On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods

Minghui Zhu and Sonia Martínez

Abstract

We consider a general multi-agent convex optimization problem where the agents are to collectively minimize a global objective function subject to a global inequality constraint, a global equality constraint, and a global constraint set. The objective function is defined by a sum of local objective functions, while the global constraint set is produced by the intersection of local constraint sets. In particular, we study two cases: one where the equality constraint is absent, and the other where the local constraint sets are identical. We devise two distributed primal-dual subgradient algorithms which are based on the characterization of the primal-dual optimal solutions as the saddle points of the Lagrangian and penalty functions. These algorithms can be implemented over networks with changing topologies but satisfying a standard connectivity property, and allow the agents to asymptotically agree on optimal solutions and optimal values of the optimization problem under the Slater’s condition.

I. INTRODUCTION

Recent advances in sensing, communication and computation technologies are challenging the way in which control mechanisms are designed for their efficient exploitation in a coordinated manner. This has motivated a wealth of algorithms for information processing, cooperative control, and optimization of large-scale networked multi-agent systems performing a variety of tasks. Due to a lack of a centralized authority, the proposed algorithms aim to be executed by individual agents through local actions, with the main feature of being robust to dynamic changes of network topologies.

The authors are with Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla CA, 92093, \{mizhu,soniamd\}@ucsd.edu

Preprint submitted to IEEE Transactions on Automatic Control. Received: October 26, 2009 13:02:07 PST
In this paper, we consider a general multi-agent optimization problem where the goal is to minimize a global objective function, given as a sum of local objective functions, subject to global constraints, which include an inequality constraint, an equality constraint and a (state) constraint set. Each local objective function is convex and only known to 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. This convex optimization problem arises in many practical scenarios, such as distributed parameter estimation [22] or network utility maximization [10]. An important feature of the problem is that the objective and constraint functions depend upon a global decision vector. This requires the design of distributed algorithms where, on the one hand, agents can align their decisions through a local information exchange and, on the other hand, the common decisions will coincide with an optimal solution and the optimal value.

**Literature Review.** In [1] and [23], the authors develop a general framework for parallel and distributed computation over a set of processors. Consensus problems, a class of canonical problems on networked multi-agent systems, have been intensively studied since then; e.g., see [4], [7], [13], [18], [19], [9], [20], [6], [11], [12]. Our work is also related to the literature on network utility maximization, as in [5], [10], [21].

The recent papers [15], [17] are the most relevant to our work. In [15], the authors solve a multi-agent unconstrained convex optimization problem through a novel combination of average consensus algorithms with subgradient methods. More recently, the paper [17] further takes local constraint sets into account. To deal with these constraints, the authors in [17] present an extension of their distributed subgradient algorithm, by projecting the original algorithm onto the local constraint sets. Two cases are solved in [17]: the first assumes that the network topologies can dynamically change and satisfy a periodic strong connectivity assumption (i.e., the union of the network topologies over a bounded period of time is strongly connected), but then the local constraint sets are identical; the second requires that the communication graphs are (fixed and) complete and then the local constraint sets can be different. Another related paper is [8] where a special case of [17], the network topology is fixed and all the local constraint sets are identical, is addressed.

**Statement of Contributions.** Building on the work [17], this paper further incorporates global inequality and equality constraints. More precisely, we study two cases: one in which the equality...
constraint is absent, and the other in which the local constraint sets are identical. For the first case, we adopt a Lagrangian relaxation approach, define a Lagrangian dual problem and devise the distributed Lagrangian primal-dual subgradient algorithm (DLPDS, for short) based on the characterization of the primal-dual optimal solutions as the saddle points of the Lagrangian function. The DLPDS algorithm involves each agent updating its estimates of the saddle points via a combination of an average consensus step, a subgradient (or supgradient) step and a primal (or dual) projection step onto its local constraint set (or a compact set containing the dual optimal set). The DLPDS algorithm is shown to asymptotically converge to a pair of primal-dual optimal solutions under the Slater’s condition and the periodic strong connectivity assumption. Furthermore, each agent asymptotically agrees on the optimal value by implementing a dynamic average consensus algorithm developed in [24], which allows a multi-agent system to track time-varying average values.

For the second case, to dispense with the additional equality constraint, we adopt a penalty relaxation approach, while defining a penalty dual problem and devising the distributed penalty primal-dual subgradient algorithm (DPPDS, for short). Unlike the first case, the dual optimal set of the second case may not be bounded, and thus the dual projection steps are not involved in the DPPDS algorithm. It renders that dual estimates and thus (primal) subgradients may not be uniformly bounded. This challenge is addressed by a more careful choice of step-sizes. We show that the DPPDS algorithm asymptotically converges to a primal optimal solution and the optimal value under the Slater’s condition and the periodic strong connectivity assumption.

For the special case where the global inequality and equality constraints are not taken into account, this paper extends the results in [17] to a more general scenario where the network topologies satisfy the periodic strong connectivity assumption, and the local constraint sets can be different, while relaxing an interior-point condition requirement. We refer the readers to Section VI-C for additional information.

II. PROBLEM FORMULATION AND ASSUMPTIONS

A. Problem formulation

Consider a network of agents labeled by $V := \{1, \ldots, N\}$ that can only interact with each other through local communication. The objective of the multi-agent group is to solve the following
optimization problem:
\[
\min \sum_{i=1}^{N} f_i(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in \cap_{i=1}^{N} X_i, \quad (1)
\]
where \( f_i : \mathbb{R}^n \to \mathbb{R} \) is the convex objective function of agent \( i \), \( X_i \subseteq \mathbb{R}^n \) is the compact and convex constraint set of agent \( i \), and \( x \) is a global decision vector. Assume that \( f_i \) and \( X_i \) are only known by agent \( i \), and probably different. The function \( g : \mathbb{R}^n \to \mathbb{R}^m \) is known to all the agents with each component \( g_\ell \), for \( \ell \in \{1, \ldots, m\} \), being convex. The inequality \( g(x) \leq 0 \) is understood component-wise; i.e., \( g_\ell(x) \leq 0 \), for all \( \ell \in \{1, \ldots, m\} \), and represents a global inequality constraint. The function \( h : \mathbb{R}^n \to \mathbb{R}^o \), defined as \( h(x) := Ax - b \) with \( A \in \mathbb{R}^{o \times n} \), represents a global equality constraint, and is known to all the agents. We denote \( X := \cap_{i=1}^{N} X_i \), \( f(x) := \sum_{i=1}^{N} f_i(x) \), and \( Y := \{ x \in \mathbb{R}^n \mid g(x) \leq 0, \quad h(x) = 0 \} \). We assume that the set of feasible points is non-empty; i.e., \( X \cap Y \neq \emptyset \). Since \( X \) is compact and \( Y \) is closed, then we can deduce that \( X \cap Y \) is compact. The convexity of \( f_i \) implies that of \( f \) and thus \( f \) is continuous. In this way, the optimal value \( p^* \) of the problem (1) is finite and \( X^* \), the set of primal optimal points, is non-empty. Throughout this paper, we suppose the following Slater’s condition holds:

**Assumption 2.1 (Slater’s Condition):** There exists a vector \( \bar{x} \in X \) such that \( g(\bar{x}) < 0 \) and \( h(\bar{x}) = 0 \). And there exists a relative interior point \( \tilde{x} \) of \( X^1 \) such that \( h(\tilde{x}) = 0 \).

In this paper, we will study two particular cases of problem (1): one in which the global equality constraint \( h(x) = 0 \) is not included, and the other in which all the local constraint sets are identical. For the case where the constraint \( h(x) = 0 \) is absent, the Slater’s condition 2.1 reduces to the existence of a vector \( \bar{x} \in X \) such that \( g(\bar{x}) < 0 \).

**B. Network model**

We will consider that the multi-agent network operates synchronously. The topology of the network at time \( k \geq 0 \) will be represented by a directed weighted graph \( G(k) = (V, E(k), A(k)) \) where \( A(k) := [a^i_{j}(k)] \in \mathbb{R}^{N \times N} \) is the adjacency matrix with \( a^i_{j}(k) \geq 0 \) being the weight assigned to the edge \( (j, i) \) and \( E(k) \subset V \times V \setminus \text{diag}(V) \) is the set of edges with non-zero weights \( a^i_{j}(k) \). The in-neighbors of node \( i \) at time \( k \) are denoted by \( N_i(k) = \{ j \in V \mid (j, i) \in E(k) \text{ and } j \neq i \} \).

\(^1\tilde{x} \) is a relative interior point of \( X \), if \( \tilde{x} \in X \) and there exists an open sphere \( S \) centered at \( \tilde{x} \) such that \( S \cap \text{aff}(X) \subset X \) where \( \text{aff}(X) \) is the affine hull of \( X \).
We here make the following assumptions on the network communication graphs, which are standard in the analysis of average consensus algorithms; e.g., see [18], [19], and distributed optimization in [15], [17].

**Assumption 2.2 (Non-degeneracy):** There exists a constant $\alpha > 0$ such that $a_i^j(k) \geq \alpha$, and $a_j^i(k)$, for $i \neq j$, satisfies $a_j^i(k) \in \{0\} \cup [\alpha, 1]$, for all $k \geq 0$.

**Assumption 2.3 (Balanced Communication):** It holds that $\sum_{j=1}^{N} a_j^i(k) = 1$ for all $i \in V$ and $k \geq 0$, and $\sum_{i=1}^{N} a_j^i(k) = 1$ for all $j \in V$ and $k \geq 0$.

**Assumption 2.4 (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.

### C. Notion and notations

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

**Definition 2.1 (Saddle point):** Consider a function $\phi : X \times M \to \mathbb{R}$ where $X$ and $M$ are non-empty subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$. A pair of vectors $(x^*, \mu^*) \in X \times M$ is called a saddle point of $\phi$ over $X \times M$ if $\phi(x^*, \mu) \leq \phi(x^*, \mu^*) \leq \phi(x, \mu^*)$ hold for all $(x, \mu) \in X \times M$.

**Remark 2.1:** Equivalently, $(x^*, \mu^*)$ is a saddle point of $\phi$ over $X \times M$ if and only if $(x^*, \mu^*) \in X \times M$, and $\sup_{\mu \in M} \phi(x^*, \mu) \leq \phi(x^*, \mu^*) \leq \inf_{x \in X} \phi(x, \mu^*)$.

In this paper, we do not assume the differentiability of $f_i$ and $g_i$. At the points where the function is not differentiable, the subgradient plays the role of the gradient. For a given convex function $F : \mathbb{R}^n \to \mathbb{R}$ and a point $\bar{x} \in \mathbb{R}^n$, a subgradient of the function $F$ at $\bar{x}$ is a vector $D F(\bar{x}) \in \mathbb{R}^n$ such that $D F(\bar{x})^T (x - \bar{x}) \leq F(x) - F(\bar{x})$, $\forall x \in \mathbb{R}^n$. Similarly, for a given concave function $G : \mathbb{R}^m \to \mathbb{R}$ and a point $\bar{\mu} \in \mathbb{R}^m$, a supgradient of the function $G$ at $\bar{\mu}$ is a vector $D G(\bar{\mu}) \in \mathbb{R}^m$ such that $D G(\bar{\mu})^T (\mu - \bar{\mu}) \geq G(\mu) - G(\bar{\mu})$, $\forall \mu \in \mathbb{R}^m$.

Given a set $S$, we denote by $\co(S)$ its convex hull. We let the function $[\cdot]^+ : \mathbb{R}^m \to \mathbb{R}_{\geq 0}^m$ denote the projection operator onto the non-negative orthant in $\mathbb{R}^m$. For any vector $c \in \mathbb{R}^n$, we denote $|c| := (|c_1|, \cdots, |c_n|)^T$, while $\| \cdot \|$ is the 2-norm in the Euclidean space.

$^2$ It is also referred to as double stochasticity.
III. CASE (I): ABSENCE OF EQUALITY CONSTRAINT

In this section, we study the problem (1) in the absence of the equality constraint; i.e.,
\[
\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad x \in \cap_{i=1}^{N} X_i. \tag{2}
\]

We first provide some preliminaries, including a Lagrangian saddle-point characterization of the problem (2) and finding a superset containing the Lagrangian dual optimal set of the problem (2). After that, we present the distributed Lagrangian primal-dual subgradient algorithm and summarize its convergence properties.

A. Preliminaries

1) A Lagrangian saddle-point characterization: Firstly, the problem (2) is equivalent to
\[
\min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \quad Ng(x) \leq 0, \quad x \in X,
\]
with associated Lagrangian dual problem given by
\[
\max_{\mu \in \mathbb{R}^m} q_L(\mu), \quad \text{s.t.} \quad \mu \geq 0.
\]

Here, the Lagrangian dual function, \( q_L : \mathbb{R}^m_{\geq 0} \rightarrow \mathbb{R} \), is defined as \( q_L(\mu) := \inf_{x \in X} \mathcal{L}(x, \mu) \), where \( \mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \rightarrow \mathbb{R} \) is the Lagrangian function \( \mathcal{L}(x, \mu) = f(x) + N\mu^T g(x) \). We denote the Lagrangian dual optimal value of the Lagrangian dual problem by \( d^*_L \) and the set of Lagrangian dual optimal points by \( D^*_L \). As is well-known, under the Slater’s condition 2.1, the property of strong duality holds; i.e., \( p^* = d^*_L \), and \( D^*_L \neq \emptyset \). The following theorem is a standard result on Lagrangian duality stating that the primal and Lagrangian dual optimal solutions can be characterized as the saddle points of the Lagrangian function.

