SciencePG
Science Publishing Group

Distributed Subgradient Algorithm for Multi-Agent Convex Optimization with Global Inequality and Equality Constraints

Li Xiao[*], Junjie Bao, Xi Shi

Department of Mathematics and Information Engineering, Chongqing University of Education, Chongqing, PR China

Email address:

xsxiaoli@163.com (Li Xiao)

[*]Corresponding author

Abstract: In this paper, we present an improved subgradient algorithm for solving a general multi-agent convex optimization problem in a distributed way, where the agents are to jointly minimize a global objective function subject to a global inequality constraint, a global equality constraint and a global constraint set. The global objective function is a combination of local agent objective functions and the global constraint set is the intersection of each agent local constraint set. Our motivation comes from networking applications where dual and primal-dual subgradient methods have attracted much attention in the design of decentralized network protocols. Our main focus is on constrained problems where the local constraint sets are identical. Thus, we propose a distributed primal-dual subgradient algorithm, which is based on the description of the primal-dual optimal solutions as the saddle points of the penalty functions. We show that, the algorithm can be implemented over networks with changing topologies but satisfying a standard connectivity property, and allow the agents to asymptotically converge to optimal solution with optimal value of the optimization problem under the Slater's condition.

Keywords: Consensus, Saddle Point, Distributed Optimization, Subgradient Algorithm

1. Introduction

In recent years, distributed optimization and control have developed rapidly, and have been welcomed in the fields of industry and national defense, including smart grid, sensor network, social network and information system (Cyber-Physical system). Distributed optimization problems of multi-agent systems appear different kinds of distributed processing issues such as distributed estimation, distributed motion planning, distributed resource allocation and distributed congestion control [1-12]. The main focus is to solve a distributed optimization problem where the global objective function is composed of a sum of local objective functions, each of which is only known by one agent. Distributed optimization problems were first studied systematically in [1] where the union of the graphs was assumed to be strongly connected among each time interval of a certain bounded length and the adjacency matrices were doubly stochastic. A distributed subgradient method was introduced to solve the distributed optimization and error

bounds on the performance index functions were given. As a continuation of [1], a distributed subgradient projection algorithm was developed in [2] for distributed optimization where each agent was constrained to remain in a closed convex set and the paper gave corresponding convergence analysis on identical closed convex sets and on fully connected graphs with non-identical closed convex sets. Inspired by the works of [1, 2], the algorithms proposed in [1] and [2] were studied in the random environment [3] and [4], where the agents had the same state constraint. In [5], the communication topology was undirected and each possible communication link was functioning with a given probability. Thus, the expected communication topology is essentially fixed and undirected. Different from [1]-[5], a dual averaging subgradient algorithm was developed and analyzed for randomized graphs under the assumption that all agents remain in the same closed convex set in [6] and it was shown that the number of iterations were required by their algorithm scales inversely in the spectral gap of the network. Moreover, distributed optimization problems with asynchronous step-

sizes or inequality-equality constraints or using other algorithms were studied in [7]-[12] and corresponding conditions were given to ensure the system converge to the optimal point or its neighborhood. However, as in [1]-[5], it was assumed in [6]-[12] that the state sets of agents to be identical or the objective function finally converge to only a neighborhood of the optimal set.

In this paper our work is to extend [14] to study the penalty primal-dual subgradient projection algorithm in a more general method. In [14], the authors solved a multi-agent convex optimization problem where the agents subject to a global inequality constraint, a global equality constrain and a global constraint set. In order to solved these constraints, the author in [14] presented two different distributed projection algorithms with three assumptions that the union of the graphs is assumed to be strongly connected among each time interval of a certain bounded length and the adjacency matrices were doubly stochastic and non-degeneracy. However, [14] guaranteed the edge weight matrices of graphs were doubly stochastic (i.e., $\sum_{j=1}^{N} a_{ij}(k)=1$ for all $i \in V$ and $k \geq 0$, and $\sum_{i=1}^{N} a_{ij}(k)=1$ for all $j \in V$ and $k \geq 0$). Previous work did not perform well on the application of the distributed algorithms in multi-agent network.

Contributions: The subgradient algorithm (we proposed) is different with the approach proposed in [14] in properties and analysis. In our approach, the communication topology is without loss of generality. This paper does not recur to the assumption that the adjacency matrices are doubly stochastic, and we only require the network is weight-balanced, which makes our algorithm more practical. In this paper, we consider a general multi-agent optimization problem where the main focus is to minimize a global objective function which is a sum of local objective functions, subject to global constraints, including an inequality constraint, an equality constraint and a (state) constraint set. Each local objective function is convex and only known by one particular agent. On the other hand, the inequality (resp. equality) constraint is given by a convex (resp. affine) function and known by all agents. Each node has its own convex constraint set, and the global constraint set is defined as their intersection. Particularly, we assume that the local constraint sets are identical. Our main interest is in computing approximate saddle points of the Lagrangian function of a convex constrained optimization problem. To set the stage, we first study the computation of approximate saddle points (as opposed to asymptotically exact solutions) by using the subgradient method with a constant step-size. We consider constant step-size rule because of its simplicity and practical relevance, and because our interest is in generating approximate solutions in finite number of iterations.

The paper is organized as follows. In Section II, we give some basic preliminaries and concepts. Then, in Section III, we present our problem formulation as well as distributed consensus algorithm preliminaries. We then introduce the distributed penalty primal-dual subgradient algorithm with some supporting lemmas and continue with a convergence analysis of the algorithm in Section IV. Furthermore, the properties of the algorithm are explored by employing a numerical example in Section V. Finally, we conclude the paper with a discussion in Section VI.

2. Preliminaries and Notations

In this section, we first introduce some preliminary results about graph theory, the properties of the projection operation on a closed convex set and convex analysis (referring to [13], [14]).

A. Algebraic Graph Theory

The communication among different nodes in an information interplay network can be modeled as a weighted directed graph $G = \{V, E, A\}$, where $V = \{1, 2, ..., N\}$ is the set of nodes with i representing the i th node, $E \subseteq V \times V$ is the edge set, and $A = (a_{ij})_{N \times N}$ is the weighted adjacency matrix of G with nonnegative adjacency elements a_{ij} and zero diagonal elements. A directed edge $e_{ji} = (v_j, v_i)$ implies that node j can reach node i or node i can receive information from node j. If an edge $(j, i) \in E$, then node j is called a neighbor of node i and $a_{ij} > 0$. The neighbor node set of node i is denoted by N_i, while we define $|N_i|$ as the number of neighbors of node i. The Laplacian matrix $L = (l_{ij})_{N \times N}$ associated with the adjacency matrix A is defined by $l_{ij} = -a_{ij}, i \neq j$; $l_{ii} = \sum_{j=1, j \neq i}^{N} a_{ij}$, which ensures that $\sum_{j=1}^{N} l_{ij} = 0$. The Laplacian matrix L has a zero eigenvalue, and the corresponding eigenvector is 1_N. Note that the Laplacian matrix L of a directed graph G is asymmetric. The in-degree and out-degree of node i can be respectively defined by the Laplacian matrix as : $d_{in}(v_i) = -\sum_{j=1, j \neq i}^{N} l_{ij} = l_{ii}$ and $d_{out}(v_i) = -\sum_{j=1, j \neq i}^{N} l_{ji}$. A directed path from node j to node i is a sequence of edges $(j, i_1), (i_1, i_2), ..., (i_m, i)$ in the directed graph G with distinct nodes $i_k, k = 1, 2, ..., m$. A directed graph is strongly connected if for any two distinct nodes j and i in the set V, there always exists a directed path from node j to node i. A graph is called an in-degrees (or out-degrees) balanced graph if the in-degrees (or out-degrees) of all nodes in the directed graph are equal. A directed graph with N nodes is called a directed tree if it contains $N-1$ edges and there exists a root node with directed paths to every other node. A directed spanning tree of a directed graph is a directed tree that contains all the network nodes.

B. Basic Notations and Concepts

The following notion of saddle point plays a critical role in our paper.

Definition 1 (Saddle point): Consider a convex-concave function $L : X \times M \times V \to R$, where X, M and V are closed convex subsets in $R^{\bar{n}}$ and $X \times M \times V \to R^{\bar{m}}$. We are

interested in computing a saddle point (x^*, μ^*, λ^*) of $H(x^*, \mu^*, \lambda^*)$ over the set $X \times M \times V$, where a saddle point is defined as a vector pair (x^*, μ^*, λ^*) that satisfies

$$H(x^*, \mu, \lambda) \leq H(x^*, \mu^*, \lambda^*) \leq H(x, \mu^*, \lambda^*), \text{ for all }$$
$$x \in X, \mu \in M, \lambda \in V$$

In this paper, we do not assume the function $f^{[i]}$ at some points are not differentiable, and the subgradient plays the role of the gradient.

Definition 2 : For a given convex function $F : R^{\bar{n}} \to R$ and a point $\hat{x} \in R^{\bar{n}}$, a subgradient of the function F at $\hat{x}$ is a vector $DF(\tilde{x}) \in R^{\bar{n}}$ such that the following subgradient inequality holds for any $x \in R^{\bar{n}}$:

$$DF(\hat{x})^{\mathrm{T}}(x - \hat{x}) \leq F(x) - F(\hat{x})$$

Similarly, for a given concave function $G : R^{\bar{m}} \to R$ and a point $\breve{\mu} \in R^{\bar{m}}$, a supgradient of the function G at $\bar{\mu}$ is a vector $\hat{D}G(\bar{\mu}) \in R^{\bar{m}}$ such that the following supgradient inequality holds for any $\bar{\mu} \in R^{\bar{m}}$:

$$\hat{D}G(\bar{\mu})^{\mathrm{T}}(\mu - \bar{\mu}) \geq G(\mu) - G(\bar{\mu})$$

We use $P_X[\bar{x}]$ to denote the projection of a vector $\bar{x}$ on a closed convex set X, i.e.

$$P_X[\bar{x}] = \arg\min_{x \in X} \| \bar{x} - x \|$$

In the subsequent development, the properties of the projection operation on a closed convex set play an important role. In particular, we use the projection inequality, i.e., for any vector x

$$(P_X[x] - x)^{\mathrm{T}}(y - P_X[x]) \geq 0 \text{ for all } y \in X \qquad (1)$$

We also use the standard non-expansiveness property, i.e.

$$\| P_X[x] - P_X[y] \| \leq \| x - y \| \text{ for any } x \text{ and } y \qquad (2)$$

In addition, we use the properties given in the following lemma.

Lemma 2.1: Let X be a nonempty closed convex set in R^n. Then, we have for any $x \in R^n$,

(a) $(P_X[x] - x)^{\mathrm{T}}(y - P_X[x]) \leq - \| P_X[x] - x \|^2$, for all $y \in X$.

(b) $\| P_X[x] - y \|^2 \leq \| x - y \|^2 - \| P_X[x] - x \|^2$, for all $y \in X$.

Proof:

(a) Let $x \in R^n$ be arbitrary. Then, for any $y \in X$, we have

$$(P_X[x] - x)^{\mathrm{T}}(x - y) = (P_X[x] - x)^{\mathrm{T}}(x - P_X[x]) + (P_X[x] - x)^{\mathrm{T}}(P_X[x] - y)$$

By the projection inequality [cf. (1)], it follows that

$$(P_X[x] - x)^{\mathrm{T}}(P_X[x] - y) \leq 0$$

implying

$$(P_X[x] - x)^{\mathrm{T}}(y - P_X[x]) \leq - \| P_X[x] - x \|^2, \text{ for all } y \in X$$

(b) For an arbitrary $x \in R^n$ and for all $y \in X$, we have

$$\| P_X[x] - y \|^2 = \| P_X[x] - x + x - y \|^2 = \| P_X[x] - x \|^2 + \| x - y \|^2 + 2(P_X[x] - x)^{\mathrm{T}}(x - y)$$

By using the inequality of part (a), we obtain

$$\| P_X[x] - y \|^2 \leq \| x - y \|^2 - \| P_X[x] - x \|^2, \text{ for all } y \in X$$

Part (b) of the preceding lemma establishes a relation between the projection error vector and the feasible directions of the convex set X at the projection vector.

The following notations besides those aforementioned will be used throughout this paper. R^n denotes the set of all n-dimensional real vector spaces. Given a set S, we denote $co(S)$ by its convex hull. We write x^{T} or A^{T} to denote the transpose of a vector x or a matrix A. We let the function $[\cdot]^+ : R^{\bar{m}} \to R^{\bar{m}}_{\geq 0}$ denote the projection operator onto the non-negative orthant in $R^{\bar{m}}$. Denote $1_m = (1, ..., 1)^{\mathrm{T}} \in R^m$ and $0_m = (0, ..., 0)^{\mathrm{T}} \in R^m$. For a vector $x \in R^{\bar{n}}$, we denote $|x| = (|x_1|, ..., |x_{\bar{n}}|)^{\mathrm{T}}$, while $\| x \|$ is the standard Euclidean norm in the Euclidean space. In this paper, the quantities (e.g., functions, scalars and sets) associated with agent i will be indexed by the superscript $[i]$.

