_i = \mathrm{col}(\omega_i^g, \omega_i^h)$.

Based on (38), to solve problem (37), a distributed algorithm with fixed step size is developed in [131], in which two update flows are considered: one is the intermediate iteration to calculate $\hat{x}_i(k)$ and $\hat{\omega}_i(k)$ by using the local information $\nabla f_i, \nabla\theta_i$ and neighbor' state ω_j; anther one is to calculate $x_i(k+1)$ and $\omega_i(k+1)$ on bias of the first intermediate iteration. Thus, the sequence $\{x_i(k)\}_{k=1}^\infty$ converges to the ith component of an optimal solution to problem (37). Different from the first-order method mentioned above, [132] proposes a distributed neurodynamic approach with neurodynamic recurrent neural networks in which all nodes can reach consensus based on the Lagrange multiplier method, and the decision vector x converges to the global optimal solution in a distributed manner for undirected and connected communication graph.

Next, we consider the case of distributed optimization problem with hybrid inseparable globally coupled inequality and equality constraints as follows

$$\begin{aligned} \min_{x} \quad & f(x) = \sum_{i=1}^{N} f_i(x), \\ \text{subject to} \quad & g_i(x) \le 0, \\ & h_i(x) = 0, \\ & x \in \bigcap_{i=1}^{N} \Omega_i, \end{aligned} \tag{39}$$

where $\Omega_i \subseteq \Re^n$ is a convex set and the other symbols are the same as those in (34).

Under the assumption of connected and undirected communication graph, problem (39) can be equivalently written as

$$\begin{aligned} \min_{\tilde{x}} \quad & f(\tilde{x}) = \sum_{i=1}^{N} f_i(x_i), \\ \text{subject to} \quad & g(\tilde{x}) \le 0, \\ & h(\tilde{x}) = 0, \\ & \mathbf{L}\tilde{x} = 0, \tilde{x} \in \Omega, \end{aligned} \tag{40}$$

where $\tilde{x} = \mathrm{col}(x_1, x_2, \ldots, x_N), g(\tilde{x}) = \mathrm{col}(g_1(x_1), g_2(x_2), \ldots, g_N(x_N))$, $h(\tilde{x}) = \mathbf{B}\tilde{x} - \mathbf{b} = 0$ with $\mathbf{B} = \mathrm{diag}\{B_1, B_2, \ldots, B_N\}$ and $\mathbf{b} = \mathrm{col}(b_1, b_2, \ldots, b_N), \Omega = \Pi_{i=1}^{N}\Omega_i$, and $\mathbf{L} = L \otimes I_n$. The equality constraint $\mathbf{L}\tilde{x} = 0$ guarantees that $x_1 = x_2 = \cdots = x_N$, then $\tilde{x}^* = \mathbf{1}_N \otimes x^*$ is an optimal solution to problem (40) if and only if $x^* \in \Re^n$ is an optimal solution to problem (39). Therefore, for the sake of analysis and calculation conveniently, it is generally to solve (40) instead of (39).

To solve the distributed optimization problem with hybrid constraints in (40), the Lagrange function is denoted as $L(\tilde{x}, \omega, v, \lambda) = f(\tilde{x}) + \omega^T g(\tilde{x}) + v^T(\mathbf{B}\tilde{x} - \mathbf{b}) + \lambda^T \mathbf{L}\tilde{x}$, where $\omega \in \Re^{qn}_{+}$, $v \in \Re^{pn}$ and $\lambda \in \Re^{Nn}$ are the Lagrange multipliers. Based on the saddle point theorem [112], $\tilde{x}^*$ is an optimal solution to problem (40) if and only if there exists a saddle point $(\tilde{x}^*, \omega^*, v^*, \lambda^*)$ such that the following inequality holds

$$L(\tilde{x}^*, \omega, v, \lambda) \leqslant L(\tilde{x}^*, \omega^*, v^*, \lambda^*) \leqslant L(\tilde{x}, \omega^*, v^*, \lambda^*). \tag{41}$$