**Theorem 3.1 (Lagrangian Saddle-point Theorem [2]):** The pair of \( (x^*, \mu^*) \in X \times \mathbb{R}^m_{\geq 0} \) is a saddle point of the Lagrangian function \( \mathcal{L} \) over \( X \times \mathbb{R}^m_{\geq 0} \) if and only if it is a pair of primal and Lagrangian dual optimal solutions and the following Lagrangian minmax equality holds:
\[
\sup_{\mu \in \mathbb{R}^m_{\geq 0}} \inf_{x \in X} \mathcal{L}(x, \mu) = \inf_{x \in X} \sup_{\mu \in \mathbb{R}^m_{\geq 0}} \mathcal{L}(x, \mu).
\]

**Lemma 3.1:** Let \( M \) be any superset of \( D^*_L \).

(a) If \( (x^*, \mu^*) \) is a saddle point of \( \mathcal{L} \) over \( X \times \mathbb{R}^m_{\geq 0} \), then \( (x^*, \mu^*) \) is also a saddle point of \( \mathcal{L} \) over \( X \times M \).
(b) There is at least one saddle point of $L$ over $X \times M$.
(c) If $(\hat{x}, \hat{\mu})$ is a saddle point of $L$ over $X \times M$, then $L(\hat{x}, \hat{\mu}) = p^*$ and $\hat{\mu}$ is Lagrangian dual optimal.

**Proof:** Due to the space limit, the proofs are omitted here and provided in the enlarged version [25].

**Remark 3.1:** Despite that (c) holds, the reverse of (a) may not be true in general. In particular, $x^*$ may be infeasible; i.e., $g_\ell(x^*) > 0$ for some $\ell \in \{1, \ldots, m\}$.

2) A upper estimate of the Lagrangian dual optimal set: In what follows, we will find a compact superset of $D_L^*$. To do that, we define the following primal problem for each agent $i$:

$$
\min_{x \in \mathbb{R}^n} f_i(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad x \in X_i.
$$

Due to the fact that $X_i$ is compact and the $f_i$ are continuous, the primal optimal value $p_i^*$ of each agent’s primal problem is finite and the set of its primal optimal solutions is non-empty. The associated dual problem is given by

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

Here, the dual function $q_i : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ is defined by $q_i(\mu) := \inf_{x \in X_i} L_i(x, \mu)$, where $L_i : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ is the Lagrangian function $L_i(x, \mu) = f_i(x) + \mu^T g(x)$, $i \in V$. The corresponding dual optimal values are denoted by $d_i^*, i \in V$. In this way, we have that $L(x, \mu) = \sum_{i=1}^N L_i(x, \mu)$.

Define now the set-valued map $Q : \mathbb{R}^m_{\geq 0} \to 2^{\mathbb{R}^m_{\geq 0}}$ by $Q(\tilde{\mu}) = \{\mu \in \mathbb{R}^m_{\geq 0} | q_i(\mu) \geq q_i(\tilde{\mu})\}$. Additionally, define a function $\gamma : X \to \mathbb{R}$ by $\gamma(x) = \min_{i \in \{1, \ldots, m\}} \{-g_\ell(x)\}$. Observe that if $x$ is a Slater vector, then $\gamma(x) > 0$. The following lemma is a direct result of Lemma 1 in [14].

**Lemma 3.2:** The set $Q(\tilde{\mu})$ is bounded for any $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$, and, in particular, for any Slater vector $\bar{x}$, it holds that $\max_{\mu \in Q(\tilde{\mu})} \|\mu\| \leq \frac{1}{\gamma(\bar{x})} (f(\bar{x}) - q_\ell(\bar{x}))$. \(\square\)

Notice that $D_L^* = \{\mu \in \mathbb{R}^m_{\geq 0} | q_i(\mu) \geq d_i^*\}$. Picking any Slater vector $\bar{x} \in X$, and letting $\bar{\mu} = \mu^* \in D_L^*$ in Lemma 3.2 gives that

$$
\max_{\mu^* \in D_L^*} \|\mu^*\| \leq \frac{1}{\gamma(\bar{x})} (f(\bar{x}) - d_L^*).
$$

(3)

Define the function $r : X \times \mathbb{R}^m_{\geq 0} \to \mathbb{R} \cup \{+\infty\}$ by $r(x, \mu) := \frac{N}{\gamma(x)} \max_{i \in V} \{f_i(x) - q_i(\mu)\}$. By the property of weak duality, it holds that $d_i^* \leq p_i^*$ and thus $f_i(x) \geq q_i(\mu)$ for any $(x, \mu) \in X \times \mathbb{R}^m_{\geq 0}$. Since $\gamma(\bar{x}) > 0$, thus $r(\bar{x}, \mu) \geq 0$ for any $\mu \in \mathbb{R}^m_{\geq 0}$. With this observation, we pick any $\bar{\mu} \in \mathbb{R}^m_{\geq 0}$
and the following set is well-defined: \( \hat{M}_i(\bar{x}, \hat{\mu}) := \{ \mu \in \mathbb{R}^m_{\geq 0} | \|\mu\| \leq r(\bar{x}, \hat{\mu}) + \theta_i \} \) for some \( \theta_i \in \mathbb{R}_{>0} \). Observe that for all \( \mu \in \mathbb{R}^m_{\geq 0} \):

\[
q_L(\mu) = \inf_{x \in \cap_{i=1}^N X_i} \sum_{i=1}^N (f_i(x) + \mu^T g(x)) \geq \sum_{i=1}^N \inf_{x \in X_i} (f_i(x) + \mu^T g(x)) = \sum_{i=1}^N q_i(\mu). \tag{4}
\]

Since \( d^*_L \geq q_L(\hat{\mu}) \), it follows from (3) and (4) that

\[
\max_{\mu^* \in D^*_L} \|\mu^*\| \leq \frac{1}{\gamma(\bar{x})} (f(\bar{x}) - q_L(\hat{\mu})) \leq \frac{1}{\gamma(\bar{x})} (f(\bar{x}) - \sum_{i=1}^N q_i(\hat{\mu})) \leq \frac{N}{\gamma(\bar{x})} \max_{i \in V} \{f_i(\bar{x}) - q_i(\hat{\mu})\} = r(\bar{x}, \hat{\mu}).
\]

Hence, we have \( D^*_L \subseteq \hat{M}_i(\bar{x}, \hat{\mu}) \) for all \( i \in V \).

Note that in order to compute \( \hat{M}_i(\bar{x}, \hat{\mu}) \), all the agents have to agree on a common Slater vector \( \bar{x} \in \cap_{i=1}^N X_i \) which should be obtained in a distributed fashion. To handle this difficulty, we now propose a distributed algorithm, namely Distributed Slater-vector Computation Algorithm, which allows each agent \( i \) to compute a superset of \( \hat{M}_i(\bar{x}, \hat{\mu}) \).

Initially, each agent \( i \) chooses a common value \( \hat{\mu} \in \mathbb{R}^m_{\geq 0} \), e.g., \( \hat{\mu} = 0 \), and computes two positive constants \( b_i(0) \) and \( c_i(0) \) such that \( b_i(0) \geq \sup_{x \in J_i} \{ f_i(x) - q_i(\hat{\mu}) \} \) and \( c_i(0) \leq \min_{1 \leq \ell \leq m} \inf_{x \in J_i} \{-g_\ell(x)\} \) where \( J_i := \{ x \in X_i | g(x) < 0 \} \). At every time \( k \geq 0 \), each agent \( i \) updates its estimates by using the following rules:

\[
b_i(k+1) = \max_{j \in N_i(k) \cup \{i\}} b_j(k), \quad c_i(k+1) = \min_{j \in N_i(k) \cup \{i\}} c_j(k).
\]

Under the periodical strong connectivity assumption 2.4, it is not difficult to verify that after at most \((N-1)B\) steps, all the agents reach the consensus, i.e., \( b_i(k) = b^* := \max_{j \in V} b_j(0) \) and \( c_i(k) = c^* := \min_{j \in V} c_j(0) \) for all \( k \geq (N-1)B \). Denote \( \hat{M}_i(\hat{\mu}) := \{ \mu \in \mathbb{R}^m_{\geq 0} | \|\mu\| \leq \frac{N b^*}{c^*} + \theta_i \} \) and \( J := \{ x \in X | g(x) < 0 \} \).

**Lemma 3.3:** It holds that \( \hat{M}_i(\hat{\mu}) \supseteq \hat{M}_i(\bar{x}, \hat{\mu}) \) for \( i \in V \).

**Proof:** Note that \( J \subseteq J_i, \forall i \in V \). Hence, we have

\[
\max_{i \in V} \sup_{x \in J} \{ f_i(x) - q_i(\hat{\mu}) \} \leq \max_{i \in V} \sup_{x \in J_i} \{ f_i(x) - q_i(\hat{\mu}) \} \leq b^*, \]

\[
\inf_{x \in J} \min_{1 \leq \ell \leq m} \{-g_\ell(x)\} \geq \min_{i \in V} \inf_{x \in J_i} \min_{1 \leq \ell \leq m} \{-g_\ell(x)\} \geq c^*.
\]

Since \( \bar{x} \in J \), then the following estimate on \( r(\bar{x}, \hat{\mu}) \) holds:

\[
r(\bar{x}, \hat{\mu}) \leq \frac{N \sup_{x \in J} \max_{i \in V} \{ f_i(x) - q_i(\hat{\mu}) \}}{\inf_{x \in J} \min_{1 \leq \ell \leq m} \{-g_\ell(x)\}} \leq \frac{N b^*}{c^*}.
\]
The desired result immediately follows. From Lemma 3.3 and the fact that $D^* \subseteq \bar{M}_i(\bar{x}, \tilde{\mu})$, we can see that the set of $M(\tilde{\mu}) := \cap_{i=1}^N M_i(\tilde{\mu})$ contains $D^*$. In addition, $M_i(\tilde{\mu})$ and $M(\tilde{\mu})$ are non-empty, compact and convex. To simplify the notations, we will use the shorthands $M_i := M_i(\tilde{\mu})$ and $M := M(\tilde{\mu})$.

3) Convexity of $L$: For each $\mu \in \mathbb{R}^m_{\geq 0}$, we define the function $L_{i\mu} : \mathbb{R}^n \to \mathbb{R}$ as $L_{i\mu}(x) := L_i(x, \mu)$. Note that $L_{i\mu}$ is convex since it is a nonnegative weighted sum of convex functions. For each $x \in \mathbb{R}^n$, we define the function $L_{ix} : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ as $L_{ix}(\mu) := L_i(x, \mu)$. It is easy to check that $L_{ix}$ is a concave (actually affine) function. Then the Lagrangian function $L$ is the sum of a collection of convex-concave local functions. This property motivates us to significantly extend primal-dual subgradient methods in [16] to the networked multi-agent scenario.

B. Distributed Lagrangian primal-dual subgradient algorithm

Here, we introduce the Distributed Lagrangian Primal-Dual Subgradient Algorithm (DLPDS, for short) to find a saddle point of the Lagrangian function $L$ over $X \times M$ and the optimal value. This saddle point will coincide with a pair of primal and Lagrangian dual optimal solutions which is not always the case; see Remark 3.1.

Through the algorithm, at each time $k$, each agent $i$ maintains the estimate of $(x^i(k), \mu^i(k))$ to the saddle point of the Lagrangian function $L$ over $X \times M$ and the estimate of $y^i(k)$ to $p^*$. To produce $x^i(k+1)$ (resp. $\mu^i(k+1)$), agent $i$ takes a convex combination of its estimate $x^i(k)$ (resp. $\mu^i(k)$) with the estimates sent from its neighboring agents at time $k$, makes a subgradient (resp. supgradient) step to minimize (resp. maximize) the local Lagrangian function $L_i$, and takes a primal (resp. dual) projection onto the local constraint $X_i$ (resp. $M_i$). Furthermore, agent $i$ generates the estimate $y^i(k+1)$ by taking a convex combination of its estimate $y^i(k)$ with the estimates of its neighbors at time $k$ and taking one step to track the variation of the local objective function $f_i$. More precisely, the DLPDS algorithm is described as follows:

Initially, each agent $i$ picks a common $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$ and computes the set $M_i$ with some $\theta_i > 0$ by using the Distributed Slater-vector Computation Algorithm. Furthermore, agent $i$ chooses any initial state $x^i(0) \in X_i$, $\mu^i(0) \in \mathbb{R}^m_{\geq 0}$, and $y^i(1) = N f_i(x^i(0))$.

