Chapter 2
Distributed Cooperative Optimization

2.1 Introduction

In this chapter, 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 to all agents. Each node has its own convex constraint set, and the global constraint set is defined as their intersection. This problem is motivated by others in distributed estimation [1, 2], distributed source localization [3], network utility maximization [4], optimal flow control in power systems [5, 6], and optimal shape changes of mobile robots [7]. An important feature of the problem is that the objective and (or) constraint functions depend upon a global decision vector. This requires the design of distributed algorithms where, on 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.

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 based on the characterization of the primal-dual optimal solutions as the saddle points of the Lagrangian function. The DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT 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 DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT ALGORITHM is shown to asymptotically converge to a pair of primal-dual optimal solutions under Slater's condition and the periodic strong connectivity

M. Zhu and S. Martínez, *Distributed Optimization-Based Control of Multi-Agent Networks in Complex Environments*, SpringerBriefs in Control, Automation and Robotics, DOI 10.1007/978-3-319-19072-3_2

assumption. Furthermore, each agent asymptotically agrees on the optimal value by implementing a DISTRIBUTED DYNAMIC AVERAGING ALGORITHM (1.4), 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. 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 DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT 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 DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM asymptotically converges to a primal optimal solution and the optimal value under Slater's condition and the periodic strong connectivity assumption.

2.2 Problem Formulation

Consider a network of agents labeled by $V \triangleq \{1, \dots, N\}$ that can only interact with each other through local communication.

[Objective] We aim to synthesize distributed algorithms which allow the multi-agent group to cooperatively solve the following optimization problem (Fig. 2.1):

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(x), \quad \text{s.t.} \;\; g(x) \le 0, \;\; h(x) = 0, \;\; x \in \cap_{i=1}^{N} X_i, \tag{2.1}$$

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 private convex constraint set of agent i, and x is a global decision vector.

Here we assume that the projection onto the set X_i is easy to compute. Assume that f_i and X_i are private information of agent i, and probably different across agents. The function $g : \mathbb{R}^n \to \mathbb{R}^m$ is known to all the agents with each component g_ℓ, for $\ell \in \{1, \dots, m\}$, being convex. The inequality $g(x) \le 0$ is understood component wise; i.e., $g_\ell(x) \le 0$, for all $\ell \in \{1, \dots, m\}$, and represents a global inequality constraint. The function $h : \mathbb{R}^n \to \mathbb{R}^\nu$, defined as $h(x) \triangleq Ax - b$ with $A \in \mathbb{R}^{\nu \times n}$, represents a global equality constraint, and is known to all the agents. We denote $X \triangleq \cap_{i=1}^{N} X_i$, $f(x) \triangleq \sum_{i=1}^{N} f_i(x)$, and $Y \triangleq \{x \in \mathbb{R}^n \mid g(x) \le 0, \;\; h(x) = 0\}$. We assume that the set of feasible points is nonempty; 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 (2.1) is finite and X^*, the set of primal optimal points, is nonempty. Throughout this chapter, we suppose the following Slater's condition holds:

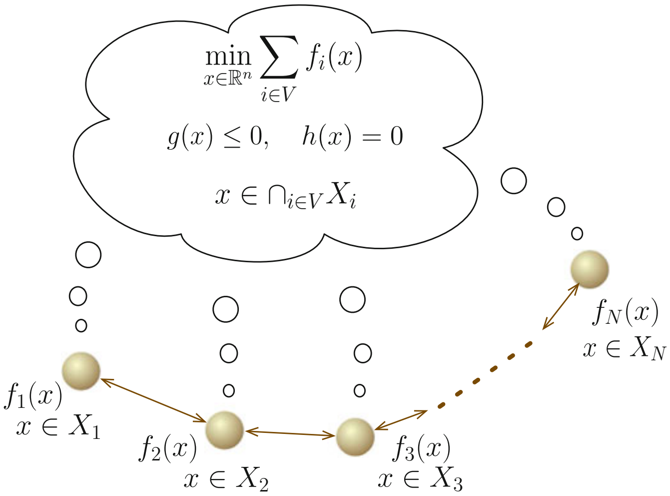

Fig. 2.1 A graphical illustration of problem (2.1)

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 such that $h(\tilde{x}) = 0$ where $\tilde{x}$ is a relative interior point of X; i.e., $\tilde{x} \in X$ and there exists an open sphere S centered at $\tilde{x}$ such that $S \cap \text{aff}(X) \subset X$ with $\text{aff}(X)$ being the affine hull of X.

In this chapter, we will study two particular cases of Problem (2.1): 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 $g(\bar{x}) < 0$. Our techniques rely on duality theory in Sect. 1.3.

2.2.1 Subgradient Notions and Notations

In this chapter, we do not assume the differentiability of the problem functions. At the points where functions are not differentiable, the subgradient plays the role of the gradient. For a given convex function $F : \mathbb{R}^r \to \mathbb{R}$ and a point $x \in \mathbb{R}^r$, a *subgradient* of the function F at x is a vector $\mathscr{D}F(x) \in \mathbb{R}^r$ such that the following subgradient inequality holds for any $y \in \mathbb{R}^r$: $\mathscr{D}F(x)^T(y - x) \le F(y) - F(x)$. Similarly, for a given concave function $G : \mathbb{R}^s \to \mathbb{R}$ and a point $\mu \in \mathbb{R}^s$, a *supgradient* of the function G at μ is a vector $\mathscr{D}G(\mu) \in \mathbb{R}^s$ such that the following supgradient inequality holds for any $\lambda \in \mathbb{R}^s$: $\mathscr{D}G(\mu)^T(\lambda - \mu) \ge G(\lambda) - G(\mu)$.

2.3 Case (i): Absence of Equality Constraint

In this section, we study the case of problem (2.1) where the equality constraint $h(x) = 0$ is absent; i.e.,

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{N} f_i(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad x \in \cap_{i=1}^{N} X_i. \tag{2.2}$$

In the following, we first provide some preliminary results, including a Lagrangian saddle point characterization of the problem (2.2) and a superset containing the Lagrangian dual optimal set of the problem (2.2). After this, the DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT ALGORITHM will be presented along with a summary of its convergence properties.

Overall Strategy and Lagrangian Saddle Point Characterization

First, the problem (2.2) 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 function $q_L : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$, is defined as $q_L(\mu) \triangleq \inf_{x\in X} \mathscr{L}(x, \mu)$, where $\mathscr{L} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ is the Lagrangian $\mathscr{L}(x, \mu) = f(x) + N\mu^T g(x)$. 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 $D_L^* \neq \emptyset$.

As explained in Theorem 1.7, saddle points of the Lagrangian correspond to min-max solutions of the primal and dual problems. Assume for simplicity that the Lagrangian is differentiable and there are no other constraints than the ones included already in the Lagrangian. Then, one can define an associated saddle point dynamics (gradient descent in one argument and gradient ascent in the other) as follows. Let $\mathscr{L}_\mu : \mathbb{R}^n \longrightarrow \mathbb{R}$ be the function $\mathscr{L}_\mu(x) := \mathscr{L}(x, \mu)$, for μ fixed, and $\mathscr{L}_x : \mathbb{R}^n \longrightarrow \mathbb{R}$ be $\mathscr{L}_x(\mu) := \mathscr{L}(x, \mu)$, for x fixed. Then, the continuous-time saddle point dynamics is given as:

$$\begin{aligned} \dot{x}(t) &= -\nabla \mathscr{L}_{\mu(t)}(x(t), \mu(t)), \\ \dot{\mu}(t) &= \nabla \mathscr{L}_{x(t)}(x(t), \mu(t)). \end{aligned} \tag{2.3}$$

If $\mathscr{L}_{\mu(t)}$ is convex and $\mathscr{L}_{x(t)}$ is concave, these dynamics converge to a saddle point of the Lagrangian from any initial and see [8]. The discrete-time counterpart can be generalized for nondifferentiable Lagrangians and to include additional projections over $x \in X$ [9].

We would like to use a distributed discrete-time version of (2.3) for the multi-agent system by defining related and separated problems for each agent, which then are globally coordinated through a consensus algorithm. To do this, note that, while $\mathscr{L}_x$ is naturally separable as a sum of factors $f_i(x) + g(x)$ for each agent, the dual function q_L is not. Then, our strategy will be to define certain sets M_i for each agent that contain the dual solution set, and, which used with a projection operation on M_i, can converge to a saddle point and a min-max solution.

This following lemma presents some preliminary analysis of saddle points toward this objective.

Lemma 2.1 (Preliminary results on saddle points) *Let M be any superset of D_L^*.*

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

Proof (a) It just follows from the definition of saddle point of $\mathscr{L}$ over $X \times M$.
(b) Observe that

$$\sup_{\mu \in \mathbb{R}^m_{\geq 0}} \inf_{x \in X} \mathscr{L}(x, \mu) = \sup_{\mu \in \mathbb{R}^m_{\geq 0}} q_L(\mu) = d_L^*, \quad \inf_{x \in X} \sup_{\mu \in \mathbb{R}^m_{\geq 0}} \mathscr{L}(x, \mu) = \inf_{x \in X \cap Y} f(x) = p^*.$$

Since the Slater's condition 2.1 implies zero duality gap, the Lagrangian minimax equality holds. From Theorem 1.7 it follows that the set of saddle points of $\mathscr{L}$ over $X \times \mathbb{R}^m_{\geq 0}$ is the Cartesian product $X^* \times D_L^*$. Recall that X^* and D_L^* are nonempty, so we can guarantee the existence of the saddle point of $\mathscr{L}$ over $X \times \mathbb{R}^m_{\geq 0}$. Then by (a), we have that (b) holds.
(c) Pick any saddle point (x^*, μ^*) of $\mathscr{L}$ over $X \times \mathbb{R}^m_{\geq 0}$. Since the Slater's condition 2.1 holds, from Theorem 1.7 one can deduce that (x^*, μ^*) is a pair of primal and Lagrangian dual optimal solutions. This implies that

$$d_L^* = \inf_{x \in X} \mathscr{L}(x, \mu^*) \leq \mathscr{L}(x^*, \mu^*) \leq \sup_{\mu \in \mathbb{R}^m_{\geq 0}} \mathscr{L}(x^*, \mu) = p^*.$$