On account of the saddle point theorem, [133] designs a continuous-time multi-agent system for solving problem (40), in which each agent optimizes its own private objective function by means of projection method for local bound constraint and the Lagrangian multiplier method for global coupled equality and inequality constraints. Through a proportional-integral consensus protocol in agents' interactions, it shows that the algorithm is convergent to the unique optimal solution to problem (40). By relaxing the local objective function to nonsmooth, [134] proposes a distributed projective primal-dual approach with fixed and connected communication topology for solving problem (39). According to the saddle point inequalities in (41), the designed algorithm can converge asymptotically to an equilibrium point which part of components is the optimal solution to primary problem (40) and the rest is optimal solution to the dual problem. For time-varying connected communication topologies, [135] presents a distributed penalty primal-dual subgradient method to ensure that all agents converge to the optimal solution. Recently, [136] proposes a novel continuous-time multi-agent neurodynamic algorithm for constrained optimization problem (39), where the objective function is the sum of nonsmooth convex local objective functions other than a continuously differentiable local objective function with local Lipschitz continuous gradient in [132]. It shows that the states of all agents achieve consensus asymptotically and converge to the optimal solution of problem (40) based on penalty function and regularization methods for undirected and connected communication graph.

8. Conclusions

This paper has reviewed models and algorithms for distributed optimization problems with different kinds of constraints in terms of the model structure, algorithm category, convergence analysis and communication topology. The reviewed distributed optimization problems include decomposable objective functions, decomposable or undecomposable variables. The models of distributed optimization problems include unconstrained optimization problem and constrained optimization problem. The constraints include bound constraints, separable and inseparable equality, inequality and hybrid constraints which contains local and global constraints. Moreover, this paper also reviews various of algorithms designed to solve the corresponding distributed optimization problems for searching the optimal solutions. However, in this paper, we only investigate convex distributed optimization problems. In fact, in many physical applications, such as machine learning and signal processing, the optimization problem modeled is generally nonconvex and the convex distributed algorithms are inapplicability. Another potential challenge is that some agents in networked cyber-physical system may loss of information or become adversarial due to malicious attacks or transmission failure, and it will result in agents failing to collaboratively search the optimal solution. Moreover, each agent requires to know its local information in advance while there is usually no prior knowledge since the information is highly unpredictable or privacy-preserving in many applications, such as the local objective functions of solar panels and wind generators, and encrypted data.

CRediT authorship contribution statement

Yanling Zheng: Conceptualization, Methodology, Writing – original draft. **Qingshan Liu:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Su, Traffic Engineering and Time-Varying Convex Optimization, Pennsylvania State University, PA, 2009.
[2] J. Chen, S. Kai, Cooperative transportation control of multiple mobile manipulators through distributed optimization, Sci. China (Inf. Sci.) 61 (12) (2018) 5–21.
[3] Z. Zhang, M. Chow, Convergence analysis of the incremental cost consensus algorithm under different communication network topologies in a smart grid, IEEE Trans. Power Syst. 27 (4) (2012) 1761–1768.
[4] T. Yang, J. Lu, D. Wu, J. Wu, G. Shi, Z. Meng, K.H. Johansson, A distributed algorithm for economic dispatch over time-varying directed networks with delays, IEEE Trans. Industr. Electron. 64 (6) (2017) 5095–5106.
[5] A. Nedic, J. Liu, Distributed optimization for control, Annu. Rev. Control Robot. Auton. Syst. 1 (2018) 77–103.
[6] K. Lynch, F. Park, Modern robotics: Mechanics, Planning, and Control, Cambridge University Press, New York, 2017.
[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2010) 1–122.
[8] N. Zeng, D. Song, H. Li, Y. You, F.E. Alsaadi, A competitive mechanism integrated multi-objective whale optimization algorithm with differential evolution, Neurocomputing 432 (2021) 170–182.
[9] N. Zeng, Z. Wang, W. Liu, H. Zhang, K. Hone, X. Liu, A dynamic neighborhood-based switching particle swarm optimization algorithm, IEEE Trans. Cybern. (2020) 1–12, https://doi.org/10.1109/TCYB.2020.3029748.
[10] W. Liu, Z. Wang, X. Liu, N. Zeng, D. Bell, A novel particle swarm optimization approach for patient clustering from emergency departments, IEEE Trans. Evol. Comput. 23 (4) (2019) 632–644.
[11] X. Ran, P. Shi, N. Angelia, A.K. Usmab, A general framework for decentralized optimization with first-order methods, Proc. IEEE 108 (11) (2020) 1869–1889.