At every $k \geq 0$, each agent $i$ generates $x^i(k+1)$, $\mu^i(k+1)$ and $y^i(k+1)$ according to the
following rules:
\[
\begin{align*}
v^i_x(k) &= \sum_{j=1}^{N} a^{i,j}(k)x^j(k), \quad v^i_\mu(k) = \sum_{j=1}^{N} a^{i,j}(k)\mu^j(k), \quad v^i_y(k) = \sum_{j=1}^{N} a^{i,j}(k)y^j(k),
\end{align*}
\]
\[
\begin{align*}
x^i(k+1) &= P_{X_i}[v^i_x(k) - \alpha(k)D^i_x(k)], \quad \mu^i(k+1) = P_{M_i}[v^i_\mu(k) + \alpha(k)D^i_\mu(k)], \\
y^i(k+1) &= v^i_y(k) + N(f_i(x^i(k)) - f_i(x^i(k-1))),
\end{align*}
\]
where \( P_{X_i} \) (resp. \( P_{M_i} \)) is the projection operator onto the set \( X_i \) (resp. \( M_i \)), the scalars \( a^{i,j}(k) \) are non-negative weights and the scalars \( \alpha(k) > 0 \) are step-sizes\(^3\). We use the shorthands \( D^i_x(k) \equiv D^i_{lv^i_x(k)}(v^i_x(k)) \), and \( D^i_\mu(k) \equiv D^i_{lv^i_\mu(k)}(v^i_\mu(k)) \).

**Remark 3.2:** The DLPDS algorithm is a generalization of primal-dual subgradient methods in [16] to the networked multi-agent scenario. It is also an extension of the distributed projected subgradient algorithm in [17] to solve multi-agent convex optimization problems with inequality constraints. Additionally, the DLPDS algorithm enables agents to find the optimal value. Furthermore, the DLPDS algorithm objective is that of reaching a saddle point of the Lagrangian function in contrast to achieving a (primal) optimal solution in [17].

The following theorem summarizes the convergence properties of the DLPDS algorithm.

**Theorem 3.2:** Consider the optimization problem (2). Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumptions 2.4 hold. Consider the sequences of \( \{x^i(k)\} \), \( \{\mu^i(k)\} \) and \( \{y^i(k)\} \) of the distributed Lagrangian primal-dual subgradient algorithm with the step-sizes \( \{\alpha(k)\} \) satisfying \( \lim_{k \to +\infty} \alpha(k) = 0 \), \( \sum_{k=0}^{+\infty} \alpha(k) = +\infty \), and \( \sum_{k=0}^{+\infty} \alpha(k)^2 < +\infty \). Then, there is a pair of primal and Lagrangian dual optimal solutions \( (x^*, \mu^*) \in X^* \times D^* \) such that \( \lim_{k \to +\infty} \|x^i(k) - x^*\| = 0 \) and \( \lim_{k \to +\infty} \|\mu^i(k) - \mu^*\| = 0 \) for all \( i \in V \). Furthermore, we have that \( \lim_{k \to +\infty} \|y^i(k) - y^*\| = 0 \) for all \( i \in V \).

**IV. CASE (II): IDENTICAL LOCAL CONSTRAINT SETS**

In this section, we consider the primal problem (1) with identical local constraint sets; i.e., \( X_i = X \) for all \( i \in V \). We first adopt a penalty relaxation and provide a penalty saddle-point characterization of the primal problem (1) with \( X_i = X \). We then introduce the distributed penalty primal-dual subgradient algorithm, followed by its convergence properties.

\(^3\)Each agent \( i \) executes the update law of \( y^i(k) \) for \( k \geq 1 \).
A. Preliminaries

1) A penalty saddle-point characterization: Note that the primal problem (1) with $X_i = X$ is trivially equivalent to the following:

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

with associated penalty dual problem given by

$$\max_{\mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^\nu} q_P(\mu, \lambda), \quad \text{s.t.} \quad \mu \geq 0, \quad \lambda \geq 0. \quad (6)$$

Here, the penalty dual function, $q_P : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$, is defined by $q_P(\mu, \lambda) := \inf_{x \in X} \mathcal{H}(x, \mu, \lambda)$, where $\mathcal{H} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ is the penalty function given by $\mathcal{H}(x, \mu, \lambda) = f(x) + N\mu^T[g(x)]^+ + N\lambda^T|h(x)|$. We denote the penalty dual optimal value by $d_P^*$ and the set of penalty dual optimal solutions by $D_P^*$. We define the penalty function $\mathcal{H}_i(x, \mu, \lambda) : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ for each agent $i$ as follows: $\mathcal{H}_i(x, \mu, \lambda) = f_i(x) + \mu^T[g(x)]^+ + \lambda^T|h(x)|$. In this way, we have that $\mathcal{H}(x, \mu, \lambda) = \sum_{i=1}^N \mathcal{H}_i(x, \mu, \lambda)$. As proven in the next lemma, the Slater’s condition 2.1 ensures zero duality gap and the existence of penalty dual optimal solutions.

**Lemma 4.1:** The values of $p^*$ and $d_P^*$ coincide, and $D_P^*$ is non-empty.

**Proof:** Consider the auxiliary Lagrangian function $\mathcal{L}_a : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu \to \mathbb{R}$ given by

$$\mathcal{L}_a(x, \mu, \lambda) = f(x) + N\mu^Tg(x) + N\lambda^T|h(x)|,$$

with the associated dual problem defined by

$$\max_{\mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^\nu} q_a(\mu, \lambda), \quad \text{s.t.} \quad \mu \geq 0. \quad (7)$$

Here, the dual function, $q_a : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu \to \mathbb{R}$, is defined by $q_a(\mu, \lambda) := \inf_{x \in X} \mathcal{L}_a(x, \mu, \lambda)$. The dual optimal value of problem (7) is denoted by $d_a^*$ and the set of dual optimal solutions is denoted $D_a^*$. Since $X$ is convex, $f$ and $g_\ell$, for $\ell \in \{1, \ldots, m\}$, are convex, $p^*$ is finite and the Slater’s condition 2.1 holds, it follows from Proposition 5.3.5 in [2] that $p^* = d_a^*$ and $D_a^* \neq \emptyset$.

We now proceed to characterize $d_P^*$ and $D_P^*$. Pick any $(\mu^*, \lambda^*) \in D_a^*$. Since $\mu^* \geq 0$, then

$$d_a^* = q_a(\mu^*, \lambda^*) = \inf_{x \in X} \{f(x) + N(\mu^*)^Tg(x) + N(\lambda^*)^T|h(x)|\}$$

$$\leq \inf_{x \in X} \{f(x) + N(\mu^*)^T[g(x)]^+ + N(\lambda^*)^T|h(x)|\} = q_P(\mu^*, |\lambda^*|) \leq d_P^*. \quad (8)$$

On the other hand, pick any $x^* \in X^*$. Then $x^*$ is feasible, i.e., $x^* \in X$, $[g(x^*)]^+ = 0$ and $|h(x^*)| = 0$. It implies that $q_P(\mu, \lambda) \leq \mathcal{H}(x^*, \mu, \lambda) = f(x^*) = p^*$ holds for any $\mu \in \mathbb{R}^m_{\geq 0}$ and $\lambda \in \mathbb{R}^\nu_{\geq 0}$, and thus $d_P^* = \sup_{\mu \in \mathbb{R}^m_{\geq 0}, \lambda \in \mathbb{R}^\nu_{\geq 0}} q_P(\mu, \lambda) \leq p^* = d_a^*$. Therefore, we have $d_P^* = p^*$. 
To prove the emptiness of $D^*_P$, we pick any $(\mu^*, \lambda^*) \in D^*_P$. From (8) and $d^*_a = d^*_P$, we can see that $(\mu^*, |\lambda^*|) \in D^*_P$ and thus $D^*_P \neq \emptyset$. ■

The following is a slight extension of Theorem 3.1 to penalty functions.

**Theorem 4.1 (Penalty Saddle-point Theorem):** The pair of $(x^*, \mu^*, \lambda^*)$ is a saddle point of the penalty function $\mathcal{H}$ over $X \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p_{\geq 0}$ if and only if it is a pair of primal and penalty dual optimal solutions and the following penalty minmax equality holds: $\sup_{(\mu, \lambda) \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p_{\geq 0}} \inf_{x \in X} \mathcal{H}(x, \mu, \lambda) = \sup_{(\mu, \lambda) \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p_{\geq 0}} \inf_{x \in X} \mathcal{H}(x, \mu, \lambda)$.

**Proof:** The proof is analogous to that of Proposition 6.2.4 in [3], and thus omitted here. For the sake of completeness, we provide the details in the enlarged version [25]. ■

2) Convexity of $\mathcal{H}$: Since $g_\ell$ is convex and $[\cdot]^+$ is convex and non-decreasing, thus $[g_\ell(x)]^+$ is convex in $x$ for each $\ell \in \{1, \ldots, m\}$. Denote $A := (a_1^T, \ldots, a_p^T)^T$. Since $|\cdot|$ is convex and $a_\ell^T x - b_\ell$ is an affine mapping, then $|a_\ell^T x - b_\ell|$ is convex in $x$ for each $\ell \in \{1, \ldots, p\}$.

We denote $w := (\mu, \lambda)$. For each $w \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p_{\geq 0}$, we define the function $\mathcal{H}_{iw} : \mathbb{R}^n \to \mathbb{R}$ as $\mathcal{H}_{iw}(x) := \mathcal{H}(x, w)$. Note that $\mathcal{H}_{iw}(x)$ is convex in $x$ by using the fact that a nonnegative weighted sum of convex functions is convex. For each $x \in \mathbb{R}^n$, we define the function $\mathcal{H}_{ix} : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p_{\geq 0} \to \mathbb{R}$ as $\mathcal{H}_{ix}(w) := \mathcal{H}(x, w)$. It is easy to check that $\mathcal{H}_{ix}(w)$ is concave (actually affine in $w$). Then the penalty function $\mathcal{H}(x, w)$ is the sum of convex-concave local functions.

**Remark 4.1:** The Lagrangian relaxation does not fit to our approach here since the Lagrangian function is not convex in $x$ by allowing $\lambda$ entries to be negative.

**B. Distributed penalty primal-dual subgradient algorithm**

We now devise the Distributed Penalty Primal-Dual Subgradient Algorithm (DPPDS, for short), that is based on the penalty saddle-point theorem 4.1, to find the optimal value and a primal optimal solution to the primal problem (1) with $X_i = X$. The main steps of the DPPDS algorithm are described as follow.

Initially, agent $i$ chooses any initial state $x^i(0) \in X$, $\mu^i(0) \in \mathbb{R}^m_{\geq 0}$, $\lambda^i(0) \in \mathbb{R}^p_{\geq 0}$, and $y^i(1) = Nf_i(x^i(0))$. At every time $k \geq 0$, each agent $i$ computes the following quantities:

$$
\begin{align*}
v^i_x(k) & = \sum_{j=1}^N a^i_j(k)x^j(k), & v^i_y(k) & = \sum_{j=1}^N a^i_j(k)y^j(k), \\
v^i_\mu(k) & = \sum_{j=1}^N a^i_j(k)\mu^j(k), & v^i_\lambda(k) & = \sum_{j=1}^N a^i_j(k)\lambda^j(k),
\end{align*}
$$
and generates $x^i(k+1)$, $y^i(k+1)$, $\mu^i(k+1)$ and $\lambda^i(k+1)$ according to the following rules:

$$
x^i(k+1) = P_X[v^i_x(k) - \alpha(k)S^i_x(k)], \quad y^i(k+1) = v^i_y(k) + N(f_i(x^i(k)) - f_i(x^i(k-1))),
\mu^i(k+1) = v^i_\mu(k) + \alpha(k)[g(v^i_x(k))], \quad \lambda^i(k+1) = v^i_\lambda(k) + \alpha(k)[h(v^i_x(k))],
$$

where $P_X$ is the projection operator onto the set $X$, the scalars $a^i_j(k)$ are non-negative weights and the positive scalars $\{\alpha(k)\}$ are step-sizes\(^4\). The vector $S^i_x(k) := Df_i(v^i_x(k)) + \sum_{\ell=1}^{m} v^i_\mu(k)D[g(\ell v^i_x(k))]+ \sum_{\ell=1}^{n} v^i_\lambda(k)D[h_\ell(v^i_x(k))]$ is a subgradient of $\mathcal{H}_{iw^i(k)}(x)$ at $x = v^i_x(k)$ where $w^i(k) := (v^i_\mu(k), v^i_\lambda(k))$.

**Remark 4.2:** As the primal-dual subgradient algorithm in [16], the DPPDS algorithm produces a pair of primal and dual estimates at each step. Main differences include: firstly, the DPPDS algorithm extends the primal-dual subgradient algorithm in [16] to the multi-agent scenario; secondly, it further takes the equality constraint into account. The presence of the equality constraint can make $D^*_p$ unbounded. Therefore, unlike the DLPDS algorithm, the DPPDS algorithm does not involve the dual projection steps onto compact sets. This may cause the subgradient $S^i_x(k)$ not to be uniformly bounded, while the boundedness of subgradients is a standard assumption in the analysis of subgradient methods, e.g., see [2], [3], [14], [15], [16], [17]. This difficulty will be addressed by a more careful choice of the step-size policy; i.e., assumption 4.1, which is stronger than the more standard diminishing step-size scheme, e.g., in the DLPDS algorithm and [17]. We require this condition in order to prove, in the absence of the boundedness of $\{S^i_x(k)\}$, the existence of a number of limits and summability of expansions toward Theorem 4.2. Thirdly, the DPPDS algorithm adopts the penalty relaxation instead of the Lagrangian relaxation in [16].