3. Problem Statement

We consider a multi-agent network model. The nodes connectively at time $k \geq 0$ can be represented by a directed weighted graph $G(k) = (V, E(k), A(k))$, where $E(k)$ is the set of activated edges at time k, i.e., edge $(j, i) \in E(k)$ if and only if node i can receive data from node j, and $A(k) = [a_{ij}(k)] \in R^{N \times N}$ is the adjacency matrix, in which $a_{ij}(k) \geq 0$ is the weight assigned to the edge (j, i) at time k. Please note that the set $E(k) \subset V \times V \setminus diag(V)$ is the set of edges with non-zero weights $a_{ij}(k)$. In this paper the agents are to correspondingly solve the following optimization problem:

$$\min_{x \in R^n} f(x) = \sum_{i=1}^{N} f^{[i]}(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in X \quad (3)$$

where $f^{[i]} : R^n \to R$ is a convex objective function of agent i, and X is a nonempty, closed, compact and convex subset of R^n. In particular, we study the cases where the local constraint sets are identical i.e., $X = X^{[i]}$ for each agent, and x is a global decision vector. Assume that $f^{[i]}$ is only known by agent i. The function $g : R^n \to R$ is known by all the agents with each component g_ℓ, for $\ell \in \{1, ..., m\}$,

being convex. The inequality $g(x) \leq 0$ is component-wise; i.e., $g_\ell(x) \leq 0$, for all $\ell \in \{1,...,m\}$, and represents a global inequality constraint. The function $h : R^n \to R$, represents a global equality constraint, and is known by all the agents. Let f^* denote the optimal value of (3) and x^* denote an optimal solution of (3). We assume that the optimal value f^* to be finite. We also represent the optimal solution set by X^*, i.e., $X^* = \{x \in R^n \mid \sum_{i=1}^n f^{[i]}(x) = f^*\}$. We will assume that in general f is non-differentiable.

To generate optional solutions to the primal problem of Eq. (3), we consider optional solutions to its dual problem. Here, the dual problem is the one arising from penalty relaxation of the inequality constraints $g(x) \leq 0$ and equality constraints $h(x) = 0$. Note that the primal problem (3) is trivially equivalent to the following:

$$\min_{x \in R^n} f(x), \quad \text{s.t.} \quad Ng(x) \leq 0, \ Nh(x) = 0, \ x \in X$$

with associated dual problem given by

$$\max_{\mu \in R^m, \lambda \in R^v} q_P(\mu, \lambda), \quad \text{s.t.} \quad \mu \geq 0, \ \lambda \geq 0$$

Here $q_P : R_{\geq 0}^m \times R_{\geq 0}^v \to R$ is the penalty dual function defined by $q_P(\mu, \lambda) = \inf_{x \in X} H(x, \mu, \lambda)$, where $H : R^n \times R_{\geq 0}^m \times R_{\geq 0}^v \to R$ is the penalty function given by $H(x, \mu, \lambda) = f(x) + N\mu^T[g(x)]^+ + N\lambda^T \mid h(x) \mid$. We often refer to vector $\mu \in R^m, \lambda \in R^v$ with $\mu \geq 0, \lambda \geq 0$ as two multiplier. We denote the dual optimal value by q^* and the dual optimal set by M^*. We define the penalty function

$H^{[i]}(x, \mu, \lambda) : R^n \times R_{\geq 0}^m \times R_{\geq 0}^v \to R$ for each agent i as follows: $H^{[i]}(x, \mu, \lambda) = f^{[i]}(x) + \mu^T[g(x)]^+ + \lambda^T \mid h(x) \mid$. In this way, we have that $H(x, \mu, \lambda) = \sum_{i=1}^N H^{[i]}(x, \mu, \lambda)$. We say that there is zero duality gap if the optimal value of the primal and the dual problems are equal, i.e., $f^* = q^*$. As proven in the following lemma, the Slater's condition in Assumption 3.1 ensures zero duality and the existence of penalty dual optimal solutions.

Assumption 3.1 (Slater's Condition): There exists a vector $\overline{x}$ such that $g(\overline{x}) < 0$ and $h(\overline{x}) = 0$. And there exists at least one interior $\overline{x}$ of X, i.e. $\overline{x} \in X$, problem (3) has finite optimal solution, and $X = \bigcap_{i=1}^N X^{[i]}$ has nonempty interior point.

Lemma 3.1: Let the Slater condition holds, the values of f^* and q^* coincide, and M^* is non-empty.

Proof: Define Lagrangian function $L_a : R^n \times R_{\geq 0}^m \times R_{\geq 0}^v \to R$ as $L_a(x, \mu, \lambda) = f(x) + N\mu^T g(x) + N\lambda^T h(x)$, with the associated dual problem defined by

$$\max_{\mu \in R^m, \lambda \in R^v} q_a(\mu, \lambda), \quad \text{s.t.} \quad \mu \geq 0 \tag{4}$$

Here, the dual function, $q_a(\mu, \lambda) = \inf_{x \in X} L_a(x, \mu, \lambda)$. The dual optimal value of problem (7) is denote by a^* and the set of dual optimal solutions is denoted by Q^*. Since X is convex, f and g_ℓ, for $\ell \in \{1,...,m\}$, are convex, and f^* is finite and the Slater's condition holds, we can conclude that $f^* = a^*$ and $Q^* \neq \varnothing$. We now proceed to characterize q^* and M^*. Pick any $q_a(\mu^*, \lambda^*) \in Q^*$. Since $\mu^* \geq 0$, then

$$\begin{aligned} a^* = q_a(\mu^*, \lambda^*) &= \inf_{x \in X}\{f(x) + N(\mu^*)^T g(x) + N(\lambda^*)^T h(x)\} \\ &\leq \inf_{x \in X}\{f(x) + N(\mu^*)^T[g(x)]^+ + N \mid \lambda^* \mid^T \mid h(x) \mid\} = q_P(\mu^*, \mid \lambda^* \mid) \leq q^* \end{aligned} \tag{5}$$

On the other hand, pick any $x^* \in X^*$. Then x^* is feasible, i.e., $x^* \in X$ $[g(x^*)]^+ = 0$ and $\mid h(x^*) \mid = 0$. It implies that $q(\mu, \lambda) \leq H(x^*, \mu, \lambda) = f(x^*) = f^*$ holds for any $\mu \in R_{\geq 0}^m$ and $\lambda \in R_{\geq 0}^v$, and thus $q^* = \sup_{\mu \in R_{\geq 0}^m, \lambda \in R_{\geq 0}^v} q(\mu, \lambda) \leq f^* = a^*$. Therefore, we have $f^* = q^*$.

To prove the non-empty of M^*, we pick any $(\mu^*, \lambda^*) \in Q^*$. From (5) and $a^* = q^*$, we can see that $(\mu^*, \mid \lambda^* \mid) \in M^*$ and thus $M^* \neq \varnothing$.

Throughout this paper, we use the following assumption for problem (3).

Assumption 3.2: Let the following conditions hold:

1) The set X is closed and convex.

2) Each function $f^{[i]} : R^n \to R$ is convex.

3) All functions $f^{[i]}$ have Lipschitz gradients with a constant $L : \| Df^{[i]}(x) - Df^{[i]}(y) \| \leq L \| x - y \|$ for all $x, y \in R^n$.

4) The gradients $Df^{[i]}(x), i \in V$ are bounded over the set X, i.e., and there exists a constant G such that $\| Df^{[i]}(x) \| \leq G$ for all $x \in X$ and all $i \in V$.

When each $f^{[i]}$ has Lipschitz gradient with a constant L_i, assumption 3.2(3) is satisfied with $L = \max_{i \in V} L$. When X is compact, the Assumption 3.2(4) holds. We here make the following assumptions on the network communication graphs $G(k)$.

Assumption 3.3 (Non-degeneracy): There exists a constant $\alpha > 0$ such that $a_{ii}(k) \geq \alpha$, and $a_{ij}(k)$, for $i \neq j$, satisfies $a_{ij}(k) \in \{0\} \cup [\alpha, 1]$, for all $k \geq 0$.

Assumption 3.4 (Weight-balanced): $G(k)$ is weight-balanced if $d_{\text{out}}(v) = d_{\text{in}}(v)$, for all $v \in V$.

Assumption 3.5 (Periodical Strong Connectivity): There is a positive integer B such that, for all $k_0 \geq 0$, the directed graph $(V, \bigcup_{k=0}^{B-1} E(k_0 + k))$ is strongly connected.

Lemma 3.2 （Saddle-point Theorem）: The pair of (x^*, μ^*, λ^*) is a saddle point of the function H over $X \times R^m_{\geq 0} \times R^v_{\geq 0}$ if and only if it is a pair of primal and penalty dual optimal solutions and the following penalty minimax equality holds:

$$\sup_{(\mu,\lambda)\in R^m_{\geq 0} \times R^v_{\geq 0}} \inf_{x\in X} H(x,\mu,\lambda) = \inf_{x\in X} \sup_{(\mu,\lambda)\in R^m_{\geq 0} \times R^v_{\geq 0}} H(x,\mu,\lambda)$$

Based on this characterization, we will use the subgradient method of the following section for finding the saddle points of the penalty function. We denote $w = (\mu, \lambda)$, for each $w \in R^m_{\geq 0} \times R^v_{\geq 0}$ and we define the function $H^{[i]}_w : R^n \to R$ as $H^{[i]}_w(x) = H^i(x, w)$. Note that $H^{[i]}_w(x)$ is convex in x by using the fact that a nonnegative weighted sum of convex functions is convex. For each $x \in R$, we define the function $H^{[i]}_x(x) : R^m_{\geq 0} \times R^v_{\geq 0} \to R$ as $H^{[i]}_x(w) = H^{[i]}(x, w)$. It is easy to check that $H^{[i]}_x(w)$ is concave in w. Then the penalty function $H(x,w)$ is the sum of convex-concave local functions.

Lemma 3.3 (Dynamic Average Consensus Algorithm) [21] : The following is a vector version of the first-order dynamic average consensus algorithm with $x^{[i]}(k), \xi^{[i]}(k) \in R^n$:

$$x^{[i]}(k+1) = \sum_{j=1}^{N} w_{ij}(k) x^{[j]}(k) + \xi^{[i]}(k)$$

We set $\Delta \xi_\ell(k) = \max_{i\in V} \xi^{[i]}_\ell(k) - \min_{i\in V} \xi^{[i]}_\ell(k)$ for $1 \leq \ell \leq n$. The sequences of $W(k) = [w_{ij}(k)]$ satisfy $\sum_{j=1}^{N} w_{ij}(k) = 1$ and $\sum_{i=1}^{N} w_{ij}(k) = 1$. Suppose that periodical strong connectivity Assumption 3.5 holds. Assume that $\lim_{k\to+\infty} \Delta\xi_\ell(k) = 0$ for all $1 \leq \ell \leq n$ and all $k \geq 0$. Then $\lim_{k\to+\infty} \| x^{[i]}(k) - x^{[i]}(k) \| = 0$ for all $i, j \in V$.

Proof: Define

$$M(t) = \max_{i\in V} x_i(t) \qquad m(t) = \min_{i\in V} x_i(t)$$
$$D(t) = M(t) - m(t) \qquad \Delta r^{[i]}(t) = r^{[i]}(t) - r^{[i]}(t-h)$$
$$\Delta r_{\max}(t) = \max_{i\in V} \Delta r_i(t) \qquad \Delta r_{\min}(t) = \min_{i\in V} \Delta r_i(t)$$

where $r^{[i]}(t)$ is referred to as the reference signal (or input) of node i at time t.