[12] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, New York, 2004.
[13] A.H. Sayed, Adaptation, learning, and optimization over networks, Found. Trends Mach. Learn. 7 (4) (2014) 311–801.
[14] J. Chen, A.H. Sayed, Diffusion adaptation strategies for distributed optimization and learning over networks, IEEE Trans. Signal Process. 60 (8) (2012) 4289–4305.
[15] N. Anna, BookReview: Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, 1989.
[16] J.N. Tsitsiklis, D.P. Bertsekas, M. Athans, Distributed asynchronous deterministic and stochastic gradient optimization algorithms, IEEE Trans. Autom. Control 9 (31) (1986) 803–812.
[17] A. Nedic, A. Ozdaglar, P.A. Parrilo, Constrained consensus and optimization in multi-agent networks, IEEE Trans. Autom. Control 55 (4) (2010) 922–938.
[18] Q. Liu, T. Huang, J. Wang, One-layer continuous- and discrete-time projection neural networks for solving variational inequalities and related optimization problems, IEEE Trans. Neural Networks Learn. Syst. 25 (7) (2017) 1308–1318.
[19] A. Nedic, Convergence rate of distributed averaging dynamics and optimization in networks, Found. Trends Syst. Control 2 (1) (2015) 1–100.
[20] A. Nedic, A. Olshevsky, M.G. Rabbat, Network topology and communication-computation tradeoffs in decentralized optimization, Proc. IEEE 106 (5) (2017) 953–976.
[21] D. Jakovetic, D. Bajovic, J. Xavier, J.M.F. Moura, Primal-dual methods for large-scale and distributed convex optimization and data analytics, Proc. IEEE 108 (11) (2020) 1–16.
[22] A. Nedic, A. Ozdaglar, Distributed subgradient methods for multi-agent optimization, IEEE Trans. Autom. Control 54 (1) (2009) 48–61.
[23] I. Matei, J.S. Baras, Performance evaluation of the consensus-based distributed subgradient method under random communication topologies, IEEE J. Sel. Top. Signal Process. 5 (4) (2011) 754–771.
[24] D. Jakovetic, J. Xavier, J. Moura, Fast distributed gradient methods, IEEE Trans. Autom. Control 59 (5) (2014) 1131–1146.
[25] C. Xi, V.S. Mai, R. Xin, E.H. Abed, U.A. Khan, Linear convergence in optimization over directed graphs with row-stochastic matrices, IEEE Trans. Autom. Control 63 (10) (2018) 3558–3565.
[26] W. Shi, Q. Ling, G. Wu, W. Yin, Extra: An exact first-order algorithm for decentralized consensus optimization, Siam J. Optim. 25 (2) (2014) 944–966.
[27] B. Ning, Q. Han, Z. Zuo, Distributed optimization for multiagent systems: An edge-based fixed-time consensus approach, IEEE Trans. Cybern. 49 (1) (2019) 122–132.
[28] A. Nedic, A. Olshevsky, W. Shi, Achieving geometric convergence for distributed optimization over time-varying graphs, SIAM J. Optim. 27 (4) (2016) 1–29.
[29] Mokhtari A., Ribeiro A., DSA: Decentralized double stochastic averaging gradient algorithm, ArXiv:1506.04216v1 [math.OC].
[30] D. Jakovetic, A unification and generalization of exact distributed first-order methods, IEEE Trans. Signal Inf. Process. Over Networks 5 (1) (2018) 31–46.
[31] A. Nedic, A. Olshevsky, Distributed optimization over time-varying directed graphs, IEEE Trans. Autom. Control 60 (3) (2013) 601–615.
[32] S. Yang, Q. Liu, J. Wang, Distributed optimization based on a multiagent system in the presence of communication delays, IEEE Trans. Syst. Man Cybern. Syst. 47 (5) (2017) 717–728.
[33] Z. Qiu, S. Liu, L. Xie, Distributed constrained optimal consensus of multi-agent systems, Automatica 68 (2016) 209–215.
[34] Q. Liu, J. Wang, A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints, IEEE Trans. Neural Networks Learn. Syst. 24 (5) (2013) 812–824.