From Theorem 1.7, we have $d_L^* = p^*$. Hence, $\mathscr{L}(x^*, \mu^*) = p^*$. On the other hand, we pick any saddle point $(\check{x}, \check{\mu})$ of $\mathscr{L}$ over $X \times M$. Then for all $x \in X$ and $\mu \in M$, it holds that $\mathscr{L}(\check{x}, \mu) \leq \mathscr{L}(\check{x}, \check{\mu}) \leq \mathscr{L}(x, \check{\mu})$. By Theorem 1.7, then $\mu^* \in D_L^* \subseteq M$. Recall $x^* \in X$, and thus we have $\mathscr{L}(\check{x}, \mu^*) \leq \mathscr{L}(\check{x}, \check{\mu}) \leq \mathscr{L}(x^*, \check{\mu})$. Since $\check{x} \in X$ and $\check{\mu} \in \mathbb{R}^m_{\geq 0}$, we have $\mathscr{L}(x^*, \check{\mu}) \leq \mathscr{L}(x^*, \mu^*) \leq \mathscr{L}(\check{x}, \mu^*)$. Combining the above two relations gives that $\mathscr{L}(\check{x}, \check{\mu}) = \mathscr{L}(x^*, \mu^*) = p^*$. From Remark 1.3 we see that $\mathscr{L}(\check{x}, \check{\mu}) \leq \inf_{x \in X} \mathscr{L}(x, \check{\mu}) = q_L(\check{\mu})$. Since $\mathscr{L}(\check{x}, \check{\mu}) = p^* = d_L^* \geq q_L(\check{\mu})$, then $q_L(\check{\mu}) = d_L^*$ and thus $\check{\mu}$ is a Lagrangian dual optimal solution. ■

Remark 2.1 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 $\ell \in \{1, \dots, m\}$. •

An 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:

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

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 nonempty. The associated dual problem is given by

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

Here, the dual function $q_i : \mathbb{R}^m_{\ge 0} \to \mathbb{R}$ is defined by $q_i(\mu) \triangleq \inf_{x\in X_i} \mathscr{L}_i(x, \mu)$, where $\mathscr{L}_i : \mathbb{R}^n \times \mathbb{R}^m_{\ge 0} \to \mathbb{R}$ is the Lagrangian function of agent i and given by $\mathscr{L}_i(x, \mu) = f_i(x) + \mu^T g(x)$. The corresponding dual optimal value is denoted by d_i^*. In this way, $\mathscr{L}$ is decomposed into a sum of local Lagrangian functions; i.e., $\mathscr{L}(x, \mu) = \sum_{i=1}^N \mathscr{L}_i(x, \mu)$.

Define now the set-valued map $Q : \mathbb{R}^m_{\ge 0} \to 2^{(\mathbb{R}^m_{\ge 0})}$ by $Q(\tilde{\mu}) = \{\mu \in \mathbb{R}^m_{\ge 0} \mid q_L(\mu) \ge q_L(\tilde{\mu})\}$. Additionally, define a function $\gamma : X \to \mathbb{R}$ by $\gamma(x) = \min_{\ell\in\{1,\dots,m\}}\{-g_\ell(x)\}$. Observe that if x is a Slater-vector, then $\gamma(x) > 0$. The following lemma is a direct result of Lemma 1 in [10].

Lemma 2.2 (Boundedness of dual solution sets) *The set $Q(\tilde{\mu})$ is bounded for any $\tilde{\mu} \in \mathbb{R}^m_{\ge 0}$, and, in particular, for any Slater-vector $\bar{x}$, it holds that* $\max_{\mu\in Q(\tilde{\mu})} \|\mu\| \le \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - q_L(\tilde{\mu}))$. □

Notice that $D_L^* = \{\mu \in \mathbb{R}^m_{\ge 0} \mid q_L(\mu) \ge d_L^*\}$. Picking any Slater-vector $\bar{x} \in X$, and letting $\tilde{\mu} = \mu^* \in D_L^*$ in Lemma 2.2 gives that

$$\max_{\mu^*\in D_L^*} \|\mu^*\| \le \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - d_L^*). \tag{2.4}$$

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

$$q_L(\mu) = \inf_{x \in \cap_{i=1}^m X_i} \sum_{i=1}^N (f_i(x) + \mu^T g(x)) \geq \sum_{i=1}^N \inf_{x \in X_i} (f_i(x) + \mu^T g(x)) = \sum_{i=1}^N q_i(\mu). \tag{2.5}$$

Since $d_L^* \geq q_L(\tilde{\mu})$, it follows from (2.4) and (2.5) that

$$\begin{aligned} \max_{\mu^* \in D_L^*} \|\mu^*\| &\leq \frac{1}{\gamma(\bar{x})}(f(\bar{x}) - q_L(\tilde{\mu})) \leq \frac{1}{\gamma(\bar{x})}\left(f(\bar{x}) - \sum_{i=1}^N q_i(\tilde{\mu})\right) \\ &\leq \frac{N}{\gamma(\bar{x})} \max_{i \in V}\{f_i(\bar{x}) - q_i(\tilde{\mu})\} = r(\bar{x}, \tilde{\mu}). \end{aligned}$$

Hence, we have $D_L^* \subseteq \bar{M}_i(\bar{x}, \tilde{\mu})$ for all $i \in V$.

Note that in order to compute $\bar{M}_i(\bar{x}, \tilde{\mu})$, 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 $\bar{M}_i(\bar{x}, \tilde{\mu})$.

Initially, each agent i chooses a common value $\tilde{\mu} \in \mathbb{R}^m_{\geq 0}$; e.g., $\tilde{\mu} = 0$, and computes two positive constants $b_i(0)$ and $c_i(0)$ such that $b_i(0) \geq \sup_{x \in J_i}\{f_i(x) - q_i(\tilde{\mu})\}$ and $c_i(0) \leq \min_{1 \leq \ell \leq m} \inf_{x \in J_i}\{-g_\ell(x)\}$ where $J_i \triangleq \{x \in X_i \mid g(x) < 0\}$.

At every time $k \geq 0$, each agent i updates its estimates by:

$$b_i(k+1) = \max_{j \in \mathcal{N}_i(k) \cup \{i\}} b_j(k), \quad c_i(k+1) = \min_{j \in \mathcal{N}_i(k) \cup \{i\}} c_j(k).$$

We denote $b^* \triangleq \max_{j \in V} b_j(0)$, $c^* \triangleq \min_{j \in V} c_j(0)$ for all $k \geq (N-1)B$, and $M^{[i]}(\tilde{\mu}) \triangleq \{\mu \in \mathbb{R}^m_{\geq 0} \mid \|\mu\| \leq \frac{Nb^*}{c^*} + \theta_i\}$, $J \triangleq \{x \in X \mid g(x) < 0\}$.

Lemma 2.3 (Convergence of the DISTRIBUTED SLATER- VECTOR COMPUTATION ALGORITHM) *Assume that the periodical strong connectivity Assumption 1.3 holds. Consider the sequences of $\{b_i(k)\}$ and $\{c_i(k)\}$ generated by the* DISTRIBUTED SLATER- VECTOR COMPUTATION ALGORITHM. *It holds that after at most $(N-1)B$ steps, all the agents reach the consensus, i.e., $b_i(k) = b^*$ and $c_i(k) = c^*$ for all $k \geq (N-1)B$. Furthermore, we have $M^{[i]}(\tilde{\mu}) \supseteq \bar{M}_i(\bar{x}, \tilde{\mu})$ for $i \in V$.*

Proof It is not difficult to verify achieving max-consensus by the periodical strong connectivity Assumption 1.3. Note that $J \subseteq J_i$, $\forall i \in V$. Hence, we have

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

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

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

$$r(\bar{x}, \tilde{\mu}) \leq \frac{N \sup_{x \in J} \max_{i \in V}\{f_i(x) - q_i(\tilde{\mu})\}}{\inf_{x \in J} \min_{1 \leq \ell \leq m}\{-g_\ell(x)\}} \leq \frac{Nb^*}{c^*}.$$

The desired result immediately follows. ∎

From Lemma 2.3 and the fact that $D_L^* \subseteq \bar{M}_i(\bar{x}, \tilde{\mu})$, we can see that the set of $M(\tilde{\mu}) \triangleq \cap_{i=1}^N M^{[i]}(\tilde{\mu})$ contains D_L^*. In addition, $M^{[i]}(\tilde{\mu})$ and $M(\tilde{\mu})$ are nonempty, compact, and convex. To simplify the notations, we will use the shorthands $M_i \triangleq M^{[i]}(\tilde{\mu})$ and $M \triangleq M(\tilde{\mu})$.

Convexity of the Lagrangian Function

For each $\mu \in \mathbb{R}^m_{\geq 0}$, we define the function $\mathscr{L}_{i\mu} : \mathbb{R}^n \to \mathbb{R}$ as $\mathscr{L}_{i\mu}(x) := \mathscr{L}_i(x, \mu)$. Note that $\mathscr{L}_{i\mu}$ is convex since it is a nonnegative weighted sum of convex functions. For each $x \in \mathbb{R}^n$, we define the function $\mathscr{L}_{ix} : \mathbb{R}^m_{\geq 0} \to \mathbb{R}$ as $\mathscr{L}_{ix}(\mu) := \mathscr{L}_i(x, \mu)$. It is easy to check that $\mathscr{L}_{ix}$ is a concave (actually affine) function. Then the Lagrangian function $\mathscr{L}$ is the sum of a collection of convex–concave local functions.

2.3.1 *The* Distributed Lagrangian Primal-Dual Subgradient Algorithm

Here, we introduce the Distributed Lagrangian Primal- Dual Subgradient Algorithm to find a saddle point of the Lagrangian function $\mathscr{L}$ 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 2.1.

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 $\mathscr{L}$ 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 $v_x^i(k)$ (resp. $v_\mu^i(k)$) 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 $\mathscr{L}_i$, 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 $v_y^i(k)$ 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 Distributed Lagrangian Primal- Dual Subgradient 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. Agent i then chooses any initial state $x^i(0) \in X_i$, $\mu^i(0) \in \mathbb{R}^m_{\geq 0}$, and $y^i(1) = Nf_i(x^i(0))$.