We propose the First-Order Dynamic Average Consensus Algorithm below to reach the dynamic average consensus:

$$x_i(t+h) = x_i(t) + \sum_{j\neq i} w_{ij}(t)(x_j(t) - x_i(t)) + \Delta r_i(t)$$

Let $s \geq 0$ and $k \in V$ be fixed. Then for every $\ell \in \{1, ..., n\}$, there exists a real number $\eta_\ell > 0$ such that for every integer $P \in [\ell B, (nB + B - 1)]$, and $i \in D_\ell$, it holds that for $t = s + ph$

$$x_i(t) \geq m(s) + \sum_{q=0}^{p-1} \Delta r_{\min}(s+qh) + \eta_\ell(x_k(s) - m(s)) \quad (6)$$

$$x_i(t) \leq M(s) + \sum_{q=0}^{p-1} \Delta r_{\max}(s+qh) + \eta_\ell(M(s) - x_k(s)) \quad (7)$$

Without loss of generality, we only consider the case where $s = 0$, being identical with the proof for a general s. Fixing some i, it holds that

$$x_i(t+h) = x_i(t) + \sum_{j\neq i} w_{ij}(t)(x_j(t) - x_i(t)) + \Delta r_i(t)$$

Let $t = 0$, we have that

$$\begin{aligned}
x_i(h) &= x_i(0) + \sum_{j\neq i} w_{ij}(0)(x_j(0) - x_i(0)) + \Delta r_i(0) \\
&= (1 - \sum_{j\neq i} w_{ij}(0))x_i(0) + \sum_{j\neq i} w_{ij}(0)x_j(0) + \Delta r_i(0) \\
&\geq (1 - \sum_{j\neq i} w_{ij}(0))m(0) + \sum_{j\neq i} w_{ij}(0)m(0) + \Delta r_{\min}(0) \\
&= m(0) + \Delta r_{\min}(0)
\end{aligned} \quad (8)$$

Since (8) holds for all i,

$$m(h) \geq m(0) + \Delta r_{\min}(0) \quad (9)$$

Applying recursive method, it follows that

$$m(t) \geq m(0) + \sum_{p=0}^{\frac{t}{h}-1} \Delta r_{\min}(ph) \quad (10)$$

Since $\sum_{j=1}^{N} w_{kj}(t) = 1$ at every $t > 0$, we have that

$$\begin{aligned}
&x_k(t+h) - m(0) - \sum_{p=0}^{\frac{t}{h}} \Delta r_{\min}(ph) \\
&= \sum_{j=1}^{N} w_{kj}(t) x_j(t) + \Delta r_k(t) - \sum_{j=1}^{N} w_{kj}(t)m(0) - \sum_{j=1}^{N} w_{kj}(t) \sum_{p=0}^{\frac{t}{h}-1} \Delta r_{\min}(ph) - \Delta r_{\min}(t) \\
&= \sum_{j=1}^{N} w_{kj}(t)(x_j(t) - m(0) - \sum_{p=0}^{\frac{t}{h}-1} \Delta r_{\min}(ph)) + \Delta r_k(t) - \Delta r_{\min}(t) \\
&\geq w_{kk}(t)(x_k(t) - m(0) - \sum_{p=0}^{\frac{t}{h}-1} \Delta r_{\min}(ph)) \\
&\geq \omega(x_k(t) - m(0) - \sum_{p=0}^{\frac{t}{h}-1} \Delta r_{\min}(ph))
\end{aligned} \quad (11)$$

where we are using the property of (10) in the last two inequalities. Applying repeatedly (11), we have that, for any integer $P \in [\ell B, (\ell B + B - 1)]$, the following holds for $t = ph$

$$x_k(t) - m(0) - \sum_{q=0}^{p-1} \Delta r_{min}(qh) \geq \omega^{p-1}(x_k(h) - m(0) - \Delta r_{min}(0))$$
$$\geq \omega^p(x_k(0) - m(0)) \geq \eta_0(x_k(0) - m(0))$$

where $\eta_0 = \omega^{NB-1}$.

Now we proceed by induction on ℓ. Suppose that (6) holds for some $0 \leq \ell \leq n$; then we should show (6) for $i \in N_{\ell+1}$. By the induction hypothesis, we have that for all integer $P \in [\ell B, (\ell B + B - 1)]$, there exists some $\eta_\ell > 0$ such that the following holds for $t = ph$

$$x_j(t) - m(0) - \sum_{\tau=0}^{p-1} \Delta r_{min}(qh) \geq \eta_\ell(x_k(0) - m(0))$$

Consequently, as in (11), we have

$$x_i(t' + h) - m(0) - \sum_{q=0}^{\frac{t'}{h}} \Delta r_{min}(qh) \geq w_{ij}(t')(x_i(t') - m(0) - \sum_{q=0}^{\frac{t'}{h}-1} \Delta r_{min}(qh))$$
$$\geq \omega \eta_\ell(x_k(0) - m(0))$$

Following along the same lines as in (11), we obtain $x_i(t' + h) - m(0) - \sum_{q=0}^{p} \Delta r_{min}(qh) \geq \eta_{\ell+1}(x_k(0) - m(0))$ for all $P \in [(\ell+1)B, (\ell B + B - 1)]$ where $\eta_{\ell+1} = \omega^{(N-\ell)B} \eta_\ell$ and $t = ph$. This establishes (6) for $i \in N_{\ell+1}$. By induction, we have shown that (6) holds. The proof for (7) is analogous.

Let $\eta = \omega^{\frac{1}{2}N(N+1)B-1}$, then $\eta \leq \eta_\ell$ for any $\ell \in \{1, ..., N-1\}$. By replacing s and t in (4) with t and $t_1 = t + (LB + B - 1)h$ respectively. We have that for every $t \geq 0$

$$m(t_1) = \min_{\ell \in \{0,...,L\}} \min_{i \in D_\ell} x_i(t_1) \geq m(t) + \sum_{q=\frac{t}{h}}^{\frac{t_1}{h}-1} \Delta r_{min}(qh) + \min_{\ell \in \{0,...,L\}} \eta_\ell(x_k(t) - m(t))$$
$$\geq m(t) + \sum_{q=\frac{t}{h}}^{\frac{t_1}{h}-1} \Delta r_{min}(qh) + \eta(x_k(t) - m(t))$$

Similarly, we can see that

$$M(t_1) \leq M(t) + \sum_{q=\frac{t}{h}}^{\frac{t_1}{h}-1} \Delta r_{max}(qh) - \eta(M(t) - x_k(t))$$

Combining the above two inequalities gives that

$$D(t_1) \leq (1-\eta)D(t) + \sum_{q=\frac{t}{h}}^{\frac{t_1}{h}-1} \Delta R(qh)$$

Denoting $T_k = k(NB-1)h$ for an integer $k \geq 1$. From (9), we know that $D(t+h) \leq D(t) + \Delta R(t)$. Thus we have

$$D(T_n) \leq (1-\eta)^n D(0) + \Omega(n)$$

where

$$\Omega(n) = (1-\eta)^{n-1} \sum_{q=0}^{\frac{T_1}{h}-1} \Delta R(qh) + ... + \sum_{q=\frac{T_{n-1}}{h}}^{\frac{T_n}{h}-1} \Delta R(qh).$$

For any $t \geq 0$, let ℓ_t be the largest integer such that $\ell_t(NB-1)h \leq t$, and $\bar{\Omega}(t) = \Omega(\ell_t) + \sum_{q=\frac{T_{\ell_t}}{h}}^{\frac{t}{h}-1} \Delta R(qh)$. Thus for

all $t \geq 0$ it follows that

$$D(t) \leq D(\ell_t) + \sum_{q=\frac{T_{\ell_t}}{h}}^{\frac{t}{h}-1} \Delta R(qh)$$
$$\leq (1-\eta)^{\ell_t} D(0) + \bar{\Omega}(t) \qquad (12)$$
$$\leq (1-\eta)^{\frac{t}{(NB-1)h}-1} D(0) + \bar{\Omega}(t)$$

Since $\Delta R(t) \leq h\theta$ and $D(t)$ are input-to-output stable with ultimate bound $\Xi \leq 4h\theta(NB-1)\frac{1}{\eta} \leq 4h\theta(NB-1)w^{-\frac{1}{2}N(N+1)B+1}$; i.e., there exist $\Gamma > 0$ and $0 \leq \lambda \leq 1$ such that

$$D(t) \leq \max\{\Gamma \lambda^{\frac{t}{h}}, \Xi\}, \qquad \forall t \geq 0$$

Choosing as initial state $x_i(0) = r_i(-h)$ for all $i \in \{1, ..., N\}$. Since $x_i(t+h) = \sum_{j \neq i} w_{ij}(t)(x_j(t) - x_i(t)) + x_i(t) + \Delta r_i(t)$, we can deduce that

$$\sum_{i=1}^{N} x_i(t+h) = \sum_{i=1}^{N} x^{[i]}(t) + \sum_{i=1}^{N} \Delta r_i(t)$$
$$= \sum_{i=1}^{N} x^{[i]}(0) + \sum_{i=1}^{N} \sum_{q=0}^{\frac{t}{h}} \Delta r_i(qh) \qquad (13)$$
$$= \sum_{i=1}^{N} x^{[i]}(0) + \sum_{i=1}^{N} (r_i(t) - r_i(-h)) = \sum_{i=1}^{N} r_i(t)$$

It follows from (13) that $m(t+h) \leq \frac{1}{N} \sum_{i=1}^{N} r_i(t) \leq M(t+h)$

and thus

$$\max_{i \in V} \limsup_{t \to \infty} | x_i(t) - \frac{1}{N} \sum_{i=1}^{N} r_i(t-h) | \leq \limsup_{t \to \infty} D(t) \leq \Xi$$

Let $\lim_{t \to \infty} \Delta R(t) = 0$, for any $h > 0$. The implementation of the Dynamic Average Consensus Algorithms ensures that $\lim_{t \to \infty} \Xi = 0$. So we can conclude that

$$\limsup_{t \to \infty} | x_i(t) - x_j(t) | \leq \limsup_{t \to \infty} D(t) \leq 0$$

Thus, $\limsup_{t \to \infty} \left| x_i(t) - x_j(t) \right| = 0$ holds.

Consider the following Distributed projected subgradient algorithm proposed in [13]: Suppose $Z \subseteq R^n$ is a closed and convex set. Let $x^{[i]}(k+1) = P_Z[v_x^{[i]}(k) - \alpha(k)d^{[i]}(k)]$. Denote $e^{[i]}(k) = P_Z[v_x^{[i]}(k) - \alpha(k)d^{[i]}(k)] - v_x^{[i]}(k)$. The following is a slight modification of Lemma 8 and its proof in [13].

Lemma 3.4: Let the non-degeneracy Assumption 3.3, the weighted-balanced Assumption 3.4, and the periodic strong connectivity Assumption 3.5 hold. Then there exist $\gamma > 0$ and $\beta \in (0,1)$ such that

$$\| x^{[i]}(k) - \hat{x}(k) \| \leq N\gamma \sum_{\tau=0}^{k-1} \beta^{k-\tau} \{ \alpha(\tau) \| d^{[i]}(\tau) \| + \| e^{[i]}(\tau) + \alpha(\tau)d^{[i]}(\tau) \| \}$$
$$+ N\gamma \beta^{k-1} \sum_{i=0}^{N} \| x^{[i]}(0) \|$$

Suppose $\{d^{[i]}(k)\}$ is uniformly bounded for each $i \in V$, and $\sum_{k=0}^{+\infty} \alpha(k)^2 < +\infty$, then we have $\sum_{k=0}^{+\infty} \alpha(k)^2 \max_{i \in V} \| x^{[i]}(k) - \hat{x}(k) \| < +\infty$.

4. Distributed Subgradient Methods

In this section, we introduce a distributed penalty primal-dual subgradient algorithm to solve the optimization problem (3), followed by its convergence properties.