[35] A. Nedic, A. Ozdaglar, Distributed subgradient methods for multi-agent optimization, IEEE Trans. Autom. Control 54 (1) (2009) 48–61.
[36] K. Srivastava, A. Nedic, Distributed asynchronous constrained stochastic optimization, IEEE J. Sel. Top. Signal Process. 5 (4) (2011) 772–790.
[37] H. Fang, J. Lei, H. Chen, Primal-dual algorithm for distributed constrained optimization, Syst. Control Lett. 96 (2016) 110–117.
[38] M. Zhu, S. Martinez, On distributed convex optimization under inequality and equality constraints, IEEE Trans. Autom. Control 57 (1) (2011) 151–164.
[39] D. Yuan, D.W.C. Ho, S. Xu, Regularized primal-dual subgradient method for distributed constrained optimization, IEEE Trans. Cybern. 46 (9) (2016) 2109–2118.
[40] P.M. Pardalos, H.E. Romeijn, Handbook of Global Optimization, Kluwer Academic Publishers, London, 2002.
[41] R.L. Francis, L.F. McGinnis, J.A. White, Facility Layout and Location: an Analytical Approach, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1992.
[42] Q. Liu, M. Wang, A projection-based algorithm for optimal formation and optimal matching of multi-robot system, Nonlinear Dyn. 104 (2021) 439–450.
[43] P. Lin, W. Ren, H. Wang, U.M. Al-Saggaf, Multiagent rendezvous with shortest distance to convex regions with empty intersection: Algorithms and experiments, IEEE Trans. Cybern. 49 (3) (2019) 1026–1034.
[44] Chen A.I., Ozdaglar A., A fast distributed proximal-gradient method, in: Proceedings of 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2012, pp. 601–608.
[45] R. Olfati-Saber, R.M. Murray, Consensus problems in networks of agents with switching topology and time-delays, IEEE Trans. Autom. Control 49 (9) (2004) 1520–1533.
[46] Olshevsky A., Tsitsiklis J.N., Convergence speed in distributed consensus and averaging, arXiv:math/0612682v2 [math.OC].
[47] Wang J., Elia N., Control approach to distributed optimization, in: Proceedings of 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2010, pp. 557–561.
[48] B. Gharesifard, J. Cortes, Distributed continuous-time convex optimization on weight-balanced digraphs, IEEE Trans. Autom. Control 59 (3) (2012) 781–786.
[49] A. Solmaz, S. Kia, C. Jorge, M. Sonia, Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication, Automatica 55 (2015) 254–264.
[50] I. Lobel, A. Ozdaglar, Distributed subgradient methods for convex optimization over random networks, Autom. Control IEEE Trans. 56 (6) (2011) 1291–1306.
[51] S. Sundhar Ram, A. Nedi, V.V. Veeravalli, Distributed stochastic subgradient projection algorithms for convex optimization, J. Optim. Theory Appl. 147 (3) (2010) 516–545.
[52] K. Yuan, Q. Ling, W. Yin, On the convergence of decentralized gradient descent, Mathematics 26 (3) (2013) 953–976.
[53] K. Markku, R.H. Charles, Large-scale convex optimization via saddle point computation, Oper. Res. (1999) 93–101.
[54] A. Nedic, A. Ozdaglar, Subgradient methods for saddle-point problems, J. Optim. Theory Appl. 142 (1) (2009) 205–228.
[55] N. Maculan, S.C.S. Porto, C.C. Ribeiro, C.D.S. Cid, Parallel and distributed computation, RAIRO Oper. Res. 33 (1) (1999) 87–92.
[56] A. Mokhtari, W. Shi, Q. Ling, A. Ribeiro, Dqm: Decentralized quadratically approximated alternating direction method of multipliers, IEEE Trans. Signal Process. 64 (19) (2016) 5158–5173.
[57] C. Xi, V.S. Mai, R. Xin, E.H. Abed, U.A. Khan, Linear convergence in optimization over directed graphs with row-stochastic matrices, IEEE Trans. Autom. Control 63 (10) (2018) 3558–3565.