**Remark 4.3:** Observe that $\mu^i(k) \geq 0$ and $\lambda^i(k) \geq 0$. Furthermore, $([g(v^i_x(k))], |h(v^i_x(k))|)$ is a subgradient of $\mathcal{H}_{iw^i(k)}(w^i(k))$; i.e. the following penalty supgradient inequality holds for any $\mu \in \mathbb{R}_{\geq 0}^m$ and $\lambda \in \mathbb{R}_{\geq 0}^n$:

$$
([g(v^i_x(k))]+)^T(\mu - v^i_\mu(k)) + |h(v^i_x(k))|^T(\lambda - v^i_\lambda(k)) \geq \mathcal{H}_i(v^i_x(k), \mu, \lambda) - \mathcal{H}_i(v^i_x(k), v^i_\mu(k), v^i_\lambda(k)).
$$

Given a step-size sequence $\{\alpha(k)\}$, we define $s(k) := \sum_{\ell=0}^{k} \alpha(\ell)$ and assume that:

\(^4\)Each agent $i$ executes the update law of $y^i(k)$ for $k \geq 1$.}

Limited circulation. For review only

DRAFT

Preprint submitted to IEEE Transactions on Automatic Control. Received: October 26, 2009 13:02:07 PST
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 \( \lim_{k \to +\infty} \alpha(k+1)s(k) = 0 \), \( \sum_{k=0}^{+\infty} \alpha(k+1)^2s(k) < +\infty \), \( \sum_{k=0}^{+\infty} \alpha(k+1)^2s(k)^2 < +\infty \).

The following theorem is the main result of this section, characterizing the convergence of the DPPDS algorithm.

Theorem 4.2: Consider the problem (1) with \( X_i = X \). Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 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 \( \lim_{k \to +\infty} \|y^i(k) - p^*\| = 0 \) for all \( i \in V \).

Remark 4.4: Note that \( v^i_x(k) \in X \) (due to the fact that \( X \) is convex). A step-size sequence that satisfies the step-size assumption 4.1 is the harmonic series \( \{\alpha(k) = \frac{1}{k+1}\}_{k \in \mathbb{Z} \geq 0} \). The analysis can be found in the enlarged version [25].

V. CONVERGENCE ANALYSIS

We next provide the proofs of Theorem 3.2 and 4.2. We start our analysis by providing some useful properties of the sequences weighted by \( \{\alpha(k)\} \).

Lemma 5.1: Let \( K \geq 0 \). Consider the sequence \( \{\delta(k)\} \) defined by \( \delta(k) := \frac{\sum_{\ell=K}^{k-1} \alpha(\ell)\rho(\ell)}{\sum_{\ell=K}^{+\infty} \alpha(\ell)} \).

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 \( \lim_{k \to +\infty} \delta(k) = \rho^* \).

Proof: (a) For any \( \Pi > 0 \), there exists \( k_1 \geq K \) such that \( \rho(k) \geq \Pi \) for all \( k \geq k_1 \). Then the following holds for all \( k \geq k_1 + 1 \):

\[
\delta(k) \geq \frac{1}{\sum_{\ell=K}^{k-1} \alpha(\ell)} \left( \sum_{\ell=K}^{k-1} \alpha(\ell)\rho(\ell) + \sum_{\ell=K}^{k-1} \alpha(\ell)\Pi \right) = \Pi + \frac{1}{\sum_{\ell=K}^{k-1} \alpha(\ell)} \left( \sum_{\ell=K}^{k-1} \alpha(\ell)\rho(\ell) - \sum_{\ell=K}^{k-1} \alpha(\ell)\Pi \right).
\]

Take the limit on \( k \) in the above estimate and we have \( \liminf_{k \to +\infty} \delta(k) \geq \Pi \). Since \( \Pi \) is arbitrary, then \( \lim_{k \to +\infty} \delta(k) = +\infty \).
(b) For any $\epsilon > 0$, there exists $k_2 \geq K$ such that $\|\rho(k) - \rho^*\| \leq \epsilon$ for all $k \geq k_2 + 1$. Then we have
\[
\|\delta(k) - \rho^*\| = \| \sum_{\tau = k_2}^{k - 1} \alpha(\tau) (\rho(\tau) - \rho^*) \| \leq \sum_{\tau = K}^{k_2 - 1} \alpha(\tau) \|\rho(\tau) - \rho^*\| + \epsilon.
\]
Take the limit on $k$ in the above estimate and we have $\lim_{k \to +\infty} \|\delta(k) - \rho^*\| \leq \epsilon$. Since $\epsilon$ is arbitrary, then $\lim_{k \to +\infty} \|\delta(k) - \rho^*\| = 0$.

A. Proofs of Theorem 3.2

We now proceed to show Theorem 3.2. To do that, we first rewrite the DLPDS algorithm into the following form:
\[
x^i(k + 1) = v^i_x(k) + e^i_x(k), \quad \mu^i(k + 1) = v^i_\mu(k) + e^i_\mu(k), \quad y^i(k + 1) = v^i_y(k) + u^i(k),
\]
where $e^i_x(k)$ and $e^i_\mu(k)$ are projection errors described by
\[
e^i_x(k) := P_{X_i}[v^i_x(k) - \alpha(k)D^i_x(k)] - v^i_x(k), \quad e^i_\mu(k) := P_{\mu}[v^i_\mu(k) + \alpha(k)D^i_\mu(k)] - v^i_\mu(k),
\]
and $u^i(k) := N(f_i(x^i(k)) - f_i(x^i(k - 1)))$ is the local input which allows agent $i$ to track the variation of the local objective function $f_i$. In this manner, the update law of each estimate is decomposed in two parts: a convex sum to fuse the information of each agent with those of its neighbors, plus some local error or input. With this decomposition, all the update laws are put into the same form as the dynamic average consensus algorithm in the Appendix. This observation allows us to divide the analysis of the DLPDS algorithm in two steps. Firstly, we show all the estimates asymptotically achieve consensus by utilizing the property that the local errors and inputs are diminishing. Secondly, we further show that the consensus vectors coincide with a pair of primal and Lagrangian dual optimal solutions and the optimal value.

**Lemma 5.2 (Lipschitz continuity of $L_i$):** Consider $L_{i\mu}$ and $L_{ix}$. Then there are $L > 0$ and $R > 0$ such that $\|DL_{i\mu}(x)\| \leq L$ and $\|DL_{ix}(\mu)\| \leq R$ for each pair of $x \in co(\cup_{i=1}^N X_i)$ and $\mu \in co(\cup_{i=1}^N M_i)$. Furthermore, for each $\mu \in co(\cup_{i=1}^N M_i)$, the function $L_{i\mu}$ is Lipschitz continuous with Lipschitz constant $L$ over $co(\cup_{i=1}^N X_i)$, and for each $x \in co(\cup_{i=1}^N X_i)$, the function $L_{ix}$ is Lipschitz continuous with Lipschitz constant $R$ over $co(\cup_{i=1}^N M_i)$.
Proof: Due to the space limit, the proof is omitted here. The readers can find it in the enlarged version [25].

The following lemma provides a basic iteration relation used in the convergence proof for the DLPDS algorithm.

**Lemma 5.3 (Basic iteration relation):** Let the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. For any $x \in X$, any $\mu \in M$ and all $k \geq 0$, the following estimates hold:

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

\[
- \sum_{i=1}^{N} 2\alpha(k)(\mathcal{L}_i(v_x^i(k), v_{\mu}^i(k))) - \mathcal{L}_i(x, v_{\mu}^i(k)),
\]

(10)

\[
\sum_{i=1}^{N} \| e_{\mu}^i(k) - \alpha(k)D_{\mu}^i(k) \|^2 \leq \sum_{i=1}^{N} \alpha(k)^2 \| D_{\mu}^i(k) \|^2 + \sum_{i=1}^{N} \{ \| \mu^i(k) - \mu \|^2 - \| \mu^i(k+1) - \mu \|^2 \}
\]

\[
+ \sum_{i=1}^{N} 2\alpha(k)(\mathcal{L}_i(v_{\mu}^i(k), v_{\mu}^i(k))) - \mathcal{L}_i(v_{\mu}^i(k), \mu)).
\]

(11)

**Proof:** By Lemma 8.1 with $Z = M$, $z = v_{\mu}^i(k) + \alpha(k)D_{\mu}^i(k)$ and $y = \mu \in M$, we have that for all $k \geq 0$

\[
\sum_{i=1}^{N} \| e_{\mu}^i(k) - \alpha(k)D_{\mu}^i(k) \|^2 \leq \sum_{i=1}^{N} \| v_{\mu}^i(k) + \alpha(k)D_{\mu}^i(k) - \mu \|^2 - \sum_{i=1}^{N} \| \mu^i(k+1) - \mu \|^2
\]

\[
= \sum_{i=1}^{N} \| v_{\mu}^i(k) - \mu \|^2 + \sum_{i=1}^{N} \alpha(k)^2 \| D_{\mu}^i(k) \|^2
\]

\[
+ \sum_{i=1}^{N} 2\alpha(k)D_{\mu}^i(k)^T(v_{\mu}^i(k) - \mu) - \sum_{i=1}^{N} \| \mu^i(k+1) - \mu \|^2
\]

\[
\leq \sum_{i=1}^{N} \alpha(k)^2 \| D_{\mu}^i(k) \|^2 + \sum_{i=1}^{N} 2\alpha(k)D_{\mu}^i(k)^T(v_{\mu}^i(k) - \mu)
\]

\[
+ \sum_{i=1}^{N} \| \mu^i(k) - \mu \|^2 - \sum_{i=1}^{N} \| \mu^i(k+1) - \mu \|^2.
\]

(12)

One can show (11) by substituting the following Lagrangian supgradient inequality into (12):

\[
D_{\mu}^i(k)^T(\mu - v_{\mu}^i(k)) \geq \mathcal{L}_i(v_{\mu}^i(k), \mu) - \mathcal{L}_i(v_{\mu}^i(k), v_{\mu}^i(k)).
\]
Similarly, the equality (10) can be shown by using the following Lagrangian subgradient inequality: \( D_x^i(k)^T (x - v_x^i(k)) \leq L_i(x, v_x^i(k)) = L_i(v_x^i(k), v_{\mu}^i(k)) \).

**Lemma 5.4 (Achieving consensus):** Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Consider the sequences of \( \{x^i(k)\} \), \( \{\mu^i(k)\} \) and \( \{y^i(k)\} \) of the DLPDS algorithm with the step-size sequence \( \{\alpha(k)\} \) satisfying \( \lim_{k \to +\infty} \alpha(k) = 0 \). Then there exist \( x^* \in X \) and \( \mu^* \in M \) such that \( \lim_{k \to +\infty} \|x^i(k) - x^*\| = 0 \), \( \lim_{k \to +\infty} \|\mu^i(k) - \mu^*\| = 0 \) for all \( i \in V \), and \( \lim_{k \to +\infty} \|y^i(k) - y^j(k)\| = 0 \) for all \( i, j \in V \).

**Proof:** Observe that \( v_x^i(k) \in \text{co}(\cup_{i=1}^N X_i) \) and \( v_{\mu}^i(k) \in \text{co}(\cup_{i=1}^N M_i) \). Then it follows from Lemma 5.2 that \( \|D_x^i(k)\| \leq L \). From Lemma 5.3 it follows that

\[
\sum_{i=1}^N \|x^i(k+1) - x\|^2 \leq \sum_{i=1}^N \|x^i(k) - x\|^2 + \sum_{i=1}^N \alpha(k)^2 L^2 + \sum_{i=1}^N 2\alpha(k)(\|L_i(v_x^i(k), v_{\mu}^i(k))\| + \|L_i(x, v_{\mu}^i(k))\|). \tag{13}
\]

Notice that \( v_x^i(k) \in \text{co}(\cup_{i=1}^N X_i) \), \( v_{\mu}^i(k) \in \text{co}(\cup_{i=1}^N M_i) \) and \( x \in X \) are bounded. Since \( L_i \) is continuous, then \( L_i(v_x^i(k), v_{\mu}^i(k)) \) and \( L_i(x, v_{\mu}^i(k)) \) are bounded. Since \( \lim_{k \to +\infty} \alpha(k) = 0 \), the last two terms on the right-hand side of (13) converge to zero as \( k \to +\infty \). Taking limits on both sides of (13), one can see that \( \limsup_{k \to +\infty} \sum_{i=1}^N \|x^i(k+1) - x\|^2 \leq \liminf_{k \to +\infty} \sum_{i=1}^N \|x^i(k) - x\|^2 \) for any \( x \in X \), and thus \( \lim_{k \to +\infty} \sum_{i=1}^N \|x^i(k) - x\|^2 \) exists for any \( x \in X \). On the other hand, taking limits on both sides of (10) we obtain \( \lim_{k \to +\infty} \sum_{i=1}^N \|e_x^i(k) + \alpha(k)D_x^i(k)\|^2 = 0 \) and therefore we deduce that \( \lim_{k \to +\infty} \|e_x^i(k)\| = 0 \) for all \( i \in V \). It follows from Proposition 8.1 in the Appendix that \( \lim_{k \to +\infty} \|x^i(k) - x^i(k)\| = 0 \) for all \( i, j \in V \). Combining this with the property that \( \lim_{k \to +\infty} \|x^i(k) - x\| \) exists for any \( x \in X \), we deduce that there exists \( x^* \in \mathbb{R}^n \) such that \( \lim_{k \to +\infty} \|x^i(k) - x^*\| = 0 \) for all \( i \in V \). Since \( x^i(k) \in X_i \) and \( X_i \) is closed, it implies that \( x^* \in X_i \) for all \( i \in V \) and thus \( x^* \in X \). Similarly, one can show that there is \( \mu^* \in M \) such that \( \lim_{k \to +\infty} \|\mu^i(k) - \mu^*\| = 0 \) for all \( i \in V \).