Distributed Penalty Primal-Dual Subgradient Algorithm

We consider a set $V = \{1,...,n\}$ of agents. Each agent chooses any initial state $X^{[i]}(0) \in X, \mu^{[i]}(0) \in R_{\geq 0}^m, \lambda^{[i]}(0) \in R_{\geq 0}^v$, and $y^{[i]}(1) = Nf^{[i]}(X^{[i]}(0))$. At any time $k \geq 0$, each agent i computes the following convex combination:

$$v_x^{[i]}(k) = x^{[i]}(k) + h \sum_{j \in N_i(k)} a_{ij}(k)(x^{[j]}(k) - x^{[i]}(k))$$
$$v_u^{[i]}(k) = \mu^{[i]}(k) + h \sum_{j \in N_i(k)} a_{ij}(k)(\mu^{[j]}(k) - \mu^{[i]}(k))$$
$$v_\lambda^{[i]}(k) = \lambda^{[i]}(k) + h \sum_{j \in N_i(k)} a_{ij}(k)(\lambda^{[j]}(k) - \lambda^{[i]}(k))$$
$$v_y^{[i]}(k) = y^{[i]}(k) + h \sum_{j \in N_i(k)} a_{ij}(k)(y^{[j]}(k) - y^{[i]}(k))$$

and updates its estimates $x^{[i]}(k+1), \mu^{[i]}(k+1), \lambda^{[i]}(k+1)$, and $y^{[i]}(k+1)$ according to the following ways:

$$\begin{aligned}
x^{[i]}(k+1) &= P_X[v_x^{[i]}(k) - \alpha(k)S_x^{[i]}(k)] \\
\mu^{[i]}(k+1) &= v_u^{[i]}(k) + \alpha(k)[g(v_x^{[i]}(k))]^+ \\
\lambda^{[i]}(k+1) &= v_\lambda^{[i]}(k) + \alpha(k) | h(v_x^{[i]}(k)) | \\
y^{[i]}(k+1) &= v_y^{[i]}(k) + N(f^{[i]}(x^{[i]}(k)) - f^{[i]}(x^{[i]}(k-1)))
\end{aligned} \qquad (14)$$

where the scalars $a_{i1}(k), a_{i2}(k),..., a_{in}(k)$ are nonnegative weights and the positive scalars $\{\alpha(k)\}$ are step-sizes, P_X is the projector onto the set X. The vector $S_x^{[i]}(k)$ is a subgradient of the agent i 's penalty function $H_{w^{[i]}(k)}^{[i]}(x)$ at $x = v_x^{[i]}(k)$, where $w^{[i]}(k) = (v_u^{[i]}(k), v_\lambda^{[i]}(k))$ is the convex combination of dual estimates of agent i and its neighbors'. $S_x^{[i]}(k)$ keeps to the following rules:

$$S_x^{[i]}(k) = Df^{[i]}(v_x^{[i]}(k)) + \sum_{\ell=1}^{m} v_u^{[i]}(k)_\ell Dg_\ell[(v_x^{[i]}(k))]^+$$
$$+ \sum_{\ell=1}^{v} v_\lambda^{[i]}(k)_\ell D | h_\ell | (v_x^{[i]}(k))$$

Remark 4.1: Since $v_x^{[i]}(k) = x^{[i]}(k) + h \sum_{j \in N_i(k)} a_{ij}(k)(x^{[j]}(k) - x^{[i]}(k))$, it follows that

$$v_x^{[i]}(k) = x^{[i]}(k) - h \sum_{j=1}^{N} l_{ij}(k) x^{[j]}(k)$$

where $L(k) = [l_{ij}(k)]$ is the Laplacian matrix such that $L1_N = 0_N, 1_N^T L = 0_N^T$.

Proof: Modifying the second term on the right-hand side in the above formula, we then have

$$v_x^{[i]}(k) = x^{[i]}(k) - hl_{ii}(k) x^{[i]}(k) - h \sum_{j=1, j \neq i}^{N} l_{ij}(k) x^{[j]}(k)$$
$$= (1 - hl_{ii}(k)) x^{[i]}(k) - h \sum_{j=1, j \neq i}^{N} l_{ij}(k) x^{[j]}(k)$$

Let $w_{ii}(k) = (1 - hl_{ii}(k)), w_{ij}(k) = -hl_{ij}(k)$, one has

$$v_x^{[i]}(k) = \sum_{j=1}^{N} w_{ij}(k) x^{[j]}(k)$$

Since graph $G(k)$ is balanced, then $\sum_{j=1}^{N} a_{ij}(k) = 0$ and $\sum_{i=1}^{N} a_{ij}(k) = 0$. We can conclude that $\sum_{j=1}^{N} w_{ij}(k) = 1$ and $\sum_{i=1}^{N} w_{ij}(k) = 1$ hold under the condition that h satisfies $1 - hl_{ii}(k) > 0$. Similarly, we obtain

$$v_\mu^{[i]}(k) = \sum_{j=1}^N w_{ij}(k)\mu^{[j]}(k) \ , \quad v_\lambda^{[i]}(k) = \sum_{j=1}^N w_{ij}(k)\lambda^{[j]}(k) \quad \text{and}$$

$$v_y^{[i]}(k) = \sum_{j=1}^N w_{ij}(k)y^{[j]}(k) \ .$$

Assumption 4.1 (Step-size assumption): The step-sizes satisfy $\lim_{k\to+\infty}\alpha(k)=0$, $\sum_{k=0}^{+\infty}\alpha(k)=+\infty$, $\sum_{k=0}^{+\infty}\alpha(k)^2<+\infty$ and $\lim_{k\to+\infty}\alpha(k+1)s(k)=0$, $\sum_{k=0}^{+\infty}\alpha(k)^2 s(k)<+\infty$, $\sum_{k=0}^{+\infty}\alpha(k+1)^2 s(k)^2<+\infty$.

In the following, we study the convergence behavior of the subgradient algorithm introduced in this section where the optimal solution and the optimal value is asymptotically agreed upon.

Theorem 4.1 (Convergence properties of the DPPDS algorithm): Consider the problem (3). Let the non-degeneracy Assumption 3.3, the weight-balanced Assumption 3.4 and the periodic strong connectivity Assumption 3.5 hold. Consider the sequences of $\{x^{[i]}(k)\}$ and $\{y^{[i]}(k)\}$ of the distributed penalty primal-dual subgradient algorithm, where the step-sizes $\{\alpha(k)\}$ satisfy the step-size Assumption 4.1. Then there exists a primal optimal solution $\tilde{x}\in X^*$ such that $\lim_{k\to+\infty}\|x^{[i]}(k)-\tilde{x}\|=0$ for all $i\in V$. Furthermore, we have $\lim_{k\to+\infty}\|y^{[i]}(k)-f^*\|=0$ for all $i\in V$.

Remark 4.2: The distributed penalty primal-dual subgradient algorithm takes the equality constraint into account. The presence of the equality constrain can make M^* unbounded. Therefore, unlike other subgradient algorithm, e.g., [15], [16], the distributed penalty primal-dual subgradient algorithm does not involve the dual projection steps onto compact sets. So we do not guarantee the subgradient $S_x^{[i]}(k)$ not to be absolutely bounded, while the boundedness of subgradients is a standard assumption in the analysis of subgradient methods, e.g., see [6], [13], [17], [18], [19], [20]. The step-size of Assumption 3.1 is stronger than the more standard diminishing step-size scheme in [22] and this will correctly deal with the difficulty of the boundedness of $S_x^{[i]}(k)$. We give this condition in order to prove, in the absence of the boundedness of $S_x^{[i]}(k)$, the existence of a number of limits and summability of expansion toward Theorem 4.1. Finally, we adopt the penalty relaxation instead of the Lagrangian relaxation in this paper.

Remark 4.3 (Penalty subgradient inequality): Observe that $\mu^{[i]}(k)\ge 0, \lambda^{[i]}(k)\ge 0$ and $v_x^{[i]}(k)\in X$ (due to the fact that X is convex and $v_x^{[i]}(k)=\sum_{j=1}^N w_{ij}(k)x^{[i]}(k)$). Moreover, $(g[(v_x^{[i]}(k))]^+, |h(v_x^{[i]}(k))|)$ is a supgradient of $H_{w^{[i]}(k)}^{[i]}(w^{[i]}(k))$; i.e. the following penalty supgradient inequality holds for any $\mu\in R_{\ge 0}^m$ and $\lambda\in R_{\ge 0}^v$:

$$\begin{aligned} &(g[(v_x^{[i]}(k))]^+)^T(\mu - v_\mu^{[i]}(k)) + |h(v_x^{[i]}(k))|^T (\lambda - v_\lambda^{[i]}(k))\\ &\ge H^{[i]}(v_x^{[i]}(k),\mu,\lambda) - H^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k), v_\lambda^{[i]}(k)) \end{aligned} \quad (15)$$

Proof: Observe   that $H^{[i]}(x,\mu,\lambda)=f^{[i]}(x)+\mu^T[g(x)]^+ + \lambda^T|h(x)|$ holds for all $\mu^{[i]}(k)\ge 0$, $\lambda^{[i]}(k)\ge 0$, $v_x^{[i]}(k)\in X$ and i is arbitrarily. Thus,

$$H^{[i]}(v_x^{[i]}(k),\mu,\lambda)=f^{[i]}(v_x^{[i]}(k))+\mu^T[g(v_x^{[i]}(k))]^+ + \lambda^T|h(v_x^{[i]}(k))|$$

and

$$\begin{aligned} H^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k), v_\lambda^{[i]}(k)) = &f^{[i]}(v_x^{[i]}(k)) + (v_\mu^{[i]}(k))^T[g(v_x^{[i]}(k))]^+\\ &+ (v_\lambda^{[i]}(k))^T|h(v_x^{[i]}(k))| \end{aligned}$$

Followed by the properties of supgradient, we obtain

$$\begin{aligned} &H^{[i]}(v_x^{[i]}(k),\mu,\lambda) - H^{[i]}(v_x^{[i]}(k), v_\mu^{[i]}(k), v_\lambda^{[i]}(k))\\ &\le (\mu - v_\mu^{[i]}(k))^T[g(v_x^{[i]}(k))]^+ + (\lambda - v_\lambda^{[i]}(k))^T |h(v_x^{[i]}(k))|\\ &\le (g[(v_x^{[i]}(k))]^+)^T(\mu - v_\mu^{[i]}(k)) + |h(v_x^{[i]}(k))|^T (\lambda - v_\lambda^{[i]}(k)) \end{aligned}$$

Remark 4.4: In this paper, we apply the harmonic series $\left\{\alpha(k)=\dfrac{1}{k+1}\right\}_{k\in Z\ge 0}$ into our subgradient algorithm. It's easy to check that $\left\{\alpha(k)=\dfrac{1}{k+1}\right\}_{k\in Z\ge 0}$ satisfies the step-size Assumption 4.1 (for more details, one may refer to [14]).

A. Convergence Analysis

In the following, we will prove convergence of the distributed penalty primal-dual subgradient algorithm. First, we rewrite our algorithm into the following form:

$$\mu^{[i]}(k+1) = v_\mu^{[i]}(k) + u_\mu^{[i]}(k), \qquad \lambda^{[i]}(k+1) = v_\mu^{[i]}(k) + u_\mu^{[i]}(k)$$

$$x^{[i]}(k+1) = v_x^{[i]}(k) + e_x^{[i]}(k), \qquad y^{[i]}(k+1) = v_y^{[i]}(k) + u_y^{[i]}(k)$$

where $e_x^{[i]}(k)$ is projection error described by

$$e_x^{[i]}(k) = P_X[v_x^{[i]}(k) - \alpha(k)S_x^{[i]}(k)] - v_x^{[i]}(k)$$

and $u_\mu^{[i]}(k)=\alpha(k)[g(v_x^{[i]}(k))]^+$, $u_\lambda^{[i]}(k)=\alpha(k)|h(v_x^{[i]}(k))|$, $u_y^{[i]}(k)=N(f^{[i]}(x^{[i]}(k))-f^{[i]}(x^{[i]}(k-1)))$ are some local inputs. Denote the maximum deviations of dual estimates by $M_\mu(k)=\max_{i\in V}\|\mu^{[i]}(k)\|$ and $M_\lambda(k)=\max_{i\in V}\|\lambda^{[i]}(k)\|$. We further denote the averages of primal and dual estimates as $\bar{x}(k)=\dfrac{1}{N}\sum_{i=1}^N x^{[i]}(k), \bar{\mu}(k)=\dfrac{1}{N}\sum_{i=1}^N\mu^{[i]}(k)$ and $\bar{\lambda}(k)=\dfrac{1}{N}\sum_{i=1}^N\lambda^{[i]}(k)$. Since X is compact, and $f^{[i]},[g(\cdot)]^+$ and h are continuous, there exist $F,G^+,H>0$ such that for all $x\in X$, it holds that $\|f^{[i]}(x)\|\le F$ for all $i\in V$, $\|[g(x)]^+\|\le G^+$ and $\|h(x)\|\le H$. Since X is a compact set

and $f^{[i]}$, $[g(x)]^+$, $|h_\ell(\cdot)|$ are convex, then it follows from Proposition 5.4.2 in [19] that there exist $D_F, D_{G^+}, D_H > 0$ such that for all $x \in X$, we have that $\| Df^{[i]}(x) \| \leq D_F$ $(i \in V)$, $m \| D[g_\ell(x)]^+ \| \leq D_{G^+}$ $(1 \leq \ell \leq m)$ and $\| D|h_\ell|(x) \| \leq D_H$ $(1 \leq \ell \leq v)$.

Lemma 4.1: Let $K \geq 0$. Consider the sequence $\{\delta(k)\}$ defined by $\{\delta(k)\} = \dfrac{\sum_{\ell=k}^{k-1} \alpha(\ell)\rho(\ell)}{\sum_{\ell=k}^{k-1} \alpha(\ell)}$, where $k \geq K+1$, $\alpha(k) > 0$ and $\sum_{k=K}^{+\infty} \alpha(k) = +\infty$.