[58] F. Bnzit, V. Blondel, P. Thiran, J. Tsitsiklis, M. Vetterli, Weighted gossip: Distributed averaging using non-doubly stochastic matrices, in: Proceedings of IEEE International Symposium on Information Theory, 2010, pp. 1753–1757.
[59] C. Xi, U.A. Khan, Distributed subgradient projection algorithm over directed graphs, IEEE Trans. Autom. Control. 62 (8) (2017) 3986–3992.
[60] K.I. Tsianos, S. Lawlor, M.G. Rabbat, Push-sum distributed dual averaging for convex optimization, Proceedings of IEEE Conference on Decision & Control (2012) 5453–5458.
[61] V.S. Mai, E.H. Abed, Distributed optimization over weighted directed graphs using row stochastic matrix, Proceedings of American Control Conference (2016) 7165–7170.
[62] Q. Lu, X. Liao, H. Li, T. Huang, A nesterov-like gradient tracking algorithm for distributed optimization over directed networks, IEEE Trans. Syst. Man Cybern.: Syst. (2020) 1–13, https://doi.org/10.1109/TSMC.2019.2960770.
[63] S. Yang, J. Wang, Q. Liu, Consensus of heterogeneous nonlinear multi-agent systems with duplex control laws, IEEE Trans. Autom. Control 64 (12) (2019) 5140–5147.
[64] Z. Li, Z. Duan, G. Chen, L. Huang, Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint, Circuits Syst. I Regular Papers IEEE Trans. 57 (1) (2010) 213–224.
[65] W. Yu, G. Chen, M. Cao, Some necessary and sufficient conditions for second-order consensus in multi-agent dynamical systems, Automatica 46 (6) (2010) 1089–1095.
[66] Y. Xia, J. Wang, A recurrent neural network for solving nonlinear convex programs subject to linear constraints, IEEE Trans. Neural Networks 16 (2) (2005) 379–386.
[67] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[68] Zhou H., Zeng X., Hong Y., An exact penalty method for constrained distributed optimization, in: Proceedings of 36th Chinese Control Conference, Dalian, China, 2017.
[69] S. Liang, L. Wang, G. Yin, Dual averaging push for distributed convex optimization over time-varying directed graph, IEEE Trans. Autom. Control 65 (4) (2020) 1785–1791.
[70] Q. Liu, J. Wang, A second-order multi-agent network for bound-constrained distributed optimization, IEEE Trans. Autom. Control 60 (12) (2015) 3310–3315.
[71] C. Corduneanu, Stability and Control: Theory, Methods and Applications, Taylor Francis, 2002.
[72] S. Yang, J. Wang, Q. Liu, Cooperativecompetitive multiagent systems for distributed minimax optimization subject to bounded constraints, IEEE Trans. Autom. Control 64 (4) (2018) 1358–1372.
[73] Z. Yan, J. Wang, G. Li, A collective neurodynamic optimization approach to bound-constrained nonconvex optimization, Neural Networks 55 (2014) 20–29.
[74] H. Che, J. Wang, A collaborative neurodynamic approach to global and combinatorial optimization, Neural networks 114 (2019) 15–27.
[75] Z. Yan, J. Fan, J. Wang, A collective neurodynamic approach to constrained global optimization, IEEE Trans. Neural Networks Learn. Syst. 28 (5) (2017) 1206–1215.
[76] Q. Liu, S. Yang, J. Wang, A collective neurodynamic approach to distributed constrained optimization, IEEE Trans. Neural Networks Learn. Syst. 28 (8) (2017) 1747–1758.
[77] J. Lei, H. Chen, H. Fang, Primal-dual algorithm for distributed constrained optimization, Syst. Control Lett. 96 (2015) 110–117.
[78] F. Guo, C. Wen, J. Mao, G. Li, Y.D. Song, A distributed hierarchical algorithm for multi-cluster constrained optimization, Automatica 77 (2017) 230–238.