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:

$$v_x^i(k) = \sum_{j=1}^{N} a_j^i(k)x^j(k), \quad v_\mu^i(k) = \sum_{j=1}^{N} a_j^i(k)\mu^j(k), \quad v_y^i(k) = \sum_{j=1}^{N} a_j^i(k)y^j(k),$$
$$x^i(k+1) = P_{X_i}[v_x^i(k) - \alpha(k)\mathscr{D}_x^i(k)], \quad \mu^i(k+1) = P_{M_i}[v_\mu^i(k) + \alpha(k)\mathscr{D}_\mu^i(k)],$$
$$y^i(k+1) = v_y^i(k) + N(f_i(x^i(k)) - f_i(x^i(k-1))),$$

where P_{X_i} (resp. P_{M_i}) is the projection operator onto the set X_i (resp. M_i), the scalars $a_j^i(k)$ are nonnegative weights and the scalars $\alpha(k) > 0$ are step-sizes.[1] We use the shorthands $\mathscr{D}_x^i(k) \equiv \mathscr{D}\mathscr{L}_{iv_\mu^i(k)}(v_x^i(k))$, and $\mathscr{D}_\mu^i(k) \equiv \mathscr{D}\mathscr{L}_{iv_x^i(k)}(v_\mu^i(k))$.

The following theorem summarizes the convergence properties of the DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT ALGORITHM where it is guaranteed that agents asymptotically agree upon a pair of primal-dual optimal solutions.

Theorem 2.1 (Convergence of the DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT ALGORITHM) *Consider the optimization problem* (2.2). *Let the nondegeneracy Assumption 1.1, the double stochasticity Assumption 1.2, and the periodic strong connectivity Assumption 1.3 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* $(x^*, \mu^*) \in X^* \times D_L^*$ *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,* $\lim_{k\to+\infty} \|y^i(k) - p^*\| = 0$, *for all* $i \in V$.

2.3.2 *A Numerical Example for the* DISTRIBUTED LAGRANGIAN PRIMAL-DUAL SUBGRADIENT ALGORITHM

In order to illustrate the performance of the DISTRIBUTED LAGRANGIAN PRIMAL-DUAL SUBGRADIENT ALGORITHM, we here study a numerical example of network utility maximization, e.g., in [4]. Consider five agents and one link where each agent sends data through the link at a rate of z_i, and the link capacity is 5. The global decision vector $x := (z_1 \ z_2 \ z_3 \ z_4 \ z_5)^T \in \mathbb{R}^5$ is the resource allocation vector. Each agent i is associated with a concave utility function $f_i(z_i) := \sqrt{z_i}$, representing the utility

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

agent i obtained through sending data at a rate of z_i. Agents aim to maximize the network utility and it can be formulated as follows:

$$\min_{x\in\mathbb{R}^5} \sum_{i\in V} -\sqrt{z_i} \quad \text{s.t.} \quad z_1+z_2+z_3+z_4+z_5 \le 5, \quad x \in \cap_{i\in V} X_i, \tag{2.6}$$

where local constraint sets X_i are given by:

$$\begin{aligned}
X_1 &:= [0.5,\ \ 5.5] \times [0.5,\ \ 5.5] \times [0.5,\ \ 5.5] \times [0.5,\ \ 5.5] \times [0.5,\ \ 5.5],\\
X_2 &:= [0.55,\ \ 5.25] \times [0.55,\ \ 5.25] \times [0.55,\ \ 5.25] \times [0.55,\ \ 5.25] \times [0.55,\ \ 5.25],\\
X_3 &:= [0.5,\ \ 6] \times [0.5,\ \ 6] \times [0.5,\ \ 6] \times [0.5,\ \ 6] \times [0.5,\ \ 6],\\
X_4 &:= [0.5,\ \ 5] \times [0.5,\ \ 5] \times [0.5,\ \ 5] \times [0.5,\ \ 5] \times [0.5,\ \ 5],\\
X_5 &:= [0.525,\ \ 5.75] \times [0.525,\ \ 5.75] \times [0.525,\ \ 5.75] \times [0.525,\ \ 5.75] \times [0.525,\ \ 5.75].
\end{aligned}$$

One can verify that the optimal solution is $[1 \quad 1 \quad 1 \quad 1 \quad 1]^T$. We use the DISTRIBUTED LAGRANGIAN PRIMAL-DUAL SUBGRADIENT ALGORITHM to solve problem (2.6) by choosing step-size $\alpha(k) = \frac{1}{k+1}$. Figures 2.2 and 2.3 show the simulation results of the DISTRIBUTED LAGRANGIAN PRIMAL-DUAL SUBGRADIENT ALGORITHM, demonstrating that the agents take 10^4 iterates to agree upon value 1 for z_1 and z_2.

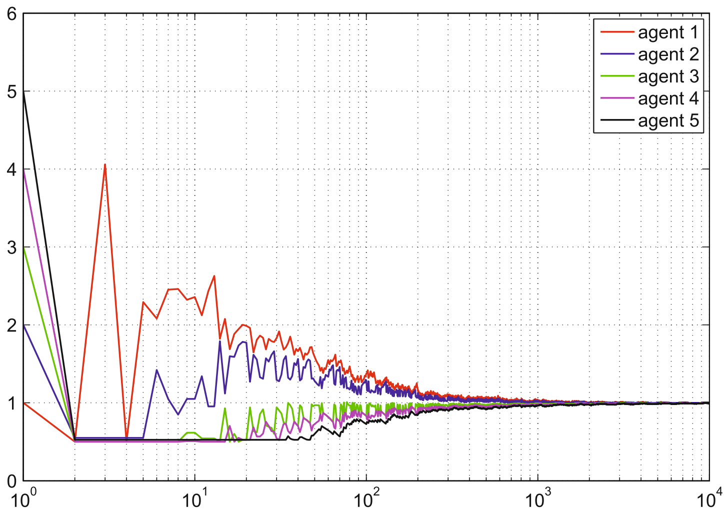

Fig. 2.2 The estimates on z_1 generated by different agents in the DISTRIBUTED LAGRANGIAN PRIMAL-DUAL SUBGRADIENT ALGORITHM (DLPDS)

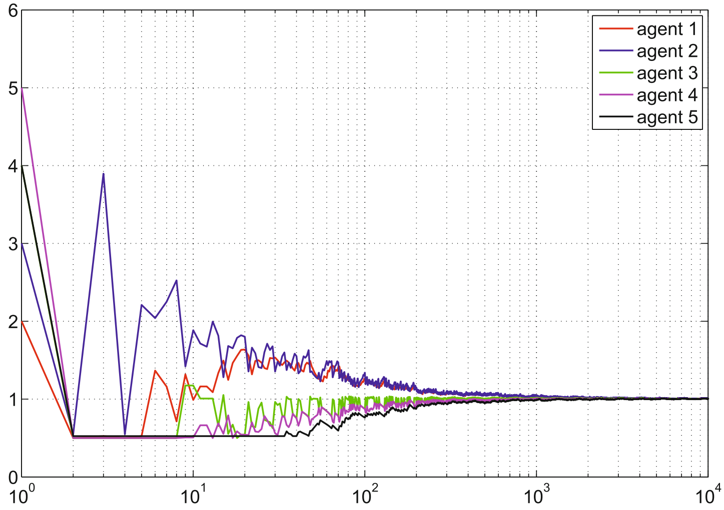

Fig. 2.3 The estimates on z_2 generated by different agents in the DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT ALGORITHM (DLPDS)

2.4 Case (ii): Identical Local Constraint Sets

In the previous section, we study the case where the equality constraint is absent in problem (2.1). Here, we turn our attention to another case of problem (2.1), where $h(x) = 0$ is taken into account but we require that local constraint sets are identical; i.e., $X_i = X$ for all $i \in V$. We first adopt a penalty relaxation formulation and provide a penalty saddle point characterization of the primal problem (2.1) with $X_i = X$. We then introduce the DISTRIBUTED PENALTY PRIMAL- DUAL SUBGRADIENT ALGORITHM, followed by its convergence properties.

Overall Strategy and a Penalty Saddle Point Characterization

As in the previous section, our strategy will be to define an appropriate dynamics to converge to a saddle point or a min-max solution of the problem. However, to deal with the equality constraint, we will employ a penalty function, which includes positive terms penalizing the violation of the equality and inequality constraints. The identical local constraint sets will also help in guaranteeing the convergence of the method. More precisely, consider the following.

The primal problem (2.1) with $X_i = X$ is trivially equivalent to the following:

$$\min_{x\in\mathbb{R}^n} f(x), \quad \text{s.t.} \quad Ng(x) \leq 0, \quad Nh(x) = 0, \quad x \in X, \tag{2.7}$$

with associated penalty dual problem given by

$$\max_{\mu\in\mathbb{R}^m, \lambda\in\mathbb{R}^\nu} q_P(\mu, \lambda), \quad \text{s.t.} \quad \mu \geq 0, \quad \lambda \geq 0. \tag{2.8}$$

Here, the penalty dual function, $q_P : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$, is defined by

$$q_P(\mu, \lambda) \triangleq \inf_{x \in X} \mathscr{H}(x, \mu, \lambda),$$

where $\mathscr{H} : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ is the *penalty function* given by $\mathscr{H}(x, \mu, \lambda) = f(x) + N\mu^T[g(x)]^+ + N\lambda^T|h(x)|$. 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 $\mathscr{H}_i(x, \mu, \lambda) : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ for each agent i as follows: $\mathscr{H}_i(x, \mu, \lambda) = f_i(x) + \mu^T[g(x)]^+ + \lambda^T|h(x)|$. In this way, we have that $\mathscr{H}(x, \mu, \lambda) = \sum_{i=1}^N \mathscr{H}_i(x, \mu, \lambda)$. 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 2.4 (Strong duality and nonemptiness of the penalty dual optimal set) *The values of p^* and d^*_P coincide, and D^*_P is nonempty.*

Proof Consider the auxiliary Lagrangian function $\mathscr{L}_a : \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu \to \mathbb{R}$ given by $\mathscr{L}_a(x, \mu, \lambda) = f(x) + N\mu^T g(x) + N\lambda^T h(x)$, with the associated dual problem defined by

$$\max_{\mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^\nu} q_a(\mu, \lambda), \quad \text{s.t.} \quad \mu \geq 0. \tag{2.9}$$

Here, the dual function, $q_a : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu \to \mathbb{R}$, is defined by

$$q_a(\mu, \lambda) \triangleq \inf_{x \in X} \mathscr{L}_a(x, \mu, \lambda).$$

The dual optimal value of problem (2.9) is denoted by d^*_a and the set of dual optimal solutions is denoted D^*_a. Since X is convex, f and g_ℓ, 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 [11] 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

$$\begin{aligned} d^*_a &= q_a(\mu^*, \lambda^*) = \inf_{x \in X}\{f(x) + N(\mu^*)^T g(x) + N(\lambda^*)^T h(x)\} \\ &\leq \inf_{x \in X}\{f(x) + N(\mu^*)^T[g(x)]^+ + N|\lambda^*|^T|h(x)|\} = q_P(\mu^*, |\lambda^*|) \leq d^*_P. \end{aligned} \tag{2.10}$$

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 $q_P(\mu, \lambda) \leq \mathscr{H}(x^*, \mu, \lambda) = f(x^*) = p^*$ 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 $(\mu^*, \lambda^*) \in D^*_a$. From (2.10) 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 1.7 to penalty functions.