    (a) If $\lim_{k \to +\infty} \rho(k) = +\infty$, then $\lim_{k \to +\infty} \delta(k) = +\infty$.

    (b) If $\lim_{k \to +\infty} \rho(k) = \rho^*$, then $\lim_{k \to +\infty} \delta(k) = \rho^*$.

The proof of Lemma 4.1 can be referred to Lemma 5.1 in [14].

Lemma 4.2 (Diminishing and summable properties): Suppose the weighted-balanced Assumption 3.4 and the step-size Assumption 4.1 hold.

    (a) It holds that $\lim_{k \to +\infty} \alpha(k)M_\mu(k) = 0$, $\lim_{k \to +\infty} \alpha(k)M_\lambda(k) = 0$, $\lim_{k \to +\infty} \alpha(k) \| S_x^{[i]}(k) \| = 0$, and the sequences of $\{\alpha(k)^2 M_\mu(k)^2\}$, $\{\alpha(k)^2 M_\lambda(k)^2\}$ and $\{\alpha(k)^2 \| S_x^{[i]}(k) \|^2\}$ are summable.

    (b) The sequences of $\{\alpha(k) \| \hat{\mu}(k) - v_\mu^{[i]}(k) \|\}$, $\{\alpha(k) \| \hat{\lambda}(k) - v_\lambda^{[i]}(k) \|\}$, $\{\alpha(k)M_\mu(k) \| \hat{x}(k) - v_x^{[i]}(k) \|\}$, $\{\alpha(k)M_\lambda(k) \| \hat{x}(k) - v_x^{[i]}(k) \|\}$ and $\{\alpha(k) \| \hat{x}(k) - v_x^{[i]}(k) \|\}$ are summable.

Proof: (a) Noticing that

$$M_\mu(k) = \max_{i \in V} \| \mu^{[i]}(k) \|, v_\mu^{[i]}(k) = \sum_{j=1}^N w_{ij}(k)\mu^{[j]}(k), \sum_{j=1}^N w_{ij}(k) = 1, \sum_{i=1}^N w_{ij}(k) = 1$$

Then, we show that

$$\| v_\mu^{[i]}(k) \| = \| \sum_{j=1}^N w_{ij}(k)\mu^{[j]}(k) \| \leq \sum_{j=1}^N w_{ij}(k) \| \mu^{[j]}(k) \| \leq \sum_{j=1}^N w_{ij}(k)M_\mu(k) = M_\mu(k)$$

Recalling that $v_x^{[i]}(k) \in X$, $\mu^{[i]}(k+1) = v_\mu^{[i]}(k) + \alpha(k)[g(v_x^{[i]}(k))]^+$. This implies that the following inequalities hold for all $k \geq 0$:

$$\| \mu^{[i]}(k+1) \| = \| v_\mu^{[i]}(k) + \alpha(k)[g(v_x^{[i]}(k))]^+ \| \leq \| v_\mu^{[i]}(k) \| + G^+\alpha(k) \leq M_\mu(k) + G^+\alpha(k)$$

Then we deduce the following recursive estimate on $M_\mu(k+1) \leq M_\mu(k) + G^+\alpha(k)$. Repeatedly applying the above estimates yields that

$$M_\mu(k+1) \leq M_\mu(0) + G^+ s(k) \tag{16}$$

where $s(k) = \alpha(0) + \alpha(1) + \cdots \alpha(k)$.

Similar arguments can be employed to show that

$$M_\lambda(k+1) \leq M_\lambda(0) + Hs(k) \tag{17}$$

Since $\lim_{k \to +\infty} \alpha(k+1)s(k) = 0$ and $\lim_{k \to +\infty} \alpha(k) = 0$, then we know that $\lim_{k \to +\infty} \alpha(k+1)M_\mu(k+1) = 0$ and $\lim_{k \to +\infty} \alpha(k+1)M_\lambda(k+1) = 0$. Noticing that

$$\| S_x^{[i]}(k) \| = \| Df^{[i]}(v_x^{[i]}(k)) + \sum_{l=1}^m v_\mu^{[i]}(k)_\ell Dg_\ell[(v_x^{[i]}(k))]^+ + \sum_{\ell=1}^v v_\lambda^{[i]}(k)_\ell D|h_\ell|(v_x^{[i]}(k)) \|$$

$$\leq \| Df^{[i]}(v_x^{[i]}(k)) \| + \sum_{l=1}^m \| v_\mu^{[i]}(k)_\ell Dg_\ell[(v_x^{[i]}(k))]^+ \| + \sum_{\ell=1}^v \| v_\lambda^{[i]}(k)_\ell D|h_\ell|(v_x^{[i]}(k)) \|$$

Then, the following estimate on $S_x^{[i]}(k)$ holds:

$$\| S_x^{[i]}(k) \| \leq D_F + D_{G^+}M_\mu(k) + D_H M_\lambda(k) \tag{18}$$

Recalling that $\lim_{k \to +\infty} \alpha(k) = 0$, $\lim_{k \to +\infty} \alpha(k)M_\mu(k) = 0$ and $\lim_{k \to +\infty} \alpha(k)M_\lambda(k) = 0$. Then we have $\lim_{k \to +\infty} \alpha(k) \| S_x^{[i]}(k) \| = 0$. By (16), we obtain

$$\sum_{k=0}^{+\infty} \alpha(k)^2 M_\mu^2(k) \leq \alpha(0)^2 M_\mu^2(0) + \sum_{k=1}^{+\infty} \alpha(k)^2 (M_\mu(0) + G^+ s(k-1))^2$$

It follows from the step-size Assumption 4.1 that $\sum_{k=0}^{+\infty}\alpha(k)^2 M_\mu^2(k)<+\infty$. Similarly, one can show that $\sum_{k=0}^{+\infty}\alpha(k)^2 M_\lambda^2(k)<+\infty$. Multiplying both sides of $M_\mu(k+1)\le M_\mu(k)+G^+\alpha(k)$ by $\alpha(k)$ and square, then we deduce the following recursive estimate:

$$\sum_{k=0}^{+\infty}\alpha(k)^2\|S_x^{[i]}(k)\|^2\le\alpha(0)^2(D_F+D_G+M_\mu(0)+D_H M_\lambda(0))^2$$
$$+\sum_{k=1}^{+\infty}\alpha(k)^2(D_F+D_G+(M_\mu(0)+G^+s(k-1))+D_H(M_\lambda(0)+Hs(k-1)))^2$$

Then the summability of $\{\alpha(k)^2\}$, $\{\alpha(k+1)^2 s(k)\}$ and $\{\alpha(k+1)^2 s(k)^2\}$ testifies that of $\{\alpha(k)^2\|S_x^{[i]}(k)\|^2\}$.

(b) Noticing that

$$\|\hat{\mu}(k)-v_\mu^{[i]}(k)\|\le\sum_{j=1}^N w_{ij}(k)\|\hat{\mu}(k)-\mu^{[j]}(k)\|\le\max_{i\in V}\|\hat{\mu}(k)-\mu^{[i]}(k)\| \tag{19}$$

Then following from Lemma 3.4 with $Z=R_{\ge 0}^m$ and $d^{[i]}(k)=-[g(v_x^{[i]}(k))]^+$, we have the summability of $\{\alpha(k)\max_{i\in V}\|\hat{\mu}(k)-\mu^{[i]}(k)\|\}$. Then $\{\alpha(k)\|\hat{\mu}(k)-v_\mu^{[i]}(k)\|\}$ is summable. Similarly, it holds that $\sum_{k=0}^{+\infty}\alpha(k)\|\hat{\lambda}(k)-v_\lambda^{[i]}(k)\|<+\infty$.

We now consider the evolution of $x^{[i]}(k)$. Recalling that $v_x^{[i]}(k)\in X$. By Lemma 2.1, let $Z=X$, $z=v_x^{[i]}(k)-\alpha(k)S_x^{[i]}(k)$ and $y=v_x^{[i]}(k)$, we get

$$\|x^{[i]}(k+1)-v_x^{[i]}(k)\|^2\le\|v_x^{[i]}(k)-\alpha(k)S_x^{[i]}(k)-v_x^{[i]}(k)\|^2$$
$$-\|x^{[i]}(k+1)-(v_x^{[i]}(k)-\alpha(k)S_x^{[i]}(k))\|^2$$

Regrouping the above estimates, we obtain

$$\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|\le\alpha(k)\|S_x^{[i]}(k)\|$$

With the above relation, from Lemma 3.4 with $Z=X$ and $d^{[i]}(k)=S_x^{[i]}(k)$, the following holds for some $\gamma>0$ and $0<\beta<1$:

$$\|x^{[i]}(k)-\hat{x}(k)\|\le N\gamma\beta^{k-1}\sum_{i=0}^N\|x^{[i]}(0)\|+2N\gamma\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)\|S_x^{[i]}(\tau)\| \tag{20}$$

Multiplying both side of (20) by $\alpha(k)M_\mu(k)$ and using (18), for all $i\in V$, it yields

$$\alpha(k)M_\mu(k)\|x^{[i]}(k)-\hat{x}(k)\|\le N\gamma\sum_{i=0}^N\|x^{[i]}(0)\|\alpha(k)M_\mu(k)\beta^{k-1}$$
$$+2N\gamma\alpha(k)M_\mu(k)\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)(D_F+D_G+M_\mu(\tau)+D_H M_\lambda(\tau))$$

By applying the relation of $ab\le\dfrac{1}{2}(a^2+b^2)$ and sorting out, we get

$$\alpha(k)M_\mu(k)\max_{i\in V}\|x^{[i]}(k)-\hat{x}(k)\|\le N\gamma(\frac{1}{2}\sum_{i=0}^N\|x^{[i]}(0)\|+(D_F+D_{G^+}+D_H)\sum_{\tau=0}^{k-1}\beta^{k-\tau})$$
$$\times\alpha(k)^2 M_\mu(k)^2+\frac{1}{2}N\gamma\sum_{i=0}^N\|x^{[i]}(0)\|\beta^{2(k-1)}$$
$$+N\gamma\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)^2(D_F+D_{G^+}M_\mu^2(\tau)+D_H M_\lambda^2(\tau))$$

Part (a) gives that $\{\alpha(k)^2 M_\mu(k)^2\}$ is summable. Meanwhile, $\sum_{i=0}^N\|x^{[i]}(0)\|,D_F,D_{G^+},D_H$ are bounded, and $\sum_{\tau=0}^{k-1}\beta^{k-\tau}\le\sum_{k=0}^{+\infty}\beta^k=\dfrac{1}{1-\beta}$, then we can say that the first term on the right-hand side in the above estimate is summable.

Recalling that $N\gamma\sum_{i=0}^{N}\|x^{[i]}(0)\|\beta^{2(k-1)} \leq \dfrac{N\gamma\sum_{i=0}^{N}\|x^{[i]}(0)\|}{1-\beta}$, it's easy to check that the second term is also summable. Part (a)

gives that $\lim_{k\to+\infty}\alpha(k)^2((D_F+D_{G^+}M_\mu^2(k)+D_H M_\lambda^2(k)))=0$ and $\{\alpha(k)^2((D_F+D_{G^+}M_\mu^2(k)+D_H M_\lambda^2(k)))\}$ is summable. Then

according to the Lemma 7 in [13] with $\gamma_\ell = N\gamma\alpha(\ell)^2(D_F+D_{G^+}M_\mu^2(\ell)+D_H M_\lambda^2(\ell))=0$, ensure that the third term is

summable. In summary, $\{\alpha(k)M_\mu(k)\max_{i\in V}\|x^{[i]}(k)-\hat{x}(k)\|\}$ is summable. Following the same lines in (19), one can show the

summability of $\{\alpha(k)M_\mu(k)\|v_x^{[i]}(k)-\hat{x}(k)\|\}$. Similarly, $\{\alpha(k)M_\lambda(k)\|v_x^{[i]}(k)-\hat{x}(k)\|\}$ and $\{\alpha(k)\|v_x^{[i]}(k)-\tilde{x}(k)\|\}$ are

summable.