[79] X. Zeng, P. Yi, Y. Hong, Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach, IEEE Trans. Autom. Control 62 (2015) 5227–5233.
[80] Y. Lou, G. Shi, K. Johansson, Y. Hong, Approximate projected consensus for convex intersection computation: Convergence analysis and critical error angle, IEEE Trans. Autom. Control 59 (7) (2014) 1722–1736.
[81] Y. Lou, Y. Hong, S. Wang, Distributed continuous-time approximate projection protocols for shortest distance optimization problems, Automatica 69 (2016) 1722–1736.
[82] Q. Liu, S. Yang, Y. Hong, Constrained consensus algorithms with fixed step size for distributed convex optimization over multiagent networks, IEEE Trans. Autom. Control 62 (8) (2017) 4259–4265.
[83] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming. Theory and Algorithms, third ed., Wiley, Hoboken, NJ, USA, 2006.
[84] B.D.O. Anderson, S. Mou, A.S. Morse, U. Helmke, Decentralized gradient algorithm for solution of a linear equation, Numer. Algebra Control Optim. 6 (3) (2015) 319–328.
[85] S. Mou, L. Lin, D. Wang, A. Fullmer, S. Morse, A distributed algorithm for efficiently solving linear equations and its applications, special issue jcw, Syst. Control Lett. 99 (2016) 21–27.
[86] X. Wang, S. Mou, D. Sun, Improvement of a distributed algorithm for solving linear equations, IEEE Trans. Ind. Electron. 64 (4) (2017) 3113–3117.
[87] G. Shi, B.D.O. Anderson, U. Helmke, Network flows that solve linear equations, IEEE Trans. Autom. Control 62 (6) (2017) 2659–2674.
[88] X. Wang, J. Zhou, S. Mou, M. Corless, A distributed algorithm for least squares solutions, IEEE Trans. Autom. Control 64 (10) (2019) 4217–4222.
[89] X. Jiang, S. Qin, X. Xue, A penalty-like neurodynamic approach to constrained nonsmooth distributed convex optimization, Neurocomputing 377 (2020) 225–233.
[90] Y. Xu, T. Han, K. Cai, Z. Lin, G. Yan, M. Fu, A distributed algorithm for resource allocation over dynamic digraphs, IEEE Trans. Signal Process. 65 (10) (2017) 2600–2612.
[91] D. Wang, Y. Chen, V. Gupta, J. Lian, Distributed constrained optimization for multi-agent systems over a directed graph with piecewise stepsize, J. Franklin Inst. 357 (8) (2020) 4855–4868.
[92] K. Li, Q. Liu, Z. Zeng, Distributed optimisation based on multi-agent system for resource allocation with communication time-delay, IET Control Theory Appl. 14 (2) (2019) 549–557.
[93] K. Li, Q. Liu, S. Yang, J. Cao, G. Lu, Cooperative optimization of dual multiagent system for optimal resource allocation, IEEE Trans. Syst. Man Cybern.: Syst. 50 (11) (2020) 4676–4687.
[94] X. Le, S. Chen, F. Li, Z. Yan, J. Xi, Distributed neurodynamic optimization for energy internet management, IEEE Trans. Syst. Man Cybern.: Syst. 49 (8) (2019) 1624–1633.
[95] Q. Liu, X. Le, K. Li, A distributed optimization algorithm based on multiagent network for economic dispatch with region partitioning, IEEE Trans. Cybern. (2019) 1–10, https://doi.org/10.1109/TCYB.2019.2948424.
[96] T.H. Chang, M. Hong, X. Wang, Multi-agent distributed optimization via inexact consensus admm, IEEE Trans. Signal Process. 63 (2) (2014) 482–497.
[97] T. Lin, S. Ma, S. Zhang, On the global linear convergence of the admm with multi-block variables, Siam J. Optim. 25 (3) (2015) 1–23.
[98] X. Le, S. Chen, Z. Yan, J. Xi, A neurodynamic approach to distributed optimization with globally coupled constraints, IEEE Trans. Cybern. 48 (11) (2018) 3149–3158.
[99] C. Gu, Z. Wu, J. Li, Regularized dual gradient distributed method for constrained convex optimization over unbalanced directed graphs, Numer. Algorithms 84 (1) (2020) 91–115.
[100] R.T. Rockafellar, Augmented lagrangians and applications of the proximal point algorithm in convex programming, Math. Operat. Res. 1 (2) (1976) 97–116.