Since \( \lim_{k \to +\infty} \|x^i(k) - x^*\| = 0 \) and \( f_i \) is continuous, then \( \lim_{k \to +\infty} \|u^i(k)\| = 0 \). It follows from Proposition 8.1 that \( \lim_{k \to +\infty} \|y^i(k) - y^j(k)\| = 0 \) for all \( i, j \in V \).
From Lemma 5.4, we know that the sequences of \(\{x^i(k)\}\) and \(\{\mu^i(k)\}\) of the DLPDS algorithm asymptotically agree on to some point in \(X\) and some point in \(M\), respectively. Denote by \(\Theta \subseteq X \times M\) the set of such limit points. Denote \(\hat{x}(k) := \frac{1}{N} \sum_{i=1}^{N} x^i(k)\) and \(\hat{\mu}(k) := \frac{1}{N} \sum_{i=1}^{N} \mu^i(k)\). The following lemma further characterizes that the points in \(\Theta\) are saddle points of the Lagrangian function \(L\) over \(X \times M\).

**Lemma 5.5 (Saddle-point characterization of \(\Theta\))**: Each point in \(\Theta\) is a saddle point of the Lagrangian function \(L\) over \(X \times M\).

**Proof**: Denote \(\Delta_x(k) := \max_{i,j \in V} \|x^i(k) - x^j(k)\|\). Notice that
\[
\|v^i_x(k) - \hat{x}(k)\| = \|\sum_{j=1}^{N} a^i_j(k)x^j(k) - \frac{1}{m} x^i(k)\|
\]
\[
= \|\sum_{j \neq i} a^i_j(k)(x^j(k) - x^i(k)) - \frac{1}{N}(x^i(k) - x^i(k))\|
\]
\[
\leq \sum_{j \neq i} a^i_j(k)\|x^j(k) - x^i(k)\| + \frac{1}{N}\|x^i(k) - x^i(k)\| \leq 2\Delta_x(k).
\]
Denote \(\Delta_{\mu}(k) := \max_{i,j \in V} \|\mu^i(k) - \mu^j(k)\|\). Similarly, we have \(\|v^i_\mu(k) - \hat{\mu}(k)\| \leq 2\Delta_{\mu}(k)\).

We will show this lemma by contradiction. Suppose that there is \((x^*, \mu^*) \in \Theta\) which is not a saddle point of \(L\) over \(X \times M\). Then at least one of the following equalities holds:

\[
\exists x \in X \quad \text{s.t.} \quad L(x^*, \mu^*) > L(x, \mu^*), \tag{14}
\]
\[
\exists \mu \in M \quad \text{s.t.} \quad L(x^*, \mu) > L(x^*, \mu^*). \tag{15}
\]
Suppose first that (14) holds. Then, there exists \(\zeta > 0\) such that \(L(x^*, \mu^*) = L(x, \mu^*) + \zeta\). Consider the sequences of \(\{x^i(k)\}\) and \(\{\mu^i(k)\}\) which converge respectively to \(x^*\) and \(\mu^*\) defined above. The estimate (10) leads to
\[
\sum_{i=1}^{N} \|x^i(k+1) - x\|^2 \leq \sum_{i=1}^{N} \|x^i(k) - x\|^2 + \alpha(k)^2 \sum_{i=1}^{N} \|D^i_x(k)\|^2
\]
\[- 2\alpha(k) \sum_{i=1}^{N} (A_i(k) + B_i(k) + C_i(k) + D_i(k) + E_i(k) + F_i(k)),
\]
where
\[
A_i(k) = L_i(v^i_x(k), v^i_\mu(k)) - L_i(\hat{x}(k), v^i_\mu(k)), \quad B_i(k) = L_i(\hat{x}(k), v^i_\mu(k)) - L_i(\hat{x}(k), \hat{\mu}(k)),
\]
\[
C_i(k) = L_i(\hat{x}(k), \hat{\mu}(k)) - L_i(x^*, \hat{\mu}(k)), \quad D_i(k) = L_i(x^*, \hat{\mu}(k)) - L_i(x^*, \mu^*),
\]
\[
E_i(k) = L_i(x^*, \mu^*) - L_i(x, \mu^*), \quad F_i(k) = L_i(x, \mu^*) - L_i(x, v^i_\mu(k)).
\]
It follows from the Lipschitz continuity property of $L_i$; see Lemma 5.2, that
\[
\|A_i(k)\| \leq L\|v_i^j(k) - \hat{x}(k)\| \leq 2L\Delta x(k), \quad \|B_i(k)\| \leq R\|v_i^j(k) - \hat{\mu}(k)\| \leq 2R\Delta \mu(k),
\]
\[
\|C_i(k)\| \leq L\|\hat{x}(k) - x^*\| \leq \frac{L}{N} \sum_{i=1}^{N} \|x^i(k) - x^*\|,
\]
\[
\|D_i(k)\| \leq R\|\hat{\mu}(k) - \mu^*\| \leq \frac{R}{N} \sum_{i=1}^{N} \|\mu_i^j(k) - \mu^*\|,
\]
\[
\|F_i(k)\| \leq R\|\mu^* - v_i^j(k)\| \leq R\|\mu^* - \hat{\mu}(k)\| + R\|\hat{\mu}(k) - v_i^j(k)\|
\leq \frac{R}{N} \sum_{i=1}^{N} \|\mu^*(k) - \mu_i^j(k)\| + 2R\Delta \mu(k).
\]
Since $\lim_{k \to +\infty} \|x^i(k) - x^*\| = 0$, $\lim_{k \to +\infty} \|\mu_i^j(k) - \mu^*\| = 0$, $\lim_{k \to +\infty} \Delta x(k) = 0$ and $\lim_{k \to +\infty} \Delta \mu(k) = 0$, then all $A_i(k), B_i(k), C_i(k), D_i(k), F_i(k)$ converge to zero as $k \to +\infty$. Then there exists $k_0 \geq 0$ such that for all $k \geq k_0$, it holds that
\[
\sum_{i=1}^{N} \|x^i(k + 1) - x\|^2 \leq \sum_{i=1}^{N} \|x^i(k) - x\|^2 + N\alpha(k)^2L^2 - \zeta\alpha(k).
\]
Following a recursive argument, we have that for all $k \geq k_0$, it holds that
\[
\sum_{i=1}^{N} \|x^i(k + 1) - x\|^2 \leq \sum_{i=1}^{N} \|x^i(k_0) - x\|^2 + NL^2 \sum_{\tau=k_0}^{k} \alpha(\tau)^2 - \zeta \sum_{\tau=k_0}^{k} \alpha(\tau).
\]
Since $\sum_{k=k_0}^{+\infty} \alpha(k) = +\infty$ and $\sum_{k=k_0}^{+\infty} \alpha(k)^2 < +\infty$ and $x^i(k_0) \in X_i$, $x \in X$ are bounded, the above estimate yields a contradiction by taking $k$ sufficiently large. In other words, (14) cannot hold. Following a parallel argument, one can show that (15) cannot hold either. This ensures that each $(x^*, \mu^*) \in \Theta$ is a saddle point of $L$ over $X \times M$.

The combination of (c) in Lemmas 3.1 and Lemma 5.5 gives that, for each $(x^*, \mu^*) \in \Theta$, we have that $L(x^*, \mu^*) = p^*$ and $\mu^*$ is Lagrangian dual optimal. We still need to verify that $x^*$ is a primal optimal solution. We are now in the position to show Theorem 3.2 based on two claims.

**Proofs of Theorem 3.2:**

**Claim 1:** Each point $(x^*, \mu^*) \in \Theta$ is a point in $X^* \times D_{L^*}^i$.

**Proof:** The Lagrangian dual optimality of $\mu^*$ follows from (c) in Lemma 3.1 and Lemma 5.5. To characterize the primal optimality of $x^*$, we define an auxiliary sequence $\{z(k)\}$ by $z(k) := \frac{\sum_{\tau=0}^{k-1} \alpha(\tau) \hat{x}(\tau)}{\sum_{\tau=0}^{k-1} \alpha(\tau)}$. Since $\lim_{k \to +\infty} \hat{x}(k) = x^*$, it follows from Lemma 5.1 (b) that $\lim_{k \to +\infty} z(k) = x^*$. 

---

Limited circulation. For review only

Preprint submitted to IEEE Transactions on Automatic Control. Received: October 26, 2009 13:02:07 PST
Since \((x^*, \mu^*)\) is a saddle point of \(L\) over \(X \times M\), then \(L(x^*, \mu) \leq L(x^*, \mu^*)\) for any \(\mu \in M\); i.e., the following relation holds for any \(\mu \in M\):

\[
g(x^*)^T(\mu - \mu^*) \leq 0. \tag{16}
\]

Choose \(\mu_a = \mu^* + \min_{i \in V} \theta_i \frac{x^*}{\|x^*\|}\) where \(\theta_i > 0\) is given in the definition of \(M_i\). Then \(\mu_a \geq 0\) and \(\|\mu_a\| \leq \|\mu^*\| + \min_{i \in V} \theta_i\) implying \(\mu_a \in M\). Letting \(\mu = \mu_a\) in (16) gives that \(\frac{\min_{i \in V} \theta_i}{\|\mu^*\|} g(x^*)^T \mu_a \leq 0\). Since \(\theta_i > 0\), we have \(g(x^*)^T \mu_a \leq 0\). On the other hand, we choose \(\mu_b = \frac{1}{2} \mu^*\) and then \(\mu_b \in M\). Letting \(\mu = \mu_b\) in (16) gives that \(-\frac{1}{2} g(x^*)^T \mu^* \leq 0\) and thus \(g(x^*)^T \mu^* \geq 0\). The combination of the above two estimates guarantees the property of \(g(x^*)^T \mu^* = 0\).

We now proceed to show \(g(x^*) \leq 0\) by contradiction. Assume that \(g(x^*) \leq 0\) does not hold. Denote \(J^+(x^*) := \{1 \leq \ell \leq m \mid g_\ell(x^*) > 0\} \neq \emptyset\) and \(\eta := \min_{\ell \in J^+(x^*)} \{g_\ell(x^*)\}\). Then \(\eta > 0\). Since \(g\) is continuous and \(v^i_x(k)\) converges to \(x^*\), there exists \(K \geq 0\) such that \(g_\ell(v^i_x(k)) \geq \frac{\eta}{2}\) for all \(k \geq K\) and all \(\ell \in J^+(x^*)\). Since \(v^i_x(k)\) converges to \(\mu^*\), without loss of generality, we say that \(\|v^i_\mu(k) - \mu^*\| \leq \frac{1}{2} \min_{i \in V} \theta_i\) for all \(k \geq K\). Choose \(\hat{\mu}\) such that \(\hat{\mu}_\ell = \mu^*_i\) for \(\ell \notin J^+(x^*)\) and \(\hat{\mu}_\ell = \mu^*_i + \min_{i \in V} \theta_i\) for \(\ell \in J^+(x^*)\). Since \(\mu^*_i \geq 0\) and \(\theta_i > 0\), thus \(\hat{\mu} \geq 0\). Furthermore, \(\|\hat{\mu}\| \leq \|\mu^*\| + \min_{i \in V} \theta_i\), then \(\hat{\mu} \in M\). Equating \(\mu\) to \(\hat{\mu}\) and letting \(D^i_\mu(k) = g(v^i_x(k))\) in the estimate (12), the following holds for \(k \geq K\):

\[
N |J^+(x^*)| \eta \min_{i \in V} \theta_i \alpha(k) \leq 2 \alpha(k) \sum_{i=1}^{N} \sum_{\ell \in J^+(x^*)} g_\ell(v^i_x(k))(\hat{\mu} - v^i_\mu(k))_\ell \\
\leq \sum_{i=1}^{N} \|\mu^i(k) - \hat{\mu}\|^2 - \sum_{i=1}^{N} \|\mu^i(k + 1) - \hat{\mu}\|^2 + NR^2 \alpha(k)^2 \\
- 2 \alpha(k) \sum_{i=1}^{N} \sum_{\ell \in J^+(x^*)} g_\ell(v^i_x(k))(\hat{\mu} - v^i_\mu(k))_\ell. \tag{17}
\]

Summing (17) over \([K, k - 1]\) with \(k \geq K + 1\), dividing by \(\sum_{\tau = K}^{k-1} \alpha(\tau)\) on both sides, and using \(- \sum_{i=1}^{N} \|\mu^i(k) - \hat{\mu}\|^2 \leq 0\), we obtain

\[
N |J^+(x^*)| \eta \min_{i \in V} \theta_i \leq \frac{1}{\sum_{\tau = K}^{k-1} \alpha(\tau)} \left\{ \sum_{i=1}^{N} \|\mu^i(K) - \hat{\mu}\|^2 + NR^2 \sum_{\tau = K}^{k-1} \alpha(\tau)^2 \\
- \sum_{\tau = K}^{k-1} 2 \alpha(\tau) \sum_{i=1}^{N} \sum_{\ell \notin J^+(x^*)} g_\ell(v^i_x(\tau))(\hat{\mu} - v^i_\mu(\tau))_\ell \right\}. \tag{18}
\]

Since \(\mu^i(K) \in M_i\), \(\hat{\mu} \in M\) are bounded and \(\sum_{\tau = K}^{+\infty} \alpha(\tau) = +\infty\), then the limit of the first term on the right hand side of (18) is zero as \(k \to +\infty\). Since \(\sum_{\tau = K}^{+\infty} \alpha(\tau)^2 < +\infty\), then the
limit of the second term is zero as \( k \to +\infty \). Since \( \lim_{k \to +\infty} v_\mu^i(k) = \mu^* \), thus
\[
\lim_{k \to +\infty} \frac{1}{2} \sum_{i=1}^{N} \sum_{\ell \geq f^+(x^*)} g_{\ell}(v_\mu^i(k)) (\mu - v_\mu^i(k)) = 0.
\]
Then it follows from Lemma 5.1 (b) that then the limit of the third term is zero as \( k \to +\infty \). Then we have \( N |J^+(x^*)| \eta \min_{i \in V} \theta_i \leq 0 \). Recall that \( |J^+(x^*)| > 0 \), \( \eta > 0 \) and \( \theta_i > 0 \). Then we reach a contradiction, implying that \( g(x^*) \leq 0 \).