Theorem 2.2 (Penalty Saddle point Theorem) *The pair of* (x^*, μ^*, λ^*) *is a saddle point of the penalty function* $\mathscr{H}$ *over* $X \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$ *if and only if it is a pair of primal and penalty dual optimal solutions and the following* penalty minimax equality *holds:*

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

Proof The proof is analogous to that of Proposition 6.2.4 in [12], and for the sake of completeness, we provide the details here. It follows from Proposition 2.6.1 in [12] that (x^*, μ^*, λ^*) is a saddle point of $\mathscr{H}$ over $X \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$ if and only if the penalty minimax equality holds and the following conditions are satisfied:

$$\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathscr{H}(x^*, \mu, \lambda) = \min_{x\in X}\{\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathscr{H}(x, \mu, \lambda)\}, \tag{2.11}$$

$$\inf_{x\in X} \mathscr{H}(x, \mu^*, \lambda^*) = \max_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \{\inf_{x\in X} \mathscr{H}(x, \mu, \lambda)\}. \tag{2.12}$$

Notice that $\inf_{x\in X} \mathscr{H}(x, \mu, \lambda) = q_P(\mu, \lambda)$; and if $x \in Y$, then the following holds:

$$\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathscr{H}(x, \mu, \lambda) = f(x),$$

otherwise, $\sup_{(\mu,\lambda)\in\mathbb{R}^m_{\geq 0}\times\mathbb{R}^\nu_{\geq 0}} \mathscr{H}(x, \mu, \lambda) = +\infty$. Hence, the penalty minimax equality is equivalent to $d_P^* = p^*$. Condition (2.11) is equivalent to the fact that x^* is primal optimal, and condition (2.12) is equivalent to (μ^*, λ^*) being a penalty dual optimal solution. ■

Convexity of the Penalty Function

Since g_ℓ is convex and $[\cdot]^+$ is convex and nondecreasing, $[g_\ell(x)]^+$ is convex in x for each $\ell \in \{1, \dots, m\}$. Denote $A \triangleq (a_1^T, \dots, a_\nu^T)^T$. 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 $\ell \in \{1, \dots, \nu\}$.

We denote $w \triangleq (\mu, \lambda)$. For each $w \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$, we define the function $\mathscr{H}_{iw} : \mathbb{R}^n \to \mathbb{R}$ as $\mathscr{H}_{iw}(x) := \mathscr{H}_i(x, w)$. Note that $\mathscr{H}_{iw}(x)$ 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 $\mathscr{H}_{ix} : \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0} \to \mathbb{R}$ as $\mathscr{H}_{ix}(w) := \mathscr{H}_i(x, w)$. It is easy to check that $\mathscr{H}_{ix}(w)$ is concave (actually affine) in w. Then the penalty function $\mathscr{H}(x, w)$ is the sum of convex–concave local functions.

Remark 2.2 The Lagrangian relaxation does not fit into our approach here since the Lagrangian function is not convex in x by allowing λ entries to be negative. •

2.4.1 *The* Distributed Penalty Primal-Dual Subgradient Algorithm

We now proceed to devise the Distributed Penalty Primal-Dual Subgradient Algorithm, which is based on the penalty saddle point Theorem 2.2, to find the optimal value and a primal optimal solution to the primal problem (2.1) with $X_i = X$. The main steps of the Distributed Penalty Primal-Dual Subgradient Algorithm are described as follows.

Initially, agent i chooses any initial state $x^i(0) \in X$, $\mu^i(0) \in \mathbb{R}^m_{\geq 0}$, $\lambda^i(0) \in \mathbb{R}^\nu_{\geq 0}$, and $y^i(1) = N f_i(x^i(0))$. After this, at every $k \geq 0$, agent i computes the following convex combinations:

$$v_x^i(k) = \sum_{j=1}^N a_j^i(k) x^j(k), \quad v_y^i(k) = \sum_{j=1}^N a_j^i(k) y^j(k),$$
$$v_\mu^i(k) = \sum_{j=1}^N a_j^i(k) \mu^j(k), \quad v_\lambda^i(k) = \sum_{j=1}^N a_j^i(k) \lambda^j(k),$$

and generates $x^i(k+1)$, $y^i(k+1)$, $\mu^i(k+1)$, and $\lambda^i(k+1)$ according to the following:

$$\begin{aligned}
&x^i(k+1) = P_X[v_x^i(k) - \alpha(k)\mathcal{S}_x^i(k)],\\
&y^i(k+1) = v_y^i(k) + N(f_i(x^i(k)) - f_i(x^i(k-1))),\\
&\mu^i(k+1) = v_\mu^i(k) + \alpha(k)[g(v_x^i(k))]^+, \quad \lambda^i(k+1) = v_\lambda^i(k) + \alpha(k)|h(v_x^i(k))|,
\end{aligned}$$

where P_X is the projection operator onto the set X, the scalars $a_j^i(k)$ are nonnegative weights, and the positive scalars $\{\alpha(k)\}$ are step-sizes.[2] The vector

$$\mathcal{S}_x^i(k) \triangleq \mathcal{D} f_i(v_x^i(k)) + \sum_{\ell=1}^m v_\mu^i(k)_\ell \mathcal{D}[g_\ell(v_x^i(k))]^+ + \sum_{\ell=1}^\nu v_\lambda^i(k)_\ell \mathcal{D}|h_\ell|(v_x^i(k))$$

is a subgradient of $\mathcal{H}_{iw^i(k)}(x)$ at $x = v_x^i(k)$ where $w^i(k) \triangleq (v_\mu^i(k), v_\lambda^i(k))$.

Given a step-size sequence $\{\alpha(k)\}$, we define the sum over $[0, k]$ by $s(k) \triangleq \sum_{\ell=0}^k \alpha(\ell)$, which should satisfy the following; see Remark 2.4 on how to define such a sequence.

Assumption 2.2 (*Step-size assumption*) The step-sizes satisfy

$$\lim_{k\to+\infty} \alpha(k) = 0, \quad \sum_{k=0}^{+\infty} \alpha(k) = +\infty, \quad \sum_{k=0}^{+\infty} \alpha(k)^2 < +\infty,$$

[2] Each agent i executes the update law of $y^i(k)$ for $k \geq 1$.

$$\lim_{k\to+\infty} \alpha(k+1)s(k) = 0, \quad \sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k) < +\infty, \quad \sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k)^2 < +\infty.$$

The following theorem is the main result of this section, characterizing the convergence properties of the DISTRIBUTED PENALTY PRIMAL- DUAL SUBGRADIENT ALGORITHM, where an optimal solution and the optimal value are asymptotically achieved.

Theorem 2.3 (Convergence of the DISTRIBUTED PENALTY PRIMAL- DUAL SUBGRADIENT ALGORITHM) *Consider the problem* (2.1) *with* $X_i = X$. *Let the non-degeneracy Assumption 1.1, the double stochasticity Assumption 1.2, and the periodic strong connectivity Assumption 1.3 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 2.2. 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 2.3 Observe that $\mu^i(k) \geq 0$, $\lambda^i(k) \geq 0$ and $v_x^i(k) \in X$ (due to the fact that X is convex). Furthermore, $([g(v_x^i(k))]^+, |h(v_x^i(k))|)$ is a supgradient of $\mathscr{H}_{iv_x^i(k)}(w^i(k))$; i.e., the following *penalty supgradient inequality* holds for any $\mu \in \mathbb{R}^m_{\geq 0}$ and $\lambda \in \mathbb{R}^\nu_{\geq 0}$:

$$\begin{aligned}&([g(v_x^i(k))]^+)^T(\mu - v_\mu^i(k)) + |h(v_x^i(k))|^T(\lambda - v_\lambda^i(k))\\ &\geq \mathscr{H}_i(v_x^i(k), \mu, \lambda) - \mathscr{H}_i(v_x^i(k), v_\mu^i(k), v_\lambda^i(k)).\end{aligned} \tag{2.13}$$

•

Remark 2.4 A step-size sequence that satisfies the step-size Assumption 2.2 is the harmonic series $\{\alpha(k) = \frac{1}{k+1}\}_{k\in\mathbb{Z}_{\geq 0}}$. It is obvious that $\lim_{k\to+\infty} \dfrac{1}{k+1} = 0$, and well-known that $\sum_{k=0}^{+\infty} \frac{1}{k+1} = +\infty$ and $\sum_{k=0}^{+\infty} \frac{1}{(k+1)^2} < +\infty$. We now proceed to check the property of $\lim_{k\to+\infty} \alpha(k+1)s(k) = 0$. For any $k \geq 1$, there is an integer $n \geq 1$ such that $2^{n-1} \leq k < 2^n$. It holds that

$$\begin{aligned}s(k) &\leq s(2^n) = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \cdots + \left(\frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^n}\right)\\ &\leq 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{3}\right) + \cdots + \left(\frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^{n-1}+1}\right)\\ &\leq 1 + 1 + 1 + \cdots + 1 = n \leq \log_2 k + 1.\end{aligned}$$