[101] D.P. Bertsekas, Extended monotropic programming and duality, J. Optim. Theory Appl. 139 (2008) 209–225.
[102] C. Nikolaos, D. Darinka, M.Z. Michael, An augmented lagrangian method for distributed optimization, Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 152 (2014) 405–434.
[103] R. Shang, K. Dai, L. Jiao, R. Stolkin, Improved memetic algorithm based on route distance grouping for multiobjective large scale capacitated arc routing problems, IEEE Trans. Cybern. 46 (4) (2017) 1000–1013.
[104] J. Wang, R. Dou, R.R. Muddada, W. Zhang, Management of a holistic supply chain network for proactive resilience: Theory and case study, Comput. Ind. Eng. 125 (12) (2017) 668–677.
[105] D.P. Palomar, M. Chiang, Alternative distributed algorithms for network utility maximization: Framework and applications, IEEE Trans. Autom. Control 52 (12) (2007) 2254–2269.
[106] B. Huang, L. Liu, H. Zhang, Y. Li, Q. Sun, Distributed optimal economic dispatch for microgrids considering communication delays, IEEE Trans. Syst. Man Cybern.: Syst. 49 (8) (2019) 1634–1642.
[107] J. Qin, Q. Ma, Y. Shi, L. Wang, Recent advances in consensus of multi-agent systems: A brief survey, IEEE Trans. Industr. Electron. 64 (6) (2017) 4972–4983.
[108] Z. Tang, D.J. Hill, T. Liu, A novel consensus-based economic dispatch for microgrids, IEEE Trans. Smart Grid 9 (4) (2018) 3920–3922.
[109] C. Li, X. Yu, T. Huang, X. He, Distributed optimal consensus over resource allocation network and its application to dynamical economic dispatch, IEEE Trans. Neural Networks Learn. Syst. 29 (6) (2018) 2407–2418.
[110] W. Du, L. Yao, D. Wu, X. Li, G. Liu, T. Yang, Proceedings of accelerated distributed energy management for microgrids, IEEE Power Energy Society General Meeting (PESGM) 2018 (2018) 1–5.
[111] S. Yang, S. Tan, J. Xu, Consensus based approach for economic dispatch problem in a smart grid, IEEE Trans. Power Syst. 28 (4) (2013) 4416–4426.
[112] D. Bertsekas, Convex Optimization Theory, Athena Scientific, Belmont, MA, 2009.
[113] M. Zhu, S. Martinez, On distributed convex optimization under inequality and equality constraints, IEEE Trans. Autom. Control 57 (1) (2012) 151–164.
[114] W. Jia, S. Qin, X. Xue, A generalized neural network for distributed nonsmooth optimization with inequality constraint, Neural Networks 119 (2019) 46–56.
[115] Li X., Yi X., Xie L., Distributed online optimization for multi-agent networks with coupled inequality constraints, ArXiv:1805.05573v3 [math.OC].
[116] T. Chang, A. Nedic, A. Scaglione, Distributed constrained optimization by consensus-based primal-dual perturbation method, IEEE Trans. Autom. Control 59 (6) (2014) 1524–1538.
[117] P. Xie, K. You, S. Song, C. Wu, Distributed random-fixed projected algorithm for constrained optimization over digraphs, IFAC Papersonline 50 (1) (2017) 14436–14441.
[118] D. Yuan, S. Xu, H. Zhao, Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms, IEEE Trans. Syst. Man Cybern. Part B Cybernetics Society 41 (6) (2011) 1715–1724.
[119] B. Zhou, X. Liao, T. Huang, H. Wang, G. Chen, Distributed multi-agent optimization with inequality constraints and random projections, Neurocomputing 197 (2016) 195–204.
[120] D. Yuan, S. Xu, B. Zhang, L. Rong, Distributed primal–dual stochastic subgradient algorithms for multi-agent optimization under inequality constraints, Int. J. Robust Nonlinear Control 23 (16) (2013) 2434–2456.
[121] S. Liang, X. Zeng, Y. Hong, Distributed nonsmooth optimization with coupled inequality constraints via modified lagrangian function, IEEE Trans. Autom. Control 63 (6) (2017) 1753–1759.
[122] X. Li, L. Xie, Y. Hong, Distributed continuous-time nonsmooth convex optimization with coupled inequality constraints, IEEE Trans. Control Network Syst. 7 (1) (2020) 74–84.