Lemma 4.3 (Basic iteration relation): The following estimates hold for any $x\in X$ and $(\mu,\lambda)\in R_{\geq0}^m\times R_{\geq0}^v$:

$$\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2 \leq \sum_{i=1}^{N}\alpha(k)^2\|S_x^{[i]}(k)\|^2 + \sum_{i=1}^{N}(\|x^{[i]}(k)-x\|^2-\|x^{[i]}(k+1)-x\|^2)$$
$$-\sum_{i=1}^{N}2\alpha(k)(H^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))) \tag{21}$$

and

$$0\leq\sum_{i=1}^{N}(\|\mu^{[i]}(k)-\mu\|^2-\|\mu^{[i]}(k+1)-\mu\|^2)+\sum_{i=1}^{N}(\|\lambda^{[i]}(k)-\lambda\|^2-\|\lambda^{[i]}(k+1)-\lambda\|^2)$$
$$+\sum_{i=1}^{N}2\alpha(k)(H^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(v_x^{[i]}(k),\mu,\lambda)) \tag{22}$$
$$+\sum_{i=1}^{N}\alpha(k)^2(\|[g(v_x^{[i]}(k))]^+\|^2+\|h(v_x^{[i]}(k))\|)^2$$

Proof: By Lemma 2.1, we can deduce that $\|P_Z[z]-z\|^2\leq\|z-y\|^2-\|P_Z[z]-y\|^2$. Let
$Z=X, z=v_x^{[i]}(k)-\alpha(k)S_x^{[i]}(k), y=x\in X$, we have

$$\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2 \leq \sum_{i=1}^{N}\|v_x^{[i]}(k)-\alpha(k)S_x^{[i]}(k)-x\|^2 - \sum_{i=1}^{N}\|x^{[i]}(k+1)-x\|^2$$

Expanding and regrouping the above formula, we obtain

$$\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2 \leq \sum_{i=1}^{N}\alpha(k)^2\|S_x^{[i]}(k)\|^2 - \sum_{i=1}^{N}2\alpha(k)S_x^{[i]}(k)^{\mathrm{T}}(v_x^{[i]}(k)-x)$$
$$+\sum_{i=1}^{N}\|x^{[i]}(k)-x\|^2 - \sum_{i=1}^{N}\|x^{[i]}(k+1)-x\|^2$$

Owing to the subgradient inequality $S_x^{[i]}(k)^{\mathrm{T}}(v_x^{[i]}(k)-x)\leq H^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))$,
it follows that:

$$\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2 \leq \sum_{i=1}^{N}\alpha(k)^2\|S_x^{[i]}(k)\|^2 + \sum_{i=1}^{N}(\|x^{[i]}(k)-x\|^2-\|x^{[i]}(k+1)-x\|^2)$$
$$-\sum_{i=1}^{N}2\alpha(k)(H^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x,v_\mu^{[i]}(k),v_\lambda^{[i]}(k)))$$

Lemma 4.4 (Achieving consensus): Let assumption 3.3-3.5 holds. Consider the sequences of $\{x^{[i]}(k)\},\{\lambda^{[i]}(k)\},\{\mu^{[i]}(k)\}$

and $\{y^{[i]}(k)\}$ of the distributed penalty primal-dual subgradient algorithm with the step-size sequence $\{\alpha(k)\}$ and the

associated $\{s(k)\}$ satisfy $\lim_{k\to+\infty}\alpha(k)=0$, $\lim_{k\to+\infty}\alpha(k+1)s(k)=0$. Then there exists $\tilde{x}\in X$ such that $\lim_{k\to+\infty}\|x^{[i]}(k)-\tilde{x}\|=0$ for all

$i\in V$. Furthermore, $\lim_{k\to+\infty}\|\mu^{[i]}(k)-\mu^{[j]}(k)\|=0$, $\lim_{k\to+\infty}\|\lambda^{[i]}(k)-\lambda^{[j]}(k)\|=0$ and $\lim_{k\to+\infty}\|y^{[i]}(k)-y^{[j]}(k)\|=0$ for all $i,j\in V$.

Proof: By Lemma 4.3, we see that

$$\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2 \leq \sum_{i=1}^{N}\alpha(k)^2\|S_x^{[i]}(k)\|^2 - \sum_{i=1}^{N}2\alpha(k)S_x^{[i]}(k)^{\mathrm{T}}(v_x^{[i]}(k)-x)$$
$$+\sum_{i=1}^{N}\|x^{[i]}(k)-x\|^2 - \sum_{i=1}^{N}\|x^{[i]}(k+1)-x\|^2$$

Owing to $0\leq\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2$, one can show that

$$\sum_{i=1}^{N}\|x^{[i]}(k+1)-x\|^2 \leq \sum_{i=1}^{N}\alpha(k)^2\|S_x^{[i]}(k)\|^2 + \sum_{i=1}^{N}2\alpha(k)\|S_x^{[i]}(k)\|\|(v_x^{[i]}(k)-x)\| + \sum_{i=1}^{N}\|x^{[i]}(k)-x\|^2$$

Since $\lim_{k\to+\infty}\alpha(k)\|S_x^{[i]}(k)\|=0$, taking the limit on k in the above inequality, then it follows that $\forall x\in X$:

$$\lim_{k\to+\infty}\sup\sum_{i=1}^{N}\|x^{[i]}(k+1)-x\|^2 \le \lim_{k\to+\infty}\inf\sum_{i=1}^{N}\|x^{[i]}(k)-x\|^2$$

and thus $\lim_{k\to+\infty}\sum_{i=1}^{N}\|x^{[i]}(k)-x\|^2$ exists for any $x\in X$:

$$\lim_{k\to+\infty}\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)S_x^{[i]}(k)\|^2 \le \lim_{k\to+\infty}\left\{\sum_{i=1}^{N}\alpha(k)^2\|S_x^{[i]}(k)\|^2+\sum_{i=1}^{N}\{\|x^{[i]}(k)-x\|^2-\|x^{[i]}(k+1)-x\|^2\}\right.$$
$$\left.-\sum_{i=1}^{N}2\alpha(k)(H^{[i]}(v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x,v_\mu^{[i]}(k),v_\lambda^{[i]}(k)))\right\}$$

On the other hand, taking limits on both sides of the above inequality, we have $\lim_{k\to+\infty}\sum_{i=1}^{N}\|e_x^{[i]}(k)+\alpha(k)D_x^{[i]}(k)\|^2=0$ and therefore we deduce that $\lim_{k\to+\infty}\|e_x^{[i]}(k)\|=0$ for all $i\in V$. It follows from Lemma 4.1 that $\lim_{k\to+\infty}\|x^{[i]}(k)-x^{[j]}(k)\|=0$ for all $i,j\in V$. Combining this with the property, we can deduce that $\lim_{k\to+\infty}\|x^{[i]}(k)-\tilde{x}\|=0$ for all $i\in V$.

Since $\lim_{k\to+\infty}x^{[i]}(k)=\tilde{x}$, $f^{[i]}$ is continuous, for $u_y^{[i]}(k)=N(f^{[i]}(x^{[i]}(k))-f^{[i]}(x^{[i]}(k-1)))$, $u_\mu^{[i]}(k)=\alpha(k)[g(v_x^{[i]}(k))]^+$ and $u_\lambda^{[i]}(k)=\alpha(k)|h(v_x^{[i]}(k))|$, we can deduce that $\lim_{k\to+\infty}u_y^{[i]}(k)=0$, $\lim_{k\to+\infty}u_\mu^{[i]}(k)=0$, $\lim_{k\to+\infty}u_\lambda^{[i]}(k)=0$.

Claim 1: For any $x^*\in X^*$ and $(\mu^*,\lambda^*)\in M^*$, the sequences of $\{\alpha(k)[\sum_{i=1}^{N}H^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H(x^*,\hat{\mu}(k),\hat{\lambda}(k))]\}$ and $\{\alpha(k)[\sum_{i=1}^{N}H^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-H(\hat{x}(k),\mu^*,\lambda^*]\}$ are summable

Proof: Observing that

$$H^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H(x^*,\hat{\mu}(k),\hat{\lambda}(k))$$
$$=f^{[i]}(x^*)+(v_\mu^{[i]}(k))^{\mathrm{T}}[g(x^*)]^++(v_\lambda^{[i]}(k))^{\mathrm{T}}|h(x^*)|-f^{[i]}(x^*)+(\hat{\mu}(k))^{\mathrm{T}}[g(x^*)]^++(\hat{\lambda}(k))^{\mathrm{T}}|h(x^*)|$$
$$=((v_\mu^{[i]}(k))^{\mathrm{T}}-(\hat{\mu}(k))^{\mathrm{T}})[g(x^*)]^++((v_\lambda^{[i]}(k))^{\mathrm{T}}-(\hat{\lambda}(k))^{\mathrm{T}})|h(x^*)|$$

Recalling that $\|f^{[i]}(x)\|\le F,\|[g(x)]^+\|\le G^+,\|h(x)\|\le H$, we then have

$$\|H^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H(x^*,\hat{\mu}(k),\hat{\lambda}(k))\|$$
$$\le\|((v_\mu^{[i]}(k))^{\mathrm{T}}-(\hat{\mu}(k))^{\mathrm{T}})[g(x^*)]^+\|+\|((v_\lambda^{[i]}(k))^{\mathrm{T}}-(\hat{\lambda}(k))^{\mathrm{T}})|h(x^*)|\|$$
$$\le\|v_\mu^{[i]}(k)-\hat{\mu}(k)\|\|g(x^*)^+\|+\|v_\lambda^{[i]}(k)-\hat{\lambda}(k)\|\|h(x^*)\|$$
$$\le G^+\|v_\mu^{[i]}(k)-\hat{\mu}(k)\|+H\|v_\lambda^{[i]}(k)-\hat{\lambda}(k)\|$$

By using the summability of $\{\alpha(k)\|\hat{\mu}(k)-v_\mu^{[i]}(k)\|\}$ and $\{\alpha(k)\|\hat{\lambda}(k)-v_\lambda^{[i]}(k)\|\}$ in part (b) of Lemma 4.2, we have that $\{\alpha(k)\sum_{i=1}^{N}\|H^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\|\}$ are summable. Similarly, the following estimates hold:

$$H^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-H(\hat{x}(k),\mu^*,\lambda^*)$$
$$=f^{[i]}(v_x^{[i]}(k))+(\mu^*)^{\mathrm{T}}[g(v_x^{[i]}(k))]^++(\lambda^*)^{\mathrm{T}}|h(v_x^{[i]}(k))|-f^{[i]}(\hat{x}(k)+(\mu^*)^{\mathrm{T}}[g(\hat{x}(k)]^++(\lambda^*)^{\mathrm{T}}|h(\hat{x}(k)|$$

Due to $\|Df^{[i]}(x)\|\le D_F(i\in V)$, $m\|D[g_\ell(x)]^+\|\le D_{G^+}(1\le\ell\le m)$ and $v\|D|h_\ell|(x)\|\le D_H(1\le\ell\le v)$ holds for all $x\in X$, the following estimates hold:

$$\|H^{[i]}(v_x^{[i]}(k),\mu^*,\lambda^*)-H(\hat{x}(k),\mu^*,\lambda^*)\|\le\|f^{[i]}(v_x^{[i]}(k))-f^{[i]}(\hat{x}(k))\|$$
$$+\|(\mu^*)^{\mathrm{T}}([g(v_x^{[i]}(k))]^+-[g(\hat{x}(k)]^+)\|$$
$$+\|(\lambda^*)^{\mathrm{T}}(|h(v_x^{[i]}(k))|-|h(\hat{x}(k))|)\|$$
$$\le(D_F+D_{G^+}\|\mu^*\|+D_H\|\lambda^*\|)\|v_x^{[i]}(k)-\hat{x}(k)\|$$

Then the property of $\sum_{k=0}^{+\infty}\alpha(k)\|\hat{x}(k)-v_x^{[i]}(k)\|<+\infty$ in part (b) in Lemma 4.2 implies the summability of the sequence $\{\alpha(k)\sum_{i=1}^{N}\|H^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x^*,\hat{\mu}(k),\hat{\lambda}(k))\|\}$ and that of

$\{\alpha(k)[\sum_{i=1}^{N}(H^{[i]}(x^*,v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H^{[i]}(x^*,\hat{\mu}(k),\hat{\lambda}(k)))]\}$.

Claim 2: Denote the weighted version of the local penalty function $H^{[i]}$ over $[0,k-1]$ as $\hat{H}^{[i]}(k)=\dfrac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)H(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))$. The following property holds: $\lim_{k\to+\infty}\sum_{i=1}^{N}\hat{H}^{[i]}(k)=f^*$.