Then we have $\limsup_{k\to+\infty} \frac{s(k)}{k+2} \leq \lim_{k\to+\infty} \frac{\log_2 k+1}{k+2} = 0$. Obviously, $\liminf_{k\to+\infty} \frac{s(k)}{k+2} \geq 0$. Then we have the property of $\lim_{k\to+\infty} \alpha(k+1)s(k) = 0$. Since $\log_2 k \leq (\log_2 k)^2 < (k+2)^{\frac{1}{2}}$, then

$$\sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k)^2 \leq \sum_{k=0}^{+\infty} \frac{(\log_2 k+1)^2}{(k+2)^2} = \sum_{k=0}^{+\infty} \left(\frac{(\log_2 k)^2}{(k+2)^2} + \frac{2\log_2 k}{(k+2)^2} + \frac{1}{(k+2)^2} \right)$$
$$\leq \sum_{k=0}^{+\infty} \frac{1}{(k+2)^{\frac{3}{2}}} + \sum_{k=0}^{+\infty} \frac{2}{(k+2)^{\frac{3}{2}}} + \sum_{k=0}^{+\infty} \frac{1}{(k+2)^2} < +\infty.$$

Additionally, we have $\sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k) \leq \sum_{k=0}^{+\infty} \alpha(k+1)^2 s(k)^2 < +\infty$. •

2.4.2 *A Numerical Example for the* DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM

Consider a network with five agents and their objective functions are defined as

$$f_1(x) := \frac{1}{5}\big((a-5)^2 + (b-2.5)^2 + (c-5)^2 + (d+2.5)^2 + (e+5)^2\big),$$
$$f_2(x) := \frac{1}{5}\big((a-2.5)^2 + (b-5)^2 + (c+2.5)^2 + (d+5)^2 + (e-5)^2\big),$$
$$f_3(x) := \frac{1}{5}\big((a-5)^2 + (b+2.5)^2 + (c+5)^2 + (d-5)^2 + (e-2.5)^2\big),$$
$$f_4(x) := \frac{1}{5}\big((a+2.5)^2 + (b+5)^2 + (c-5)^2 + (d-2.5)^2 + (e-5)^2\big),$$
$$f_5(x) := \frac{1}{5}\big((a+5)^2 + (b-5)^2 + (c-2.5)^2 + (d-5)^2 + (e+2.5)^2\big),$$

where the global decision vector $x := [a \quad b \quad c \quad d \quad e]^T \in \mathbb{R}^5$. The global equality constraint function is given by $h(x) := a + b + c + d + e - 5$, and the global constraint set is as follows: $X := [-5, \quad 5] \times [-5, \quad 5] \times [-5, \quad 5] \times [-5, \quad 5] \times [-5, \quad 5]$. Consider the optimization as follows:

$$\min_{x\in\mathbb{R}^5} \sum_{i\in V} f_i(x), \quad \text{s.t.} \quad h(x) = 0, \quad x \in X.$$

One can verify that the optimal solution is $[1 \quad 1 \quad 1 \quad 1 \quad 1]^T$. We employ the DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM to solve the above optimization problem with the step-size $\alpha(k) = \frac{1}{k+1}$. Its simulation results are included in Figs. 2.4 and 2.5. Observe that the estimates of a and b generated by different agents asymptotically achieve value 1.

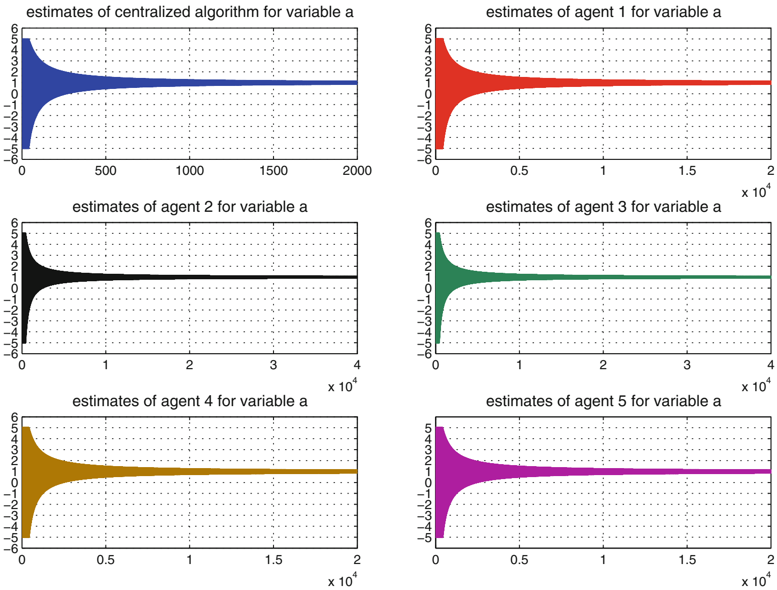

Fig. 2.4 The estimates on a generated by different agents in the DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM

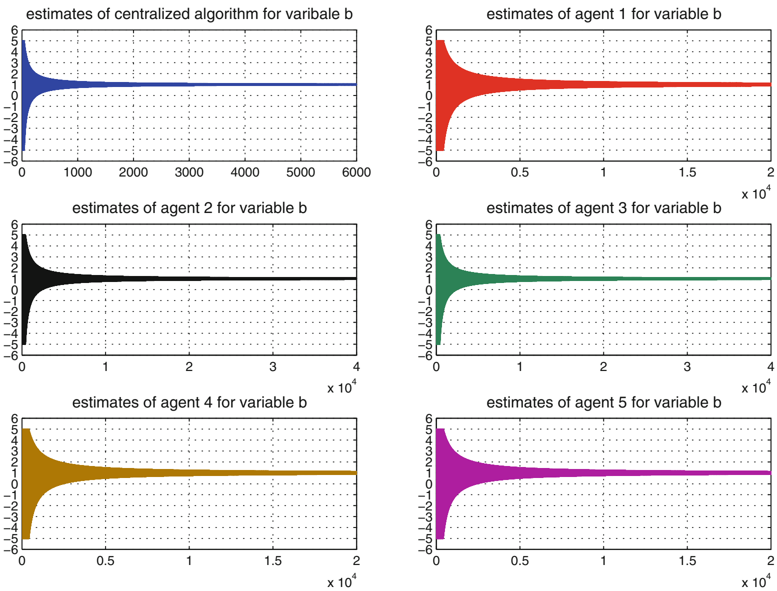

Fig. 2.5 The estimates on b generated by different agents in the DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM

2.5 Appendix

We next provide the proofs for the main results, Theorems 2.1 and 2.3, of this chapter. Before doing that, let us state an instrumental result as follows. Consider the following distributed projected subgradient algorithm proposed in [13]: $x^i(k+1) = P_Z[v_x^i(k) - \alpha(k)d_i(k)]$. Denote by $e^i(k) := P_Z[v_x^i(k) - \alpha(k)d_i(k)] - v_x^i(k)$. The following is a slight modification of Lemma 8 and its proof in [13].

Lemma 2.5 *Let the nondegeneracy Assumption 1.1, the double stochasticity Assumption 1.2, and the periodic strong connectivity Assumption 1.3 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*

$$\|x^i(k) - \hat{x}(k)\| \le N\gamma \sum_{\tau=0}^{k-1} \beta^{k-\tau} \{\alpha(\tau)\|d_i(\tau)\|$$
$$+ \|e^i(\tau) + \alpha(\tau)d_i(\tau)\|\} + N\gamma\beta^{k-1} \sum_{i=0}^{N} \|x^i(0)\|.$$

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

We start our analysis on Theorems 2.1 and 2.3 by providing the properties of the sequences weighted by $\{\alpha(k)\}$.

Lemma 2.6 (Convergence of weighted sequences) *Let $K \ge 0$. Consider the sequence $\{\delta(k)\}$ defined by $\delta(k) \triangleq \frac{\sum_{\ell=K}^{k-1} \alpha(\ell)\rho(\ell)}{\sum_{\ell=K}^{k-1} \alpha(\ell)}$ where $k \ge K + 1$, $\alpha(k) > 0$ 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 $k_1 \ge K$ such that $\rho(k) \ge \Pi$ for all $k \ge k_1$. Then the following holds for all $k \ge k_1 + 1$:

$$\delta(k) \ge \frac{1}{\sum_{\ell=K}^{k-1} \alpha(\ell)} \left(\sum_{\ell=K}^{k_1-1} \alpha(\ell)\rho(\ell) + \sum_{\ell=k_1}^{k-1} \alpha(\ell)\Pi \right)$$
$$= \Pi + \frac{1}{\sum_{\ell=K}^{k-1} \alpha(\ell)} \left(\sum_{\ell=K}^{k_1-1} \alpha(\ell)\rho(\ell) - \sum_{\ell=K}^{k_1-1} \alpha(\ell)\Pi \right).$$

Take the limit on k in the above estimate and we have $\liminf_{k\to+\infty} \delta(k) \geq \Pi$. Since Π is arbitrary, then $\lim_{k\to+\infty} \delta(k) = +\infty$.

(b) For any $\varepsilon > 0$, there exists $k_2 \geq K$ such that $\|\rho(k) - \rho^*\| \leq \varepsilon$ for all $k \geq k_2 + 1$. Then we have

$$\|\delta(k) - \rho^*\| = \|\frac{\sum_{\tau=K}^{k-1} \alpha(\tau)(\rho(\tau) - \rho^*)}{\sum_{\tau=K}^{k-1} \alpha(\tau)}\|$$

$$\leq \frac{1}{\sum_{\tau=K}^{k-1} \alpha(\tau)} \left(\sum_{\tau=K}^{k_2-1} \alpha(\tau)\|\rho(\tau) - \rho^*\| + \sum_{\tau=k_2}^{k-1} \alpha(\tau)\varepsilon \right) \leq \frac{\sum_{\tau=K}^{k_2-1} \alpha(\tau)\|\rho(\tau) - \rho^*\|}{\sum_{\tau=K}^{k-1} \alpha(\tau)} + \varepsilon.$$

Take the limit on k in the above estimate and we have $\limsup_{k\to+\infty} \|\delta(k) - \rho^*\| \leq \varepsilon$. Since ε is arbitrary, then $\lim_{k\to+\infty} \|\delta(k) - \rho^*\| = 0$. ∎

2.5.1 *Convergence Analysis of the* Distributed Lagrangian Primal-Dual Subgradient Algorithm

We now proceed to show Theorem 2.1. To do that, we first rewrite the Distributed Lagrangian Primal- Dual Subgradient Algorithm into the following form:

$$x^i(k+1) = v_x^i(k) + e_x^i(k), \quad \mu^i(k+1) = v_\mu^i(k) + e_\mu^i(k), \quad y^i(k+1) = v_y^i(k) + u^i(k),$$

where $e_x^i(k)$ and $e_\mu^i(k)$ are projection errors described by

$$e_x^i(k) \triangleq P_{X_i}[v_x^i(k) - \alpha(k)\mathscr{D}_x^i(k)] - v_x^i(k),$$
$$e_\mu^i(k) \triangleq P_{M_i}[v_\mu^i(k) + \alpha(k)\mathscr{D}_\mu^i(k)] - v_\mu^i(k),$$

and $u^i(k) \triangleq N(f_i(x^i(k)) - f_i(x^i(k-1)))$ 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 Chap. 1. This observation allows us to divide the analysis of the Distributed Lagrangian Primal- Dual Subgradient Algorithm in two steps. First, we show all the estimates asymptotically achieve consensus by utilizing the property that the local errors and inputs are diminishing. Second, we further show that the consensus vectors coincide with a pair of primal and Lagrangian dual optimal solutions and the optimal value.