[123] A. Olama, N. Bastianello, P.R.C. Mendes, E. Camponogara, Relaxed hybrid consensus admm for distributed convex optimisation with coupling constraints, IET Control Theory Appl. 13 (17) (2019) 2828–2837.
[124] I. Notarnicola, G. Notarstefano, Constraint coupled distributed optimization: a relaxation and duality approach, IEEE Trans. Control Network Syst. 7 (1) (2020) 483–492.
[125] Q. Liu, S. Yang, J. Wang, A collective neurodynamic approach to distributed constrained optimization, IEEE Trans. Neural Networks Learn. Syst. 28 (8) (2017) 1747–1758.
[126] R. Sheng, W. Ni, Distributed constrained optimization via continuous-time mirror design, Adv. Differ. Equ. 376 (2018).
[127] Nikookhoy S., Lu J., Tang C.Y., Distributed convex optimization with identical constraints, in: Proceedings of 50th IEEE Decision and Control and European Control Conference, Orlando, FL, USA, 2011, pp. 2926–2931.
[128] Doan T.T., Tang C.Y., Continuous-time constrained distributed convex optimization, in: Proceedings of 2012 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton, 2012.
[129] Farina F., Garulli A., Giannitrapani A., Notarstefano G., Asynchronous distributed method of multipliers for constrained nonconvex optimization, in: Proceedings of European Control Conference, Limassol, Cyprus, 2018.
[130] S. Liang, Y. Le, G. Yin, Distributed dual subgradient algorithms with iterate-averaging feedback for convex optimization with coupled constraints, IEEE Trans. Cybern. 51 (5) (2021) 2529–2539.
[131] S. Liang, L. Wang, G. Yin, Distributed smooth convex optimization with coupled constraints, IEEE Trans. Autom. Control 65 (1) (2020) 347–353.
[132] X. Le, S. Chen, Z. Yan, J. Xi, A neurodynamic approach to distributed optimization with globally coupled constraints, IEEE Trans. Cybern. 48 (11) (2018) 3149–3158.
[133] S. Yang, Q. Liu, J. Wang, A multi-agent system with a proportional-integral protocol for distributed constrained optimization, IEEE Trans. Autom. Control 62 (7) (2017) 3461–3466.
[134] Y. Zhu, W. Yu, G. Wen, G. Chen, Projected primal-dual dynamics for distributed constrained nonsmooth convex optimization, IEEE Trans. Cybern. 50 (4) (2020) 1776–1782.
[135] Zhu M., Martinez S., On distributed optimization under inequality and equality constraints via penalty primal-dual methods, in: Proceedings of American Control Conference, Marriott Waterfront, Baltimore, MD, USA, 2010, pp. 2434–2439.
[136] L. Ma, W. Bian, A novel multiagent neurodynamic approach to constrained distributed convex optimization, IEEE Trans. Cybern. 51 (3) (2021) 1322–1333.

Yanling Zheng received the B.S. degree in mathematics and applied mathematics from Huanggang Normal University, Huanggang, China, in 2015, and the M.S. degree in mathematics from Zhejiang Normal University, Zhejiang, China, in 2019. She is currently pursuing the Ph.D. degree in mathematics with Southeast University, Nanjing, China. Her current research interests include optimization theory and applications and multi-agent systems.

Qingshan Liu received the B.S. degree in mathematics from Anhui Normal University, Wuhu, China, in 2001, the M.S. degree in applied mathematics from Southeast University, Nanjing, China, in 2005, and the Ph.D. degree in automation and computer-aided engineering, Chinese University of Hong Kong, Hong Kong, in 2008. He is currently a Professor with the School of Mathematics, Southeast University, Nanjing, China. His current research interests include optimization theory and applications, artificial neural networks, computational intelligence, and multi-agent systems.

Dr. Liu serves as an Associate Editor for the IEEE Transactions on Cybernetics and IEEE Transactions on Neural Networks and Learning Systems, and is an Editorial Board Member of Neural Networks and Neural Processing Letters.