Proof: Summing (21) over $[0,k-1]$ and replacing x by $x^*\in X^*$, we can deduce that

$$\sum_{\ell=0}^{k-1}\alpha(\ell)\sum_{i=1}^{N}(H^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-H^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell)))$$
$$\leq\sum_{i=1}^{N}\|x^{[i]}(0)-x^*\|^2+\sum_{\ell=0}^{k-1}\sum_{i=1}^{n}\alpha(\ell)\|S_x^{[i]}(\ell)\|^2 \tag{23}$$

The summability of $\{\alpha(k)^2\|S_x^{[i]}(k)\|^2\}$ in Part (b) implies that the right-hand side of (21) is finite as $k\to+\infty$, and thus

$$\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)[\sum_{i=1}^{N}(H^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-H^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell)))]\leq 0 \tag{24}$$

On the other hand, (x^*,μ^*,λ^*) is a saddle point of H over $X\times R_{\geq0}^m\times R_{\geq0}^v$. Since $(\hat{\mu}(k),\hat{\lambda}(k))\in R_{\geq0}^m\times R_{\geq0}^v$, then we have $H(x^*,\hat{\mu}(k),\hat{\lambda}(k))\leq H(x^*,\mu^*,\lambda^*)=f^*$. Claim 1 and (24) renders that

$$\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)[\sum_{i=1}^{N}(H^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-f^*]$$
$$\leq\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)[\sum_{i=1}^{N}(H^{[i]}(v_x^{[i]}(\ell),v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-H^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell)))]$$
$$+\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)[\sum_{i=1}^{N}H^{[i]}(x^*,v_\mu^{[i]}(\ell),v_\lambda^{[i]}(\ell))-H(x^*,\hat{\mu}(\ell),\hat{\lambda}(\ell))]$$
$$+\limsup_{k\to+\infty}\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}(H(x^*,\hat{\mu}(\ell),\hat{\lambda}(\ell))-f^*)\leq 0$$

and thus $\limsup_{k\to+\infty}\sum_{i=1}^{N}\hat{H}^{[i]}(k)\leq f^*$. On the other hand, $\hat{x}(k)\in X$ (due to X is convex) implies $H(\hat{x}(k),\mu^*,\lambda^*)\geq H(x^*,\mu^*,\lambda^*)=f^*$. Similarly, we have the following estimates $\liminf_{k\to+\infty}\sum_{i=1}^{N}\hat{H}^{[i]}(k)\geq p^*$. Thus $\lim_{k\to+\infty}\sum_{i=1}^{N}\hat{H}^{[i]}(k)=f^*$.

Claim 3: Denote $\pi(k)=\sum_{i=1}^{N}H^{[i]}((v_x^{[i]}(k),v_\mu^{[i]}(k),v_\lambda^{[i]}(k))-H(\hat{x}(k),\hat{\mu}(k),\hat{\lambda}(k))$. And we denote the weighted version of the global penalty function H over $[0,k-1]$ as

$$\hat{H}(k)=\frac{1}{s(k-1)}\sum_{\ell=0}^{k-1}\alpha(\ell)H(\hat{x}(\ell),\hat{\mu}(\ell),\hat{\lambda}(\ell))$$

The following property holds: $\lim_{k\to+\infty}\hat{H}(k)=f^*$.

Proof: Noticing that

$$\pi(k)=\sum_{i=1}^{N}(f^{[i]}(v_x^{[i]}(k))-f^{[i]}(\hat{x}(k)))+\sum_{i=1}^{N}(v_\mu^{[i]}(k))^{\mathrm{T}}[g(v_x^{[i]}(k))]^+-v_\mu^{[i]}(k)^{\mathrm{T}}[g(\hat{x}(k))]^+)$$
$$+\sum_{i=1}^{N}((v_\mu^{[i]}(k))^{\mathrm{T}}[g(v_x^{[i]}(k))]^+-\hat{\mu}(k)^{\mathrm{T}}[g(\hat{x}(k))]^+)$$
$$+\sum_{i=1}^{N}(v_\lambda^{[i]}(k)^{\mathrm{T}}|h(v_x^{[i]}(k))|-v_\lambda^{[i]}(k)^{\mathrm{T}}|h(\hat{x}(k))|)$$
$$+\sum_{i=1}^{N}(v_\lambda^{[i]}(k)^{\mathrm{T}}|h(\hat{x}(k))|-\hat{\lambda}(k)^{\mathrm{T}}|h(\hat{x}(k))|) \tag{25}$$

By using the boundedness of subgradients and the primal estimates, we can see that

$$\| \pi(k) \| \leq \sum_{i=1}^{N} \| (f^{[i]}(v_x^{[i]}(k)) - f^{[i]}(\hat{x}(k))) \| + \sum_{i=1}^{N} \| (v_\mu^{[i]}(k))^{\mathrm{T}} [g(v_x^{[i]}(k))]^+ - v_\mu^{[i]}(k)^{\mathrm{T}} [g(\hat{x}(k))]^+) \|$$
$$+ \sum_{i=1}^{N} \| ((v_\mu^{[i]}(k))^{\mathrm{T}} [g(v_x^{[i]}(k))]^+ - \hat{\mu}(k)^{\mathrm{T}} [g(\hat{x}(k))]^+) \|$$
$$+ \sum_{i=1}^{N} \| (v_\lambda^{[i]}(k)^{\mathrm{T}} | h(v_x^{[i]}(k)) | - v_\lambda^{[i]}(k)^{\mathrm{T}} | h(\hat{x}(k)) |) \|$$
$$+ \sum_{i=1}^{N} \| (v_\lambda^{[i]}(k)^{\mathrm{T}} | h(\hat{x}(k)) | - \hat{\lambda}(k)^{\mathrm{T}} | h(\hat{x}(k)) |) \|$$

Noticing that $\| Df^{[i]}(x) \| \leq D_F (i \in V)$, $m \| D[g_\ell(x)]^+ \| \leq D_{G^+} (1 \leq \ell \leq m)$ and $v \| D | h_\ell | (x) \| \leq D_H (1 \leq \ell \leq v)$ holds for all $x \in X$, the following estimates hold:

$$\| \pi(k) \| \leq (D_F + D_G + M_\mu(k) + D_H M_\lambda(k)) \times \sum_{i=1}^{N} \| v_x^{[i]}(k) - \hat{x}(k) \|$$
$$+ G^+ \sum_{i=1}^{N} \| v_\mu^{[i]}(k) - \hat{\mu}(k) \| + H \sum_{i=1}^{N} \| v_\lambda^{[i]}(k) - \hat{\lambda}(k) \| \tag{26}$$

Then it follows from (b) in Lemma 4.2 that $\{\alpha(k) \| \pi(k) \|\}$ is summable. Notice that

$\lim\limits_{k \to +\infty} \| \hat{H}(k) - \sum_{i=1}^{N} \hat{H}^{[i]}(k) \| \leq \lim\limits_{k \to +\infty} \dfrac{\sum_{\ell=0}^{k-1} \alpha(\ell) \| \pi(\ell) \|}{s(k-1)} = 0$. Following the Claim 2, hence, $\lim\limits_{k \to +\infty} \hat{H}(k) = f^*$.

Claim 4: The limit point $\tilde{x}$ in Lemma 4.4 is a primal optimal solution.

Proof: Let $\hat{\mu}(k) = (\hat{\mu}_1(k), \cdots, \hat{\mu}_m(k))^{\mathrm{T}} \in R_{\geq 0}^m$. By $\mu^{[i]}(k+1) = v_\mu^{[i]}(k) + u_\mu^{[i]}(k)$ and $\sum_{j=1}^{N} w_{ij}(k) = 1, \sum_{i=1}^{N} w_{ij}(k) = 1$, we get

$$\sum_{i=1}^{N} \mu^{[i]}(k+1) = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(k) \mu^{[j]}(k) + \alpha(k) \sum_{i=1}^{N} [g(v_x^{[i]}(k))]^+$$
$$= \sum_{j=1}^{N} \mu^{[j]}(k) + \alpha(k) \sum_{i=1}^{N} [g(v_x^{[i]}(k))]^+ \tag{27}$$

This indicates that the sequence $\{\hat{\mu}_\ell(k)\}$ is non-decreasing in $R_{\geq 0}$. Observing that $\{\hat{\mu}_\ell(k)\}$ is lower bounded by zero. Therefore, we give the following two cases:

Case 1: The sequence $\{\hat{\mu}_\ell(k)\}$ is upper bounded. Then $\{\hat{\mu}_\ell(k)\}$ is convergent in $R_{\geq 0}$. Then it follows from Lemma 4.4 that $\lim\limits_{k \to +\infty} \| \mu^{[i]}(k) - \mu^{[j]}(k) \| = 0$ for all $i, j \in V$. This implies that there exists $\mu_\ell^* \in R_{\geq 0}$ such that $\lim\limits_{k \to +\infty} \| \mu_\ell^{[i]}(k) - \mu_\ell^* \| = 0$ for all $i \in V$. Recalling that $\sum_{i=1}^{N} \mu^{[i]}(k+1) = \sum_{i=1}^{N} \mu^{[i]}(k) + \sum_{i=1}^{N} u_\mu^{[i]}(k)$ in (27). Following a recursive step, we can get $\sum_{i=1}^{N} \mu^{[i]}(k+1) = \sum_{i=1}^{N} \mu^{[i]}(0) + \sum_{\tau=0}^{k} \alpha(\tau) \sum_{i=1}^{N} [g(v_x^{[i]}(\tau))]^+$. Since $\sum_{k=0}^{+\infty} \alpha(k) \sum_{i=1}^{N} [g_\ell(v_x^{[i]}(\tau))]^+ < +\infty$ and $\sum_{k=0}^{+\infty} \alpha(k) = +\infty$, we obtain $\lim\limits_{k \to +\infty} \inf [g_\ell(v_x^{[j]}(\tau))]^+ = 0$. Since $\lim\limits_{k \to +\infty} \| x^{[j]}(k) - \tilde{x} \| = 0$ for all $i \in V$, we have $\lim\limits_{k \to +\infty} \| \sum_{j=1}^{N} w_{ij}(k) v_x^{[i]}(k) - \sum_{j=1}^{N} w_{ij}(k) \tilde{x} \| = 0$ for all $i \in V$, then $\lim\limits_{k \to +\infty} \| v_x^{[i]}(k) - \tilde{x} \| = 0$ and thus $[g_\ell(\tilde{x})]^+ = 0$.

Case 2: The sequence $\{\hat{\mu}_\ell(k)\}$ is not upper bounded. Since $\{\hat{\mu}_\ell(k)\}$ is non-decreasing, then $\hat{\mu}_\ell(k) \to +\infty$ by $k \to +\infty$. Recalling that $\hat{H}(k) = \dfrac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) H(\hat{x}(\ell), \hat{\mu}(\ell), \hat{\lambda}(\ell))$ and $\lim\limits_{k \to +\infty} \hat{H}(k) = f^*$, then it follows from (a) in Lemma 4.1 that it is impossible that $H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) \to +\infty$. Suppose that $[g_\ell(\tilde{x})]^+ > 0$. Then we obtain

$$H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) = f(\hat{x}(k)) + N \hat{\mu}(k)^{\mathrm{T}} [g(\hat{x}(k))]^+ + N \lambda(k)^{\mathrm{T}} | h(\hat{x}(k)) |$$
$$\geq f(\hat{x}(k)) + \hat{\mu}_\ell(k) [g_\ell(\hat{x}(k))]^+ \tag{28}$$

Taking limits on both sides of (28), then we get

$$\lim\limits_{k \to +\infty} \inf H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) \geq \lim\limits_{k \to +\infty} \sup (f(\hat{x}(k)) + \hat{\mu}_\ell(k) [g_\ell(\hat{x}(k))]^+) = +\infty$$

Finally, we reach a contradiction, implying that $[g_\ell(\tilde{x})]^+ = 0$.

In both cases, we obtain $[g_\ell(\tilde{x})]^+ = 0$ for any $1 \leq \ell \leq m$. Similarly, we can further prove $| h(\tilde{x}) | = 0$. Since $\tilde{x} \in X$, then $\tilde{x}$ is feasible and thus $f(\tilde{x}) \geq f^*$. For another, since $\dfrac{\sum_{\ell=0}^{k-1} \alpha(\ell) \hat{x}(\ell)}{\sum_{\ell=0}^{k-1} \alpha(\ell)}$ is a convex combination of $\hat{x}(0), \cdots, \hat{x}(k-1)$ and

$\lim_{k \to +\infty} \hat{x}(k) = \tilde{x}$, then it follows from Claim 3 and (b) in Lemma 4.1 that:

$$f^* = \lim_{k \to +\infty} \hat{H}(k) = \lim_{k \to +\infty} \frac{\sum_{\ell=0}^{k-1} \alpha(\ell) H(\hat{x}(\ell), \hat{\mu}(\ell), \hat{\lambda}(\ell))}{\sum_{\ell=0}^{k-1} \alpha(\ell)} \geq \lim_{k \to +\infty} f\left(\frac{\sum_{\ell=0}^{k-1} \alpha(\ell) \hat{x}(\ell)}{\sum_{\ell=0}^{k-1} \alpha(\ell)}\right) = f(\tilde{x})$$

Hence, we have $f(\tilde{x}) = f^*$ and thus $\tilde{x} \in X^*$.