Lemma 2.7 (Lipschitz continuity of $\mathscr{L}_i$) *Consider $\mathscr{L}_{i\mu}$ and $\mathscr{L}_{ix}$. Then there are $L > 0$ and $R > 0$ such that $\|\mathscr{D}\mathscr{L}_{i\mu}(x)\| \leq L$ and $\|\mathscr{D}\mathscr{L}_{ix}(\mu)\| \leq R$ for each pair of $x \in \mathrm{co}(\cup_{i=1}^N X_i)$ and $\mu \in \mathrm{co}(\cup_{i=1}^N M_i)$. Furthermore, for each $\mu \in \mathrm{co}(\cup_{i=1}^N M_i)$, the function $\mathscr{L}_{i\mu}$ is Lipschitz continuous with Lipschitz constant L over $\mathrm{co}(\cup_{i=1}^N X_i)$, and for each $x \in \mathrm{co}(\cup_{i=1}^N X_i)$, the function $\mathscr{L}_{ix}$ is Lipschitz continuous with Lipschitz constant R over $\mathrm{co}(\cup_{i=1}^N M_i)$.*

Proof Observe that $\mathscr{D}\mathscr{L}_{i\mu} = \mathscr{D}f_i + \mu^T\mathscr{D}g$ and $\mathscr{D}\mathscr{L}_{ix} = g$. Since f_i and g_ℓ are convex, it follows from Proposition 5.4.2 in [11] that ∂f_i and ∂g_ℓ are bounded over the compact $\mathrm{co}(\cup_{i=1}^N X_i)$. Since $\mathrm{co}(\cup_{i=1}^N M_i)$ is bounded, so is $\partial\mathscr{L}_{i\mu}$, i.e., for any $\mu \in \mathrm{co}(\cup_{i=1}^N M_i)$, there exists $L > 0$ such that $\|\partial\mathscr{L}_{i\mu}(x)\| \leq L$ for all $x \in \mathrm{co}(\cup_{i=1}^N X_i)$. Since g_ℓ is continuous (due to its convexity) and $\mathrm{co}(\cup_{i=1}^N X_i)$ is bounded, then g and thus $\partial\mathscr{L}_{ix}$ are bounded, i.e., for any $x \in \mathrm{co}(\cup_{i=1}^N X_i)$, there exists $R > 0$ such that $\|\partial\mathscr{L}_{ix}(\mu)\| \leq R$ for all $\mu \in \mathrm{co}(\cup_{i=1}^N M_i)$.

It follows from the Lagrangian subgradient inequality that

$$\mathscr{D}\mathscr{L}_{i\mu}(x)^T(x'-x) \leq \mathscr{L}_{i\mu}(x') - \mathscr{L}_{i\mu}(x), \quad \mathscr{D}\mathscr{L}_{i\mu}(x')^T(x-x') \leq \mathscr{L}_{i\mu}(x) - \mathscr{L}_{i\mu}(x'),$$

for any $x, x' \in \mathrm{co}(\cup_{i=1}^N X_i)$. By using the boundedness of the subdifferentials, the above two inequalities give that $-L\|x-x'\| \leq \mathscr{L}_{i\mu}(x) - \mathscr{L}_{i\mu}(x') \leq L\|x-x'\|$. This implies that $\|\mathscr{L}_{i\mu}(x) - \mathscr{L}_{i\mu}(x')\| \leq L\|x-x'\|$ for any $x, x' \in \mathrm{co}(\cup_{i=1}^m X_i)$. The proof of the Lipschitz continuity of $\mathscr{L}_{ix}$ is analogous by using the Lagrangian supgradient inequality. ■

The following lemma provides a basic iteration relation used in the convergence proof of the DISTRIBUTED LAGRANGIAN PRIMAL- DUAL SUBGRADIENT ALGORITHM.

Lemma 2.8 (Basic iteration relation) *Let the double stochasticity Assumption 1.2 and the periodic strong connectivity Assumption 1.3 hold. For any $x \in X$, any $\mu \in M$ and all $k \geq 0$, the following estimates hold:*

$$\sum_{i=1}^N \|e_x^i(k) + \alpha(k)\mathscr{D}_x^i(k)\|^2 \leq -\sum_{i=1}^N 2\alpha(k)(\mathscr{L}_i(v_x^i(k), v_\mu^i(k)) - \mathscr{L}_i(x, v_\mu^i(k)))$$
$$+ \sum_{i=1}^N \alpha(k)^2\|\mathscr{D}_x^i(k)\|^2 + \sum_{i=1}^N\{\|x^i(k)-x\|^2 - \|x^i(k+1)-x\|^2\}, \tag{2.14}$$
$$\sum_{i=1}^N \|e_\mu^i(k) - \alpha(k)\mathscr{D}_\mu^i(k)\|^2 \leq \sum_{i=1}^N 2\alpha(k)(\mathscr{L}_i(v_x^i(k), v_\mu^i(k)) - \mathscr{L}_i(v_x^i(k), \mu))$$
$$+ \sum_{i=1}^N \alpha(k)^2\|\mathscr{D}_\mu^i(k)\|^2 + \sum_{i=1}^N\{\|\mu^i(k)-\mu\|^2 - \|\mu^i(k+1)-\mu\|^2\}. \tag{2.15}$$

Proof By Lemma 1.1 with $Z = M_i$, $z = v_\mu^i(k) + \alpha(k)\mathscr{D}_\mu^i(k)$ and $y = \mu \in M$, we have that for all $k \geq 0$

$$
\begin{aligned}
&\sum_{i=1}^{N} \|e_\mu^i(k) - \alpha(k)\mathscr{D}_\mu^i(k)\|^2 \leq \sum_{i=1}^{N} \|v_\mu^i(k) + \alpha(k)\mathscr{D}_\mu^i(k) - \mu\|^2 - \sum_{i=1}^{N} \|\mu^i(k+1) - \mu\|^2 \\
&= \sum_{i=1}^{N} \|v_\mu^i(k) - \mu\|^2 + \sum_{i=1}^{N} \alpha(k)^2 \|\mathscr{D}_\mu^i(k)\|^2 \\
&+ \sum_{i=1}^{N} 2\alpha(k)\mathscr{D}_\mu^i(k)^T(v_\mu^i(k) - \mu) - \sum_{i=1}^{N} \|\mu^i(k+1) - \mu\|^2 \\
&\leq \sum_{i=1}^{N} \alpha(k)^2 \|\mathscr{D}_\mu^i(k)\|^2 + \sum_{i=1}^{N} 2\alpha(k)\mathscr{D}_\mu^i(k)^T(v_\mu^i(k) - \mu) \\
&+ \sum_{i=1}^{N} \|\mu^i(k) - \mu\|^2 - \sum_{i=1}^{N} \|\mu^i(k+1) - \mu\|^2.
\end{aligned} \tag{2.16}
$$

One can show (2.15) by substituting the following Lagrangian supgradient inequality into (2.16):

$$\mathscr{D}_\mu^i(k)^T(\mu - v_\mu^i(k)) \geq \mathscr{L}_i(v_x^i(k), \mu) - \mathscr{L}_i(v_x^i(k), v_\mu^i(k)).$$

Similarly, the equality (2.14) can be shown by using the following Lagrangian subgradient inequality: $\mathscr{D}_x^i(k)^T(x - v_x^i(k)) \leq \mathscr{L}_i(x, v_\mu^i(k)) - \mathscr{L}_i(v_x^i(k), v_\mu^i(k))$. ∎

The following lemma shows that the consensus is asymptotically reached.

Lemma 2.9 (Achieving consensus) *Let the nondegeneracy Assumption 1.1, the double stochasticity Assumption 1.2, and the periodic strong connectivity Assumption 1.3 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-size sequence* $\{\alpha(k)\}$ *satisfying* $\lim_{k\to+\infty} \alpha(k) = 0$. *Then there exist* $x^* \in X$ *and* $\mu^* \in M$ *such that*

$$
\begin{aligned}
&\lim_{k\to+\infty} \|x^i(k) - x^*\| = 0, \quad \lim_{k\to+\infty} \|\mu^i(k) - \mu^*\| = 0, \quad \forall i \in V, \\
&\lim_{k\to+\infty} \|y^i(k) - y^j(k)\| = 0, \quad \forall i, j \in V.
\end{aligned}
$$

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 2.7 that $\|\mathscr{D}_x^i(k)\| \leq L$. From Lemma 2.8 it follows that

$$\sum_{i=1}^{N} \|x^i(k+1) - x\|^2 \leq \sum_{i=1}^{N} \|x^i(k) - x\|^2 + \sum_{i=1}^{N} \alpha(k)^2 L^2$$
$$+ \sum_{i=1}^{N} 2\alpha(k)(\|\mathscr{L}_i(v_x^i(k), v_\mu^i(k))\| + \|\mathscr{L}_i(x, v_\mu^i(k))\|). \tag{2.17}$$