Since \( x^* \in X \) and \( g(x^*) \leq 0 \), then \( x^* \) is a feasible solution and thus \( f(x^*) = p^* \). On the other hand, since \( z(k) \) is a convex combination of \( \hat{x}(0), \ldots, \hat{x}(k-1) \) and \( f \) is convex, thus we have the following estimate:
\[
f(z(k)) \leq \frac{1}{\sum_{\tau=0}^{k-1} \alpha(\tau)} \sum_{\tau=0}^{k-1} \alpha(\tau) \mathcal{L}(\hat{x}(\tau), \hat{\mu}(\tau)) - \sum_{\tau=0}^{k-1} N \alpha(\tau) \hat{\mu}(\tau)^T g(\hat{x}(\tau)).
\]

Recall that \( \lim_{k \to +\infty} z(k) = x^* \), \( \lim_{k \to +\infty} \mathcal{L}(\hat{x}(k), \hat{\mu}(k)) = \mathcal{L}(x^*, \mu^*) = p^* \), \( \lim_{k \to +\infty} \hat{\mu}(k)^T g(\hat{x}(k)) = g(x^*)^T \mu^* = 0 \), it follows from Lemma 5.1 (b) that \( f(x^*) \leq p^* \). Therefore, we have \( f(x^*) = p^* \), and thus \( x^* \) is a primal optimal point. \( \blacksquare \)

**Claim 2:** It holds that \( \lim_{k \to +\infty} \|y^i(k) - p^*\| = 0 \).

**Proof:** The following can be proven by induction on \( k \) for a fixed \( k' \geq 1 \):
\[
\sum_{i=1}^{N} y^i(k+1) = \sum_{i=1}^{N} y^i(k') + N \sum_{\ell=k'}^{k} \sum_{i=1}^{N} (f_i(x^i(\ell)) - f_i(x^i(\ell - 1))). \tag{19}
\]
Let \( k' = 1 \) in (19) and recall that initial state \( y^i(1) = N f_i(x^i(0)) \) for all \( i \in V \). Then we have
\[
\sum_{i=1}^{N} y^i(k+1) = \sum_{i=1}^{N} y^i(1) + N \sum_{i=1}^{N} (f_i(x^i(k)) - f_i(x^i(0))) = N \sum_{i=1}^{N} f_i(x^i(k)). \tag{20}
\]
The combination of (20) with \( \lim_{k \to +\infty} \|y^i(k) - y^i(k)\| = 0 \) gives the desired result. \( \blacksquare \)

**B. Proofs of Theorem 4.2**

In order to analyze the DPPDS algorithm, we first rewrite it into the following form:
\[
\begin{align*}
\mu^i(k+1) &= v_\mu^i(k) + u_\mu^i(k), \\
\lambda^i(k+1) &= v_\lambda^i(k) + u_\lambda^i(k), \\
x^i(k+1) &= v_x^i(k) + e_x^i(k), \\
y^i(k+1) &= v_y^i(k) + u_y^i(k),
\end{align*}
\]
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 \( u^i_\mu(k) := \alpha(k)[g(v^i_x(k))]^+ \), \( u^i_\lambda(k) := \alpha(k)h(v^i_x(k)) \), \( u^i_y(k) = N(f_i(x^i(k)) - f_i(x^i(k - 1))) \) are some local inputs. Denote by \( M_\mu(k) := \max_{i \in V} \|\mu^i(k)\| \) and \( M_\lambda(k) := \max_{i \in V} \|\lambda^i(k)\| \).

Before showing Lemma 5.6, we present some useful facts. Since \( X \) is compact, and \( f_i, [g(\cdot)]^+, h(\cdot) \) are continuous, there exist \( F, G^+, H > 0 \) such that for all \( x \in X \), it holds that \( \|f_i(x)\| \leq F \) for all \( i \in V \), \( \|[g(x)]^+\| \leq G^+ \) and \( \|h(x)\| \leq H \). Since \( X \) is a compact set and \( f_i, [g(\cdot)]^+, h(\cdot) \) are convex, then it follows from Proposition 5.4.2 in [2] that there exist \( D_F, D_{G^+}, D_H > 0 \) such that for all \( x \in X \), it holds that \( \|Df_i(x)\| \leq D_F (i \in V) \), \( m\|D[g(x)]^+\| \leq D_{G^+} (1 \leq \ell \leq m) \) and \( \nu\|D[h_x](x)\| \leq D_H (1 \leq \ell \leq \nu) \). Denote by \( \hat{x}(k) := \frac{1}{N} \sum_{i=1}^N x^i(k) \), \( \hat{\mu}(k) := \frac{1}{N} \sum_{i=1}^N \mu^i(k) \) and \( \hat{\lambda}(k) := \frac{1}{N} \sum_{i=1}^N \lambda^i(k) \).

**Lemma 5.6:** Suppose the balanced communication assumption 2.3 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 \( \{\alpha(k)^2M^2_\mu(k)\}, \{\alpha(k)^2M^2_\lambda(k)\} \) and \( \{\alpha(k)^2\|S_x^i(k)\|^2\} \) are summable.

(b) The sequences \( \{\alpha(k)||\hat{\mu}(k) - v^i_\mu(k)\|\}, \{\alpha(k)||\hat{\lambda}(k) - v^i_\lambda(k)\|\}, \{\alpha(k)||\hat{\mu}(k)M_\mu(k)||\hat{x}(k) - v^i_x(k)\|\}, \{\alpha(k)||\hat{\lambda}(k)M_\lambda(k)||\hat{x}(k) - v^i_x(k)\|\} \) and \( \{\alpha(k)||\hat{x}(k) - v^i_x(k)\|\} \) are summable.

**Proof:** (a) Notice that
\[
\|v^i_\mu(k)\| = \|\sum_{j=1}^N a^i_j(k)\mu^i(k)\| \leq \sum_{j=1}^N a^i_j(k)\|\mu^i(k)\| \leq \sum_{j=1}^N a^i_j(k)M_\mu(k) = M_\mu(k),
\]
where in the last equality we use the balanced communication assumption 2.3. Recall that \( v^i_x(k) \in X \). This implies that the following inequalities hold for all \( k \geq 0 \):
\[
\|\mu^i(k + 1)\| \leq \|v^i_\mu(k) + \alpha(k)[g(v^i_x(k))]^+\| \leq \|v^i_\mu(k)\| + G^+\alpha(k) \leq M_\mu(k) + G^+\alpha(k).
\]
From here, then we deduce the following recursive estimate on \( M_\mu(k + 1) \): \( M_\mu(k + 1) \leq M_\mu(k) + G^+\alpha(k) \). Repeatedly applying the above estimates yields that
\[
M_\mu(k + 1) \leq M_\mu(0) + G^+s(k). \tag{21}
\]

Similar arguments can be employed to show that
\[
M_\lambda(k + 1) \leq M_\lambda(0) + Hs(k). \tag{22}
\]

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 \( \lim_{k \to +\infty} \alpha(k + 1)M_\lambda(k + 1) = 0 \). Notice that the following estimate on \( S_x^i(k) \) holds:
\[
\|S_x^i(k)\| \leq D_F + D_{G^+}M_\mu(k) + D_HP_XM_\lambda(k). \tag{23}
\]
Recall 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 the result of \( \lim_{k \to +\infty} \alpha(k)\|S_{x}^{i}(k)\| = 0 \) follows. By (21), we have
\[
\sum_{k=0}^{+\infty} \alpha(k)^2 M_{\mu}^2(k) \leq \alpha(0)^2 M_{\mu}^2(0) + \sum_{k=1}^{+\infty} \alpha(k)^2(M_{\mu}(0) + G^{+} s(k-1))^2.
\]
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 \( \sum_{k=0}^{+\infty} \alpha(k)^2 M_{\lambda}^2(k) < +\infty \). Then the summability of \( \{\alpha(k)^2\} \), \( \{\alpha(k + 1)^2 s(k)\} \) and \( \{\alpha(k + 1)^2 s(k)^2\} \) verifies that of \( \{\alpha(k)^2\|S_{x}^{i}(k)\|^2\} \).

(b) Consider the dynamics of \( \mu^i(k) \) which is in the same form as the distributed projected subgradient algorithm in [17]. Recall that \( \{[g(v_{x}^{i}(k))]^{+}\} \) is uniformly bounded. Then following from Lemma 8.2 in the Appendix with \( Z = \mathbb{R}_{\geq 0}^{m} \) 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 by using the following set of inequalities:
\[
\|\hat{\mu}(k) - v_{\mu}^{i}(k)\| \leq \sum_{j=1}^{N} a_{j}^{i}(k)\|\hat{\mu}(k) - \mu^{j}(k)\| \leq \max_{i \in V} \|\hat{\mu}(k) - \mu^{i}(k)\|,
\]
where we use \( \sum_{j=1}^{N} a_{j}^{i}(k) = 1 \). Similarly, it holds that \( \sum_{k=0}^{+\infty} \alpha(k)\|\hat{\lambda}(k) - v_{\lambda}^{i}(k)\| < +\infty \).

We now consider the evolution of \( x^{i}(k) \). Recall that \( v_{x}^{i}(k) \in X \). By Lemma 8.1 with \( Z = X \), \( z = v_{x}^{i}(k) - \alpha(k)S_{x}^{i}(k) \) and \( y = v_{x}^{i}(k) \), we have
\[
\|x^{i}(k + 1) - v_{x}^{i}(k)\|^2 \leq \|v_{x}^{i}(k) - \alpha(k)S_{x}^{i}(k) - v_{x}^{i}(k)\|^2 - \|x^{i}(k + 1) - (v_{x}^{i}(k) - \alpha(k)S_{x}^{i}(k))\|^2,
\]
and thus \( \|\epsilon_{x}^{i}(k) + \alpha(k)S_{x}^{i}(k)\| \leq \alpha(k)\|S_{x}^{i}(k)\| \). With this relation, from Lemma 8.2 with \( Z = X \) and \( d_{i}(k) = S_{x}^{i}(k) \), the following holds for some \( \gamma > 0 \) and \( 0 < \beta < 1 \):
\[
\|x^{i}(k) - \hat{x}(k)\| \leq N\gamma\beta^{k-1} \sum_{i=0}^{N} \|x^{i}(0)\| + 2N\gamma \sum_{\tau=0}^{k-1} \beta^{k-\tau} \alpha(\tau)\|S_{x}^{i}(\tau)\|.
\]
Multiplying both sides of (25) by $\alpha(k)M_\mu(k)$ and using (23), we obtain
\[
\alpha(k)M_\mu(k)\|x^i(k) - \hat{x}(k)\| \leq N\gamma \sum_{i=0}^{N} \|x^i(0)\|\alpha(k)M_\mu(k)\beta^{k-1} + 2N\gamma\alpha(k)M_\mu(k) \\
\times \sum_{\tau=0}^{k-1} \beta^{k-\tau} \alpha(\tau)(D_F + D_G + M_\mu(\tau) + D_H M_\lambda(\tau)).
\]

Notice that the above inequalities hold for all $i \in V$. Then by employing the relation of $ab \leq \frac{1}{2}(a^2 + b^2)$ and regrouping similar terms, we obtain
\[
\alpha(k)M_\mu(k) \max_{i \in V} \|x^i(k) - \hat{x}(k)\| \leq N\gamma \left( \frac{1}{2} \sum_{i=0}^{N} \|x^i(0)\| + (D_F + D_G + D_H) \sum_{\tau=0}^{k-1} \beta^{k-\tau} \right) \\
\times \alpha(k)^2 M_\mu^2(k) + \frac{1}{2} N\gamma \sum_{i=0}^{N} \|x^i(0)\|\beta^{2(k-1)} + N\gamma \sum_{\tau=0}^{k-1} \beta^{k-\tau} \alpha(\tau)^2 (D_F + D_G + M_\mu^2(\tau) + D_H M_\lambda^2(\tau)).
\]

Part (a) gives that $\{\alpha(k)^2 M_\mu^2(k)\}$ is summable. Combining this fact with $\sum_{\tau=0}^{k-1} \beta^{k-\tau} \leq \frac{1}{1-b}$, then we can say that the first term on the right-hand side in the above estimate is summable. It is easy to check that the second term is also summable. It follows from Part (a) that
\[
\lim_{k \to +\infty} \alpha(k)^2 (D_F + D_G + M_\mu^2(k) + D_H M_\lambda^2(k)) = 0 \quad \text{and} \quad \{\alpha(k)^2 (D_F + D_G + M_\mu^2(k) + D_H M_\lambda^2(k))\}
\]

is summable. Then Lemma 7 in [17] with $\gamma = N\gamma\alpha(\ell)^2 (D_F + D_G + M_\mu^2(\ell) + D_H M_\lambda^2(\ell))$ ensures that the third term is summable. Therefore, the summability of $\{\alpha(k)M_\mu(k) \max_{i \in V} \|x^i(k) - \hat{x}(k)\|\}$ is guaranteed. Following the same lines in (24), one can show the summability of $\{\alpha(k)M_\mu(k)\|v_x^i(k) - \hat{x}(k)\|\}$. Following analogous arguments, we have that $\{\alpha(k)M_\lambda(k)\|v_x^i(k) - \hat{x}(k)\|\}$ are summable.