Lemma 4.5: It holds that $\lim_{k \to +\infty} \| y^{[i]}(k) - f^* \| = 0$.

Proof: Since $y^{[i]}(k+1) = v_y^{[i]}(k) + u^{[i]}(k)$, $v_y^{[i]}(k) = \sum_{j=1}^{N} w_{ij}(k) y^{[j]}(k)$, then the following holds for any $k \geq 1$

$$y^{[i]}(k+1) = \sum_{j=1}^{N} w_{ij}(k) y^{[j]}(k) + u^{[i]}(k)$$

The following can be proven by induction on k for a fixed $k' \geq 1$:

$$\sum_{i=1}^{N} y^{[i]}(k+1) = \sum_{i=1}^{N} y^{[i]}(k') + N \sum_{\ell=k'}^{k} \sum_{i=1}^{N} (f^{[i]}(x^{[i]}(\ell)) - f^{[i]}(x^{[i]}(\ell-1))) \tag{29}$$

Let $k' = 1$ in (29) and recall that initial state $y^{[i]}(1) = N f^{[i]}(x^{[i]}(0))$ for all $i \in V$. Then we obtain

$$\sum_{i=1}^{N} y^{[i]}(k+1) = \sum_{i=1}^{N} y^{[i]}(1) + N \sum_{i=1}^{N} (f^{[i]}(x^{[i]}(k)) - f^{[i]}(x^{[i]}(0))) = N \sum_{i=1}^{N} f^{[i]}(x^{[i]}(k)) \tag{30}$$

From (30), we can obtain

$$\sum_{i=1}^{N} (y^{[i]}(k+1) - y^{[i]}(k)) = N(\sum_{i=1}^{N} f^{[i]}(x^{[i]}(k)) - \sum_{i=1}^{N} f^{[i]}(x^{[i]}(k-1))) = u^{[i]}(k) \tag{31}$$

Combining (31) with $\lim_{k \to +\infty} \| y^{[i]}(k) - y^{[j]}(k) \| = 0$ gives the desired result. Based on the above five Lemmas, we then finish the prove of Theorem 4.1.

5. Numerical Example

In this section, we study a simple numerical example to illustrate the effectiveness of the proposed distributed penalty primal-dual subgradient algorithm. Consider a network with five agents. Suppose each agent i has a function $f^{[i]}: R \to R$, given by

$$f^{[i]}(x) = a^{[i]}x + b^{[i]}(x - c^{[i]})^2 + d^{[i]}(x - e^{[i]})^4$$

where the global decision vector $x = [x_1 \; x_2 \; x_3 \; x_4 \; x_5]^T \in R^5$. The global inequality constraint function is given by $g(x) = x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 - 10$, the global equality constraint function is give by $h(x) = x_1 + x_2 + x_3 + x_4 + x_5 - 5$ and the global constraint set is given as: $X = [-3 \; 3] \times [-3 \; 3] \times [-3 \; 3] \times [-3 \; 3] \times [-3 \; 3]$. $a^{[i]}, b^{[i]}, c^{[i]}, d^{[i]}, e^{[i]}$ are parameters of $f^{[i]}$, whose values are randomly choosen from the intervals $(-1,1), (0,1), (-1,1), (0,2), (-1,1)$. Consider the optimization problem as follows:

$$\min_{x \in R^5} \sum_{i \in V} f^{[i]}(x), \quad \text{s.t.} \quad h(x) = 0, \quad g(x) \leq 0, \quad x \in X \tag{32}$$

We solve problem (32) by employing the distributed penalty primal-dual subgradient algorithm (14) with the step-

size $\alpha(k) = 1/(k+1)$. Its simulation results are shown from Figs. 1 to 5. It can be seen from Fig. 1 that local input $u^{[i]}$ tends to 0 when it achieves consensus. Fig. 2 shows the state evolutions of all five agents, which demonstrate that all agents' takes 5×10^3 iterates to asymptotically achieve consensus. The state evolutions of dual solution μ and λ are shown in Figs. 3 and 4, respectively. We can observe from Fig. 5 that all the agents asymptotically achieve the optimal value.

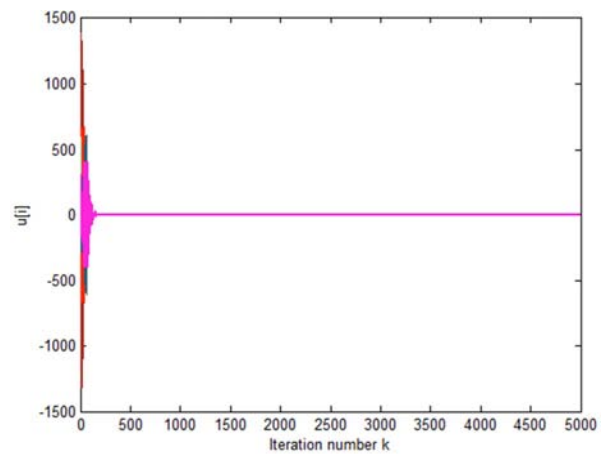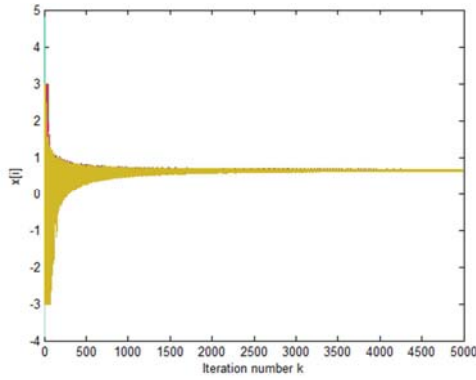**Fig. 1.** Local input $u^{[i]}$ *tends to* 0 *when achieve consensus.*
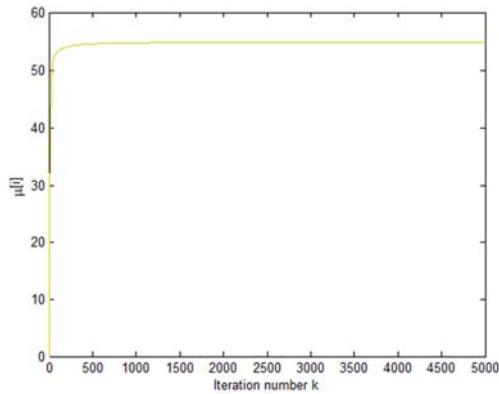
Fig. 2. *Optimal solution* x^* *of primal problem.*

Fig. 3. *Optimal solution* μ^* *of dual problem.*

Fig. 4. *Optimal solution* λ^* *of dual problem.*
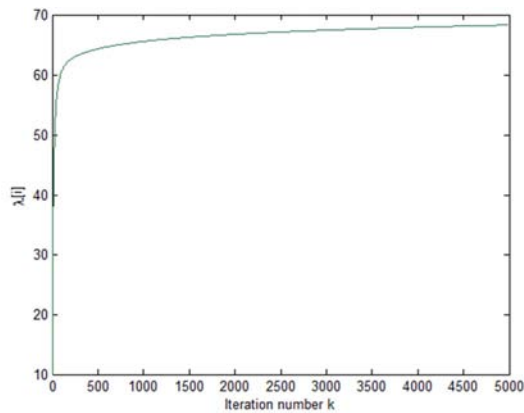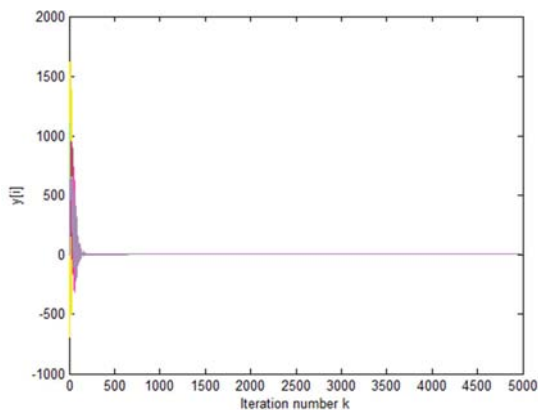

Fig. 5. *Optimal solution* f^* *of objective function* $f^{[i]}$.

6. Conclusion and Future Work

In this paper, we formulated a distributed optimization problem with local objective functions, a global equality, a global inequality and a global constraint set defined as the intersection of local constraint sets. In particular, we considered the local constraint sets to be identical. Then, we proposed a distributed penalty primal-dual subgradient algorithm for the constrained optimization with a convergence analysis. Moreover, we employed a numerical example to show that the algorithm was asymptotically converge to primal solutions and optimal values. Future work may aim at the analysis that the local constraint sets of each agent are imparities. Also, we will pay attention to the convergence rates of the algorithms in this paper.

Acknowledgements

References

[1] J. C. Duchi, A. Agarwal, and M. J. Wainwright. Dual averaging for distributed optimization: convergence analysis and network scaling, IEEE Transactions on Automatic control, 2012, 57(3): 592-606.

[2] A. Nedić, A. Ozdaglar, and P. Parrilo. Constrained consensus and optimization in multi-agent networks, IEEE Transactions on Automatic Control, 2010, 55(4): 922-938.

[3] K. Srivastava and A. Nedić. Distributed asynchronous constrained stochastic optimization, IEEE Journal of Selected Topics in Signal Processing, 2011, 5(4): 772-790.

[4] S. S. Ram, A. Nedić and V. V. Veeravalli, "Distributed Stochastic Subgradient Projection Algorithms for Convex Optimization," Journal of optimization theory and applications, 2010, 147(3): 516-545

[5] S. S. Ram, A. Nedić, and V. V. Veeravalli. Distributed stochastic subgradient projection algorithms for convex optimization, Journal of optimization theory and applications, 2010, 147(3): 516-545.

[6] A. Nedić and A. Ozdaglar. "Distributed Subgradient Methods for Multiagent Optimization", IEEE Transactions on Automatic Control, 54(1): 48-61, 2009.

[7] S. S. Ram, A. Nedić, and V. V. Veeravalli. Incremental stochastic subgradient algorithms for convex optimization, SIAM Journal on Optimization, 2009, 20(2): 691-717.

[8] J. Lu, C. Y. Tang, P. R. Regier, and et al. A gossip algorithm for convex consensus optimization over networks, American Control Conference (ACC), 2010. IEEE, 2010: 301-308.

[9] M. Zhu and S. Martínez. "An approximate dual subgradient algorithm for distributed non-convex constrained optimization", in the proc. of Conference on Decision and Control, 2010, pp. 7487-7492.

[10] M. Rabi and M. Johansson. A simple peer-to-peer algorithm for distributed optimization in sensor networks, IEEE Conference on Decision and Control. 2007, 46: 4705-4710.

[11] E. Wei, A. Ozdaglar, and A. Jadbabaie. A distributed newton method for network utility maximization, IEEE Conference on Decision and Control (CDC). 2010, 49th: 1816-1821.

[12] A. Nedić. Asynchronous broadcast-based convex optimization over a network, IEEE Transactions on Automatic Control, 2011, 56(6): 1337-1351.

[13] A. Nedić, A. Ozdaglar, and P. Parrilo. Constrained consensus and optimization in multi-agent networks, IEEE Transactions on Automatic Control, 2010, 55(4): 922-938.

[14] M. Zhu and S. Martínez. On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods, arXiv preprint arXiv:1001.2612, 2010.

[15] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson. Subgradient methods and consensus algorithms for solving convex optimization problems, In IEEE Conference on Decision and Control, pages 4185-4190, Cancun, Mexico, December 2008.

[16] A. Nedić, A. Olshevsky, A. Ozdaglar, and et al. Distributed subgradient methods and quantization effects, IEEE Conference on Decision and Control. 2008: 4177-4184.

[17] A. Nedić and A. Ozdaglar. Subgradient methods for saddle-point problems, Journal of optimization theory and applications, 2009, 142(1): 205-228.

[18] A. Nedić and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods, SIAM Journal on Optimization, 2009, 19(4): 1757-1780.

[19] D. P. Bertsekas. Convex optimization theory, Athena Scientific, 2009.

[20] D. P. Bertsekas, A. Nedić, and A. Ozdaglar. Convex analysis and optimization, Athena Scientific, 2003.

[21] M. Zhu, and S. Martínez. Discrete-time dynamic average consensus, Automatica, 2010, 46(2): 322-329.

[22] A. Nedić and A. Ozdaglar. Subgradient methods in network resource allocation: Rate analysis, Information Sciences and Systems, IEEE Conference on Decision and Control, 2008: 1189-1194.