Notice that $v_x^i(k) \in \mathrm{co}(\cup_{i=1}^N X_i)$, $v_\mu^i(k) \in \mathrm{co}(\cup_{i=1}^N M_i)$ and $x \in X$ are bounded. Since $\mathscr{L}_i$ is continuous, then $\mathscr{L}_i(v_x^i(k), v_\mu^i(k))$ and $\mathscr{L}_i(x, v_\mu^i(k))$ are bounded. Since $\{\alpha(k)\}$ diminishes, one can verify that $\lim_{k\to+\infty} \sum_{i=1}^{N} \|x^i(k) - x\|^2$ exists for any $x \in X$. On the other hand, taking limits on both sides of (2.14), we obtain

$$\lim_{k\to+\infty} \sum_{i=1}^{N} \|e_x^i(k) + \alpha(k)\mathscr{D}_x^i(k)\|^2 = 0,$$

and therefore we deduce that $\lim_{k\to+\infty} \|e_x^i(k)\| = 0$ for all $i \in V$. It follows from Theorem 1.4 that $\lim_{k\to+\infty} \|x^i(k) - x^j(k)\| = 0$ 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 Theorem 1.4 that $\lim_{k\to+\infty} \|y^i(k) - y^j(k)\| = 0$ for all $i, j \in V$. ■

From Lemma 2.9, we know that the sequences of $\{x^i(k)\}$ and $\{\mu^i(k)\}$ of the DISTRIBUTED LAGRANGIAN PRIMAL-DUAL SUBGRADIENT ALGORITHM asymptotically agree on some point in X and some point in M, respectively. Denote by $\Theta \subseteq X \times M$ the set of such limit points. Denote by the average of primal and dual estimates $\hat{x}(k) \triangleq \frac{1}{N}\sum_{i=1}^N x^i(k)$ and $\hat{\mu}(k) \triangleq \frac{1}{N}\sum_{i=1}^N \mu^i(k)$, respectively. The following lemma further characterizes that the points in Θ are saddle points of the Lagrangian function $\mathscr{L}$ over $X \times M$.

Lemma 2.10 (Saddle point characterization of Θ) *Each point in Θ is a saddle point of the Lagrangian function $\mathscr{L}$ over $X \times M$.*

Proof Denote by $\Delta_x(k) \triangleq \max_{i,j\in V} \|x^j(k) - x^i(k)\|$ the maximum deviation of primal estimates. Notice that

$$\|v_x^i(k) - \hat{x}(k)\| = \|\sum_{j=1}^{N} a_j^i(k)x^j(k) - \sum_{j=1}^{N} \frac{1}{N}x^j(k)\|$$

$$= \| \sum_{j \neq i} a^i_j(k)(x^j(k) - x^i(k)) - \sum_{j \neq i} \frac{1}{N}(x^j(k) - x^i(k)) \|$$
$$\leq \sum_{j \neq i} a^i_j(k) \|x^j(k) - x^i(k)\| + \sum_{j \neq i} \frac{1}{N} \|x^j(k) - x^i(k)\| \leq 2\Delta_x(k).$$

Denote by the maximum deviation of dual estimates $\Delta_\mu(k) \triangleq \max_{i,j \in V} \|\mu^j(k) - \mu^i(k)\|$. 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 $\mathscr{L}$ over $X \times M$. Then at least one of the following equalities holds:

$$\exists x \in X \quad \text{s.t.} \quad \mathscr{L}(x^*, \mu^*) > \mathscr{L}(x, \mu^*), \tag{2.18}$$
$$\exists \mu \in M \quad \text{s.t.} \quad \mathscr{L}(x^*, \mu) > \mathscr{L}(x^*, \mu^*). \tag{2.19}$$

Suppose first that (2.18) holds. Then, there exists $\varsigma > 0$ such that $\mathscr{L}(x^*, \mu^*) = \mathscr{L}(x, \mu^*) + \varsigma$. Consider the sequences of $\{x^i(k)\}$ and $\{\mu^i(k)\}$ which converge respectively to x^* and μ^* defined above. The estimate (2.14) leads to

$$\sum_{i=1}^N \|x^i(k+1) - x\|^2 \leq \sum_{i=1}^N \|x^i(k) - x\|^2 + \alpha(k)^2 \sum_{i=1}^N \|\mathscr{D}^i_x(k)\|^2 - 2\alpha(k)$$
$$\times \sum_{i=1}^N (A_i(k) + B_i(k) + C_i(k) + D_i(k) + E_i(k) + F_i(k)),$$

where the notations are given by:

$$A_i(k) = \mathscr{L}_i(v^i_x(k), v^i_\mu(k)) - \mathscr{L}_i(\hat{x}(k), v^i_\mu(k)),$$
$$B_i(k) = \mathscr{L}_i(\hat{x}(k), v^i_\mu(k)) - \mathscr{L}_i(\hat{x}(k), \hat{\mu}(k)),$$
$$C_i(k) = \mathscr{L}_i(\hat{x}(k), \hat{\mu}(k)) - \mathscr{L}_i(x^*, \hat{\mu}(k)), \quad D_i(k) = \mathscr{L}_i(x^*, \hat{\mu}(k)) - \mathscr{L}_i(x^*, \mu^*),$$
$$E_i(k) = \mathscr{L}_i(x^*, \mu^*) - \mathscr{L}_i(x, \mu^*), \quad F_i(k) = \mathscr{L}_i(x, \mu^*) - \mathscr{L}_i(x, v^i_\mu(k)).$$

It follows from the Lipschitz continuity property of $\mathscr{L}_i$; see Lemma 2.7, that

$$\|A_i(k)\| \leq L\|v^i_x(k) - \hat{x}(k)\| \leq 2L\Delta_x(k), \quad \|B_i(k)\| \leq R\|v^i_\mu(k) - \hat{\mu}(k)\| \leq 2R\Delta_\mu(k),$$
$$\|C_i(k)\| \leq L\|\hat{x}(k) - x^*\| \leq \frac{L}{N} \sum_{i=1}^N \|x^i(k) - x^*\|,$$
$$\|D_i(k)\| \leq R\|\hat{\mu}(k) - \mu^*\| \leq \frac{R}{N} \sum_{i=1}^N \|\mu^i(k) - \mu^*\|,$$
$$\|F_i(k)\| \leq R\|\mu^* - v^i_\mu(k)\| \leq R\|\mu^* - \hat{\mu}(k)\| + R\|\hat{\mu}(k) - v^i_\mu(k)\|$$

$$\leq \frac{R}{N}\sum_{i=1}^{N}\|\mu^*(k)-\mu^i(k)\| + 2R\Delta_\mu(k).$$

Since $\lim_{k\to+\infty}\|x^i(k)-x^*\|=0$, $\lim_{k\to+\infty}\|\mu^i(k)-\mu^*\|=0$, $\lim_{k\to+\infty}\Delta_x(k)=0$ 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

$$\sum_{i=1}^{N}\|x^i(k+1)-x\|^2 \leq \sum_{i=1}^{N}\|x^i(k)-x\|^2 + N\alpha(k)^2L^2 - \varsigma\alpha(k).$$

Following a recursive argument, we have that for all $k\geq k_0$, it holds that

$$\sum_{i=1}^{N}\|x^i(k+1)-x\|^2 \leq \sum_{i=1}^{N}\|x^i(k_0)-x\|^2 + NL^2\sum_{\tau=k_0}^{k}\alpha(\tau)^2 - \varsigma\sum_{\tau=k_0}^{k}\alpha(\tau).$$

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, (2.18) cannot hold. Following a parallel argument, one can show that (2.19) cannot hold either. This ensures that each $(x^*,\mu^*)\in\Theta$ is a saddle point of $\mathscr{L}$ over $X\times M$. ■

The combination of (c) in Lemmas 2.1 and 2.10 gives that, for each $(x^*,\mu^*)\in\Theta$, we have that $\mathscr{L}(x^*,\mu^*)=p^*$ and μ^* 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 2.1 based on two claims.

Proofs of Theorem 2.1:

Claim 2.1 *Each point $(x^*,\mu^*)\in\Theta$ is a point in $X^*\times D_L^*$.*

Proof The Lagrangian dual optimality of μ^* follows from (c) in Lemmas 2.1 and 2.10. To characterize the primal optimality of x^*, we define an auxiliary sequence $\{z(k)\}$ by $z(k)\triangleq\frac{\sum_{\tau=0}^{k-1}\alpha(\tau)\hat{x}(\tau)}{\sum_{\tau=0}^{k-1}\alpha(\tau)}$ which is a weighted version of the average of primal estimates. Since $\lim_{k\to+\infty}\hat{x}(k)=x^*$, it follows from Lemma 2.6 (b) that $\lim_{k\to+\infty}z(k)=x^*$.

Since (x^*,μ^*) is a saddle point of $\mathscr{L}$ over $X\times M$, then $\mathscr{L}(x^*,\mu)\leq\mathscr{L}(x^*,\mu^*)$ for any $\mu\in M$; i.e., the following relation holds for any $\mu\in M$:

$$g(x^*)^T(\mu-\mu^*)\leq 0. \tag{2.20}$$

Choose $\mu_a = \mu^* + \min_{i\in V}\theta_i \frac{\mu^*}{\|\mu^*\|}$ 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 (2.20) gives that

$$\frac{\min_{i\in V}\theta_i}{\|\mu^*\|} g(x^*)^T\mu^* \leq 0.$$

Since $\theta_i > 0$, we have $g(x^*)^T\mu^* \leq 0$. On the other hand, we choose $\mu_b = \frac{1}{2}\mu^*$ and then $\mu_b \in M$. Letting $\mu = \mu_b$ in (2.20) 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 $J^+(x^*) \triangleq \{1 \leq \ell \leq m \mid g_\ell(x^*) > 0\} \neq \emptyset$ and $\eta \triangleq \min_{\ell\in J^+(x^*)}\{g_\ell(x^*)\}$. Then $\eta > 0$. Since g is continuous and $v_x^i(k)$ converges to x^*, there exists $K \geq 0$ such that $g_\ell(v_x^i(k)) \geq \frac{\eta}{2}$ for all $k \geq K$ and all $\ell \in J^+(x^*)$. Since $v_\mu^i(k)$ converges to μ^*, without loss of generality, we say that $\|v_\mu^i(k) - \mu^*\| \leq \frac{1}{2}\min_{i\in V}\theta_i$ for all $k \geq K$. Choose $\hat{\mu}$ such that $\hat{\mu}_\ell = \mu_\ell^*$ for $\ell \notin J^+(x^*)$ and $\hat{\mu}_\ell = \mu_\ell^* + \frac{1}{\sqrt{m}}\min_{i\in V}\theta_i$ for $\ell \in J^+(x^*)$. Since $\mu^* \geq 0$ and $\theta_i > 0$, $\hat{\mu} \geq 0$. Furthermore, $\|\hat{\mu}\| \leq \|\mu^*\| + \min_{i\in V}\theta_i$, then $\hat{\mu} \in M$. Equating μ to $\hat{\mu}$ and letting $\mathscr{D}_\mu^i(k) = g(v_x^i(k))$ in the estimate (2.16), the following holds for $k \geq K$:

$$\begin{aligned}
N|J^+(x^*)|\eta \min_{i\in V}\theta_i\alpha(k) &\leq 2\alpha(k)\sum_{i=1}^{N}\sum_{\ell\in J^+(x^*)} g_\ell(v_x^i(k))(\hat{\mu} - v_\mu^i(k))_\ell \\
&\leq \sum_{i=1}^{N}\|\mu^i(k) - \hat{\mu}\|^2 - \sum_{i=1}^{N}\|\mu^i(k+1) - \hat{\mu}\|^2 + NR^2\alpha(k)^2 \\
&- 2\alpha(k)\sum_{i=1}^{N}\sum_{\ell\notin J^+(x^*)} g_\ell(v_x^i(k))(\hat{\mu} - v_\mu^i(k))_\ell. \qquad (2.21)
\end{aligned}$$