Lemma 5.7 (Basic iteration relation): The following estimates hold for any $x \in X$ and $(\mu, \lambda) \in \mathbb{R}_\geq 0 \times \mathbb{R}_\geq 0$:
\[
\sum_{i=1}^{N} \|e_x^i(k) + \alpha(k)S_x^i(k)\|^2 \leq \sum_{i=1}^{N} \alpha(k)^2 \|S_x^i(k)\|^2 \\
- \sum_{i=1}^{N} 2\alpha(k)(\mathcal{H}_i(v_x^i(k), v_x^i(k), v_x^i(k)) - \mathcal{H}_i(x, v_x^i(k), v_x^i(k))) \\
+ \sum_{i=1}^{N} (\|x^i(k) - x\|^2 - \|x^i(k+1) - x\|^2), \tag{26}
\]
and,
\[ 0 \leq \sum_{i=1}^{N}(\|\mu^i(k) - \mu\|^2 - \|\mu^i(k+1) - \mu\|^2) + \sum_{i=1}^{N}(\|\lambda^i(k) - \lambda\|^2 - \|\lambda^i(k+1) - \lambda\|^2) + \]
\[ \sum_{i=1}^{N}2\alpha(k)(\mathcal{H}_i(v^i_x(k), v^i_\mu(k), v^i_\lambda(k)) - \mathcal{H}_i(\hat{v}^i_x(k), \mu, \lambda)) + \sum_{i=1}^{N}\alpha(k)^2(\|g(v^i_x(k))\|^2 + \|h(v^i_x(k))\|^2). \]

(27)

**Proof:** One can finish the proof by following analogous arguments in Lemma 5.3.

**Lemma 5.8 (Achieving consensus):** Let us suppose that the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodical strong connectivity assumption 2.4 hold. Consider the sequences of \(\{x^i(k)\}, \{\mu^i(k)\}, \{\lambda^i(k)\}\) and \(\{y^i(k)\}\) of the distributed penalty primal-dual subgradient algorithm with the step-size sequence \(\{\alpha(k)\}\) and the associated \(\{s(k)\}\) satisfying \(\lim_{k \to +\infty} \alpha(k) = 0\) and \(\lim_{k \to +\infty} \alpha(k+1)s(k) = 0\). 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:** Similar to (12), we have
\[ \sum_{i=1}^{N}\|x^i(k+1) - x\|^2 \leq \sum_{i=1}^{N}\|x^i(k) - x\|^2 + \sum_{i=1}^{N}\alpha(k)^2\|S^i_x(k)\|^2 + \sum_{i=1}^{N}2\alpha(k)\|S^i_x(k)\|\|v^i_x(k) - x\|. \]

Since \(\lim_{k \to +\infty} \alpha(k)\|S^i_x(k)\| = 0\), the proofs of \(\lim_{k \to +\infty} \|x^i(k) - \tilde{x}\| = 0\) for all \(i \in V\) are analogous to those in Lemma 5.4. The remainder of the proofs can be finished by Proposition 8.1 with the properties of \(\lim_{k \to +\infty} u^i_\mu(k) = 0\), \(\lim_{k \to +\infty} u^i_\lambda(k) = 0\) and \(\lim_{k \to +\infty} u^i_y(k) = 0\) (due to \(\lim_{k \to +\infty} x^i(k) = \tilde{x}\) and \(f_i\) is continuous).

We now proceed to show Theorem 4.2 based on five claims.

**Proof of Theorem 4.2:**

**Claim 1:** For any \(x^* \in X^*\) and \((\mu^*, \lambda^*) \in D^*_p\), \(\{\alpha(k)\} [\sum_{i=1}^{N} \mathcal{H}_i(x^*, v^i_\mu(k), v^i_\lambda(k)) - \mathcal{H}(x^*, \hat{\mu}(k), \hat{\lambda}(k))\}\) and \(\{\alpha(k)\} [\sum_{i=1}^{N} \mathcal{H}_i(v^i_x(k), \mu^*, \lambda^*) - \mathcal{H}(\hat{x}(k), \mu^*, \lambda^*)\}\) are summable.

**Proof:** Observe that
\[ \|\mathcal{H}_i(x^*, v^i_\mu(k), v^i_\lambda(k)) - \mathcal{H}_i(x^*, \hat{\mu}(k), \hat{\lambda}(k))\| \]
\[ \leq \|v^i_\mu(k) - \hat{\mu}(k)\|\|g(x^*)\| + \|v^i_\lambda(k) - \hat{\lambda}(k)\|\|h(x^*)\| \]
\[ \leq G^+\|v^i_\mu(k) - \hat{\mu}(k)\| + H\|v^i_\lambda(k) - \hat{\lambda}(k)\|. \]

(28)

**DRAFT**
By using the summability of \(\{\alpha(k)\|\hat{\mu}(k) - v_i^\mu(k)\|\}\) and \(\{\alpha(k)\|\hat{\lambda}(k) - v_i^\lambda(k)\|\}\) in Part (b) of Lemma 5.6, we have that \(\{\alpha(k)\sum_{i=1}^N \|H_i(x^*, v_i^\mu(k), v_i^\lambda(k)) - \mathcal{H}_i(x^*, \hat{\mu}(k), \hat{\lambda}(k))\|\}\) and thus
\[
\{\alpha(k)\left[\sum_{i=1}^N (H_i(x^*, v_i^\mu(k), v_i^\lambda(k)) - \mathcal{H}_i(x^*, \hat{\mu}(k), \hat{\lambda}(k)))\right]\}
\] are summable. Similarly, the following estimates hold:
\[
\|H_i(v_i^\mu(k), \mu^*, \lambda^*) - \mathcal{H}_i(\hat{x}(k), \mu^*, \lambda^*)\| \leq \|f_i(v_i^\mu(k)) - f_i(\hat{x}(k))\| + \|(\mu^*)^T [g(v_i^\mu(k))^+] - [g(\hat{x}(k))^+]| + \|(\lambda^*)^T [h(v_i^\mu(k))] - |h(\hat{x}(k))]|\| \\
\leq (D_F + D_G + \|\mu^*\| + D_H \|\lambda^*\|) \|v_i^\mu(k) - \hat{x}(k)\|.
\]
Then the property of \(\sum_{k=0}^{+\infty} \alpha(k)\|\hat{x}(k) - v_i^\mu(k)\| < +\infty\) in Part (b) of Lemma 5.6 implies the summability of \(\{\alpha(k)\sum_{i=1}^N \|H_i(v_i^\mu(k), \mu^*, \lambda^*) - \mathcal{H}_i(\hat{x}(k), \mu^*, \lambda^*)\|\}\) and thus \(\{\alpha(k)\sum_{i=1}^N (H_i(v_i^\mu(k), \mu^*, \lambda^*) - \mathcal{H}_i(\hat{x}(k), \mu^*, \lambda^*))\}\).

**Claim 2:** Denote by
\[
\hat{H}_i(k) := \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell)H_i(v_i^\mu(\ell), v_i^\lambda(\ell), v_i^\mu(\ell)).
\]
The following property holds: \(\lim_{k \to +\infty} \sum_{i=1}^N \hat{H}_i(k) = p^*\).

**Proof:** Summing (26) from 0 to \(k - 1\) and replacing \(x\) by \(x^* \in X^*\) leads to
\[
\sum_{\ell=0}^{k-1} \alpha(\ell) \sum_{i=1}^N (H_i(v_i^\mu(\ell), v_i^\lambda(\ell)) - H_i(x^*, v_i^\mu(\ell), v_i^\lambda(\ell))) \\
\leq \sum_{i=1}^N \|x^i(0) - x^*\|^2 + \sum_{\ell=0}^{k-1} \sum_{i=1}^N \alpha(\ell)^2 \|S^i_x(\ell)\|^2. \tag{29}
\]
The summability of \(\{\alpha(k)^2\|S^i_x(k)\|^2\}\) in Part (b) of Lemma 5.6 implies that the right-hand side of (29) is finite as \(k \to +\infty\), and thus
\[
\limsup_{k \to +\infty} \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) \left[ \sum_{i=1}^N (H_i(v_i^\mu(\ell), v_i^\lambda(\ell)) - H_i(x^*, v_i^\mu(\ell), v_i^\lambda(\ell))) \right] \leq 0. \tag{30}
\]

Pick any \((\mu^*, \lambda^*) \in D^*_p\). It follows from Theorem 4.1 that \((x^*, \mu^*, \lambda^*)\) is a saddle point of \(\mathcal{H}\) over \(X \times \mathbb{R}^m_0 \times \mathbb{R}^\nu_0\). Since \((\hat{\mu}(k), \hat{\lambda}(k)) \in \mathbb{R}^m_0 \times \mathbb{R}^\nu_0\), then we have \(\mathcal{H}(x^*, \hat{\mu}(k), \hat{\lambda}(k)) \leq
\( \mathcal{H}(x^*, \mu^*, \lambda^*) = p^* \). Combining this relation, Claim 1 and (30) renders that

\[
\limsup_{k \to +\infty} \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) \left[ \sum_{i=1}^{N} \mathcal{H}_i(v^i_x(\ell), v^i_\mu(\ell), v^i_\lambda(\ell)) - p^* \right]
\]

\[
\leq \limsup_{k \to +\infty} \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) \left[ \sum_{i=1}^{N} \left( \mathcal{H}_i(v^i_x(\ell), v^i_\mu(\ell), v^i_\lambda(\ell)) - \mathcal{H}_i(x^*, v^i_\mu(\ell), v^i_\lambda(\ell)) \right) \right]
\]

\[
+ \limsup_{k \to +\infty} \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) \left[ \sum_{i=1}^{N} \mathcal{H}_i(x^*, v^i_\mu(\ell), v^i_\lambda(\ell)) - \mathcal{H}(x^*, \hat{\mu}(\ell), \hat{\lambda}(\ell)) \right]
\]

\[
+ \limsup_{k \to +\infty} \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \left( \mathcal{H}(x^*, \hat{\mu}(\ell), \hat{\lambda}(\ell)) - p^* \right) \leq 0,
\]

and thus \( \limsup_{k \to +\infty} \sum_{i=1}^{N} \mathcal{H}_i(k) \leq p^* \).

On the other hand, \( \hat{x}(k) \in X \) (due to the fact that \( X \) is convex) implies that \( \mathcal{H}(\hat{x}(k), \mu^*, \lambda^*) \geq \mathcal{H}(x^*, \mu^*, \lambda^*) = p^* \). Along similar lines, by using (27) with \( \mu = \mu^*, \lambda = \lambda^* \), and Claim 1, we have the following estimate: \( \liminf_{k \to +\infty} \sum_{i=1}^{N} \mathcal{H}_i(k) \geq p^* \). Then we have the desired relation. \( \blacksquare \)

**Claim 3:** Denote by \( \pi(k) := \sum_{i=1}^{N} \mathcal{H}_i(v^i_x(k), v^i_\mu(k), v^i_\lambda(k)) - \mathcal{H}(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) \) and

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

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

**Proof:** Notice that

\[
\pi(k) = \sum_{i=1}^{N} (f_i(v^i_x(k)) - f_i(\hat{x}(k))) + \sum_{i=1}^{N} \left( v^i_\mu(k)^T [g(v^i_x(k))]^+ - v^i_\mu(k)^T \left[ g(\hat{x}(k)) \right]^+ \right)
\]

\[
+ \sum_{i=1}^{N} \left( v^i_\lambda(k)^T [h(v^i_x(k))]^+ - \hat{\mu}(k)^T \left[ h(\hat{x}(k)) \right]^+ \right) + \sum_{i=1}^{N} \left( v^i_\lambda(k)^T |h(v^i_x(k))| - v^i_\lambda(k)^T |h(\hat{x}(k))| \right)
\]

\[
+ \sum_{i=1}^{N} \left( v^i_\lambda(k)^T |h(\hat{x}(k))| - \hat{\lambda}(k)^T |h(\hat{x}(k))| \right). \tag{31}
\]

By using the boundedness of subdifferentials and the primal estimates, it follows from (31) that

\[
\| \pi(k) \| \leq (D_F + D_{G^+} M_\mu(k) + D_H M_\lambda(k)) \times \sum_{i=1}^{N} \| v^i_x(k) - \hat{x}(k) \|
\]

\[
+ G^+ \sum_{i=1}^{N} \| v^i_\mu(k) - \hat{\mu}(k) \| + H \sum_{i=1}^{N} \| v^i_\lambda(k) - \hat{\lambda}(k) \|. \tag{32}
\]

Preprint submitted to IEEE Transactions on Automatic Control. Received: October 26, 2009 13:02:07 PST
Then it follows from (b) in Lemma 5.6 that \( \{\alpha(k)\parallel \pi(k)\} \) is summable. Notice that \( \|\hat{H}(k) - \sum_{i=1}^{N} \hat{H}_i(k)\| \leq \frac{\sum_{l=0}^{k-1} \alpha(l)\parallel \pi(l)\}}{s(k-1)} \), and thus \( \lim_{k \to +\infty} \|\hat{H}(k) - \sum_{i=1}^{N} \hat{H}_i(k)\| = 0 \). The desired result immediately follows from Claim 2.

\textbf{Claim 4:} The limit point \( \tilde{x} \) in Lemma 5.8 is a primal optimal solution.

\textbf{Proof:} Let \( \hat{\mu}(k) = (\hat{\mu}_1(k), \ldots, \hat{\mu}_m(k))^T \in \mathbb{R}^m_{\geq 0} \). By the balanced communication assumption 2.3, we obtain
\[
\sum_{i=1}^{N} \mu^i(k + 1) = \sum_{i=1}^{N} (\sum_{j=1}^{N} a_{ij}^i(k) \mu^j(k) + \alpha(k) \sum_{i=1}^{N} [g(v_x^i(k))]^+) + \alpha(k) \sum_{i=1}^{N} [g(v_x^i(k))]^+.
\]

This implies that the sequence \( \{\hat{\mu}_\ell(k)\} \) is non-decreasing, then \( \hat{\mu}_\ell(k) \) is lower bounded by zero. In this way, we distinguish the following two cases:

\textbf{Case 1:} The sequence \( \{\hat{\mu}_\ell(k)\} \) is upper bounded. Then \( \{\hat{\mu}_\ell(k)\} \) is convergent in \( \mathbb{R}_{\geq 0} \). Recall that \( \lim_{k \to +\infty} \|\mu^i(k) - \mu^j(k)\| = 0 \) for all \( i, j \in V \). This implies that there exists \( \mu^*_\ell \in \mathbb{R}_{\geq 0} \) such that \( \lim_{k \to +\infty} \|\mu^i_k - \mu^*_\ell\| = 0 \) for all \( i \in V \). Observe that \( \sum_{i=1}^{N} \mu^i(k + 1) = \sum_{i=1}^{N} \mu^i(0) + \sum_{\tau=0}^{k} \alpha(\tau) \sum_{i=1}^{N} [g(v_x^i(\tau))]^+ \). Thus, we have \( \sum_{k=0}^{\infty} \alpha(k) \sum_{i=1}^{N} [g(v_x^i(\tau))]^+ < +\infty \), implying that \( \lim_{k \to +\infty} \sum_{i=1}^{N} [g(v_x^i(\hat{x}(\ell)))^+] \). Since \( \lim_{k \to +\infty} \|x^i(k) - \bar{x}\| = 0 \) for all \( i \in V \), then \( \lim_{k \to +\infty} \|v_x^i(k) - \bar{x}\| = 0 \), and thus \( [g_\ell(\bar{x})]^+ = 0 \).

\textbf{Case 2:} The sequence \( \{\hat{\mu}_\ell(k)\} \) is not upper bounded. Since \( \{\hat{\mu}_\ell(k)\} \) is non-decreasing, then \( \hat{\mu}_\ell(k) \to +\infty \). It follows from Claim 3 and (a) in Lemma 5.1 that it is impossible that \( H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) \to +\infty \). Assume that \( [g_\ell(\bar{x})]^+ > 0 \). Then we have
\[
H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) = f(\hat{x}(k)) + N\hat{\mu}(k)^T [g(\hat{x}(k))]^+ + N\lambda(k)^T |h(\hat{x}(k))|
\geq f(\hat{x}(k)) + \hat{\mu}_\ell(k) [g_\ell(\hat{x}(k))]^+.
\]

Taking limits on both sides of (33) and we obtain:
\[
\liminf_{k \to +\infty} \frac{H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k))}{H(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k))} \geq \limsup_{k \to +\infty} (f(\hat{x}(k)) + \hat{\mu}_\ell(k) [g_\ell(\hat{x}(k))]^+) = +\infty.
\]

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

In both cases, we have \( [g_\ell(\bar{x})]^+ = 0 \) for any \( 1 \leq \ell \leq m \). By utilizing similar arguments, we can further prove that \( |h(\bar{x})| = 0 \). Since \( \bar{x} \in X \), then \( \bar{x} \) is feasible and thus \( f(\bar{x}) \geq p^* \). On the
other hand, since \( \frac{\sum_{\ell=0}^{k-1} \alpha(\ell) \hat{x}(\ell)}{\sum_{\ell=0}^{k-1} \alpha(\ell)} \) is a convex combination of \( \hat{x}(0), \ldots, \hat{x}(k-1) \) and \( \lim_{k \to +\infty} \hat{x}(k) = \bar{x} \), then Claim 3 and (b) in Lemma 5.1 implies that

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

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

\textbf{Claim 5:} It holds that \( \lim_{k \to +\infty} \|y^t(k) - p^*\| = 0 \).

\textit{Proof:} The proof follows the same lines in Claim 2 of Theorem 3.2 and thus omitted here.

\section{VI. DISCUSSION}

In this section, we present some possible extensions and interesting special cases.

A. \textit{Discussion on the periodic strong connectivity assumption in Theorem 3.2}

In the case that \( G(k) \) is undirected, then the periodic strong connectivity assumption 2.4 in Theorem 3.2 can be weakened into:

\textbf{Assumption 6.1 (Eventual strong connectivity):} The undirected graph \((V, \cup_{k \geq s} E(k))\) is connected for all time instant \( s \geq 0 \).

If \( G(k) \) is undirected, the periodic connectivity assumption 2.4 in Theorem 3.2 can also be replaced with the assumption in Proposition 2 of [13]; i.e., for any time instant \( k \geq 0 \), there is an agent connected to all other agents in the undirected graph \((V, \cup_{k \geq s} E(k))\).

B. \textit{Discussion on the Slater’s condition in Theorem 4.2}

If \( g_\ell \) (\( 1 \leq \ell \leq m \)) is linear, then the Slater’s condition 2.1 can be weakened to the following: there exists a relative interior point \( \bar{x} \) of \( X \) such that \( h(\bar{x}) = 0 \) and \( g(\bar{x}) \leq 0 \). For this case, the strong duality and the non-emptyness of the penalty dual optimal set can be ensured by replacing Proposition 5.3.5 [2] with Proposition 5.3.4 [2] in the proofs of Lemma 4.1. In this way, the convergence results of the DPPDS algorithm still hold for the case of linear \( g_\ell \).
C. The special case in the absence of inequality and equality constraints

The following special case of problem (1) is studied in [17]:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x), \quad \text{s.t.} \quad x \in \bigcap_{i=1}^{N} X_i. \quad (34)$$

The following Distributed Primal Subgradient Algorithm is a special case of the DLPDS algorithm, and can be utilized to solve the problem (34):

$$x^i(k+1) = P_{X_i}[v^i_x(k) - \alpha(k)Df_i(v^i_x(k))].$$

**Corollary 6.1:** Consider the problem (34), and let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Consider the sequence \( \{x^i(k)\} \) of the distributed primal subgradient algorithm with initial states \( x^i(0) \in X_i \) and the step-sizes satisfying \( \lim_{k \to +\infty} \alpha(k) = 0, \sum_{k=0}^{+\infty} \alpha(k) = +\infty \), and \( \sum_{k=0}^{+\infty} \alpha(k)^2 < +\infty \).

Then there exists an optimal solution \( x^* \) such that \( \lim_{k \to +\infty} \|x^i(k) - x^*\| = 0 \) all \( i \in V \).

**Proof:** The result is an immediate consequence of Theorem 3.2 with \( g(x) \equiv 0. \)

VII. CONCLUSION

We have studied a multi-agent optimization problem where the agents aim to minimize a sum of local objective functions subject to a global inequality constraint, a global equality constraint and a global constraint set defined as the intersection of local constraint sets. We have considered two cases: the first one in the absence of the equality constraint and the second one with identical local constraint sets. To address these cases, we have introduced two distributed subgradient algorithms which are based on Lagrangian and penalty primal-dual methods, respectively. These two algorithms were shown to asymptotically converge to primal solutions and optimal values.

Our ongoing work includes the application of the proposed algorithms or their extensions to solve network utility maximization problems with coupled and nonconcave utilities.

VIII. APPENDIX

A. Dynamic average consensus algorithms

The following is the vector version of the first-order dynamic average consensus algorithm proposed in [24] with \( x^i(k), \xi^i(k) \in \mathbb{R}^n \):

$$x^i(k+1) = \sum_{j=1}^{N} a^i_{ij}(k)x^j(k) + \xi^i(k). \quad (35)$$
Proposition 8.1: Denote $\Delta \xi^\ell(k) := \max_{i \in V} \xi_i^\ell(k) - \min_{i \in V} \xi_i^\ell(k)$ for $1 \leq \ell \leq n$. Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Assume that $\lim_{k \to +\infty} \Delta \xi^\ell(k) = 0$ for all $1 \leq \ell \leq n$ and all $k \geq 0$. Then $\lim_{k \to +\infty} \|x^i(k) - x^j(k)\| = 0$ for all $i, j \in V$.

B. A property of projection operators

The proof of the following lemma can be found in [2], [3] and [17].

Lemma 8.1: Let $Z$ be a non-empty, closed and convex set in $\mathbb{R}^n$. For any $z \in \mathbb{R}^n$, the following holds for any $y \in Z$: $\|P_Z[z] - y\|^2 \leq \|z - y\|^2 - \|P_Z[z] - z\|^2$.

C. Some properties of the distributed projected subgradient algorithm in [17]

Consider the following distributed projected subgradient algorithm proposed in [17]: $x^i(k + 1) = P_Z[v^i_x(k) - \alpha(k)d_i(k)]$. Denote by $e^i(k) := P_Z[v^i_x(k) - \alpha(k)d_i(k)] - v^i_x(k)$. The following is a slight modification of Lemma 8 and its proof in [17].

Lemma 8.2: Let the non-degeneracy assumption 2.2, the balanced communication assumption 2.3 and the periodic strong connectivity assumption 2.4 hold. Suppose $Z \in \mathbb{R}^n$ is a closed and convex set. 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 $\sum_{k=0}^{+\infty} \alpha(k) \max_{i \in V} \|x^i(k) - \hat{x}(k)\| < +\infty$.

REFERENCES

[1] D. P. Bertsekas and J. N. Tsitsiklis. \textit{Parallel and Distributed Computation: Numerical Methods}. Athena Scientific, 1997.
[2] D.P. Bertsekas. \textit{Convex optimization theory}. Athena Scientific, 2009.
[3] D.P. Bertsekas, A. Nedic, and A. Ozdaglar. \textit{Convex analysis and optimization}. Athena Scientific, 2003.
[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. \textit{IEEE Transactions on Information Theory}, 52(6):2508–2530, 2006.
[5] M. Chiang, S.H. Low, A.R. Calderbank, and J.C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. \textit{Proceedings of IEEE}, 95(1):255–312, January 2007.
[6] J. Cortés. Analysis and design of distributed algorithms for $\chi$-consensus. In \textit{IEEE Conf. on Decision and Control}, pages 3363–3368, San Diego, USA, December 2006.
[7] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

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

[9] A. Kashyap, T. Basar, and R. Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.

[10] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, 1998.

[11] M. Mehyar, D. Spanos, J. Pongsajapan, S. H. Low, and R. M. Murray. Asynchronous distributed averaging on communication networks. *IEEE/ACM Transactions on Networking*, 15(3):512–520, 2007.

[12] C. C. Moallemi and B. Van Roy. Consensus propagation. *IEEE Transactions on Information Theory*, 52(11):4753–4766, 2006.

[13] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.

[14] A. Nedic and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Control and Optimization*, 19(4):1757–1780, 2009.

[15] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

[16] A. Nedic and A. Ozdaglar. Subgradient methods for saddle-point problems. *SIAM Journal on Control and Optimization*, 142(1):205–228, 2009.

[17] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 2009. To appear.

[18] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

[19] A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009.

[20] W. Ren and R. W. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control*. Communications and Control Engineering. Springer, 2008.

[21] R. Srikant. *Mathematics of Internet Congestion Control*. Systems & Control: Foundations & Applications. Birkhäuser, 2005.

[22] S. Sundhar Ram, A. Nedic, and V. V. Veeravalli. Distributed and recursive parameter estimation in parametrized linear state-space models. *IEEE Transactions on Automatic Control*, April 2008. Submitted.

[23] J. N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Massachusetts Institute of Technology, November 1984. Available at http://web.mit.edu/jnt/www/Papers/PhD-84-jnt.pdf.

[24] M. Zhu and S. Martínez. Discrete-time dynamic average consensus. *Automatica*, 2008. Accepted.

[25] M. Zhu and S. Martínez. On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods. *IEEE Transactions on Automatic Control*, 2009. Submitted. Extended version at http://faemino.ucsd.edu/~soniamartinez/papers/data/DCO-extended.pdf.