Summing (2.21) 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

$$\begin{aligned}
N|J^+(x^*)|\eta \min_{i\in V}\theta_i \leq \frac{1}{\sum_{\tau=K}^{k-1}\alpha(\tau)}\Bigg\{&\sum_{i=1}^{N}\|\mu^i(K) - \hat{\mu}\|^2 + NR^2\sum_{\tau=K}^{k-1}\alpha(\tau)^2 \\
-\sum_{\tau=K}^{k-1}2\alpha(\tau)\sum_{i=1}^{N}&\sum_{\ell\notin J^+(x^*)} g_\ell(v_x^i(\tau))(\hat{\mu} - v_\mu^i(\tau))_\ell\Bigg\}. \qquad (2.22)
\end{aligned}$$

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 (2.22) 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 $\lim_{k\to+\infty} v_x^i(k) = x^*$ and $\lim_{k\to+\infty} v_\mu^i(k) = \mu^*$, the following holds:

$$\lim_{k\to+\infty} 2\sum_{i=1}^{N} \sum_{\ell \notin J^+(x^*)} g_\ell(v_x^i(k))(\hat{\mu} - v_\mu^i(k))_\ell = 0.$$

Then it follows from Lemma 2.6 (b) that the limit of the third term is zero as $k \to +\infty$. We have $N|J^+(x^*)|\eta \min_{i\in V} \theta_i \le 0$. Recall that $|J^+(x^*)| > 0$, $\eta > 0$ and $\theta_i > 0$. Then we reach a contradiction, implying that $g(x^*) \le 0$.

Since $x^* \in X$ and $g(x^*) \le 0$, then x^* is a feasible solution and thus $f(x^*) \ge p^*$. On the other hand, since $z(k)$ is a convex combination of $\hat{x}(0), \ldots, \hat{x}(k-1)$ and f is convex, we have the following estimate:

$$\begin{aligned} f(z(k)) &\le \frac{\sum_{\tau=0}^{k-1} \alpha(\tau) f(\hat{x}(\tau))}{\sum_{\tau=0}^{k-1} \alpha(\tau)} \\ &= \frac{1}{\sum_{\tau=0}^{k-1} \alpha(\tau)} \left\{ \sum_{\tau=0}^{k-1} \alpha(\tau)\mathscr{L}(\hat{x}(\tau), \hat{\mu}(\tau)) - \sum_{\tau=0}^{k-1} N\alpha(\tau)\hat{\mu}(\tau)^T g(\hat{x}(\tau)) \right\}. \end{aligned}$$

Recall the following convergence properties:

$$\begin{aligned} &\lim_{k\to+\infty} z(k) = x^*, \quad \lim_{k\to+\infty} \mathscr{L}(\hat{x}(k), \hat{\mu}(k)) = \mathscr{L}(x^*, \mu^*) = p^*, \\ &\lim_{k\to+\infty} \hat{\mu}(k)^T g(\hat{x}(k)) = g(x^*)^T \mu^* = 0. \end{aligned}$$

It follows from Lemma 2.6 (b) that $f(x^*) \le p^*$. Therefore, we have $f(x^*) = p^*$, and thus x^* is a primal optimal point. ■

Claim 2.2 *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 $k' \ge 1$:

$$\sum_{i=1}^{N} y^i(k+1) = \sum_{i=1}^{N} y^i(k') + N\sum_{\ell=k'}^{k}\sum_{i=1}^{N}(f_i(x^i(\ell)) - f_i(x^i(\ell-1))). \tag{2.23}$$

Let $k' = 1$ in (2.23) and recall that initial state $y^i(1) = Nf_i(x^i(0))$ for all $i \in V$. Then we have

$$\sum_{i=1}^{N} y^i(k+1) = \sum_{i=1}^{N} y^i(1) + N\sum_{i=1}^{N}(f_i(x^i(k)) - f_i(x^i(0))) = N\sum_{i=1}^{N} f_i(x^i(k)). \tag{2.24}$$

The combination of (2.24) with $\lim_{k\to+\infty} \|y^i(k) - y^j(k)\| = 0$ gives the desired result. ∎

2.5.2 *Convergence Analysis of the* DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM

In order to analyze the DISTRIBUTED PENALTY PRIMAL- DUAL SUBGRADIENT ALGORITHM, we first rewrite it into the following form:

$$\mu^i(k+1) = v^i_\mu(k) + u^i_\mu(k), \quad \lambda^i(k+1) = v^i_\lambda(k) + u^i_\lambda(k),$$
$$x^i(k+1) = v^i_x(k) + e^i_x(k), \quad y^i(k+1) = v^i_y(k) + u^i_y(k),$$

where $e^i_x(k)$ is projection error described by

$$e^i_x(k) \triangleq P_X[v^i_x(k) - \alpha(k)\mathscr{S}^i_x(k)] - v^i_x(k),$$

and the quantities $u^i_\mu(k) \triangleq \alpha(k)[g(v^i_x(k))]^+$, $u^i_\lambda(k) \triangleq \alpha(k)|h(v^i_x(k))|$ and $u^i_y(k) = N(f_i(x^i(k)) - f_i(x^i(k-1)))$ represent local inputs. Denote by the maximum deviations of dual estimates $M_\mu(k) \triangleq \max_{i\in V} \|\mu^i(k)\|$ and $M_\lambda(k) \triangleq \max_{i\in V} \|\lambda^i(k)\|$. Before showing Lemma 2.11, we present some useful facts. 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, $[g_\ell(\cdot)]^+$, $|h_\ell(\cdot)|$ are convex, then it follows from Proposition 5.4.2 in [11] that there exist $D_F, D_{G^+}, D_H > 0$ such that for all $x \in X$, it holds that $\|\mathscr{D} f_i(x)\| \le D_F$ $(i \in V)$, $m\|\mathscr{D}[g_\ell(x)]^+\| \le D_{G^+}$ $(1 \le \ell \le m)$ and $\nu\|\mathscr{D}|h_\ell|(x)\| \le D_H$ $(1 \le \ell \le \nu)$. Denote by the averages of primal and dual estimates $\hat{x}(k) \triangleq \frac{1}{N}\sum_{i=1}^N x^i(k)$, $\hat{\mu}(k) \triangleq \frac{1}{N}\sum_{i=1}^N \mu^i(k)$ and $\hat{\lambda}(k) \triangleq \frac{1}{N}\sum_{i=1}^N \lambda^i(k)$.

Lemma 2.11 (Diminishing and summable properties) *Suppose the double stochasticity Assumption 1.2 and the step-size Assumption 2.2 hold.*

(a) The following holds:

$$\lim_{k\to+\infty} \alpha(k)M_\mu(k) = 0, \quad \lim_{k\to+\infty} \alpha(k)M_\lambda(k) = 0, \quad \lim_{k\to+\infty} \alpha(k)\|\mathscr{S}^i_x(k)\| = 0.$$

Furthermore, the sequences of $\{\alpha(k)^2 M^2_\mu(k)\}$, $\{\alpha(k)^2 M^2_\lambda(k)\}$, *and* $\{\alpha(k)^2\|\mathscr{S}^i_x(k)\|^2\}$ *are summable.*

(b) The following sequences are summable:

$$\{\alpha(k)\|\hat{\mu}(k) - v_\mu^i(k)\|\}, \{\alpha(k)\|\hat{\lambda}(k) - v_\lambda^i(k)\|\}, \{\alpha(k)M_\mu(k)\|\hat{x}(k) - v_x^i(k)\|\},$$
$$\{\alpha(k)M_\lambda(k)\|\hat{x}(k) - v_x^i(k)\|\}, \{\alpha(k)\|\hat{x}(k) - v_x^i(k)\|\}.$$

Proof (a) Notice that

$$\|v_\mu^i(k)\| = \|\sum_{j=1}^N a_j^i(k)\mu^j(k)\| \le \sum_{j=1}^N a_j^i(k)\|\mu^j(k)\| \le \sum_{j=1}^N a_j^i(k)M_\mu(k) = M_\mu(k),$$

where in the last equality we use the double stochasticity Assumption 1.2. Recall that $v_x^i(k) \in X$. This implies that the following holds for all $k \ge 0$:

$$\|\mu^i(k+1)\| \le \|v_\mu^i(k) + \alpha(k)[g(v_x^i(k))]^+\| \le \|v_\mu^i(k)\| + G^+\alpha(k) \le M_\mu(k) + G^+\alpha(k).$$

From here, then we deduce the following recursive estimate on $M_\mu(k+1)$: $M_\mu(k+1) \le M_\mu(k) + G^+\alpha(k)$. Repeatedly applying the above estimates yields that

$$M_\mu(k+1) \le M_\mu(0) + G^+s(k). \tag{2.25}$$

Similar arguments can be employed to show that

$$M_\lambda(k+1) \le M_\lambda(0) + Hs(k). \tag{2.26}$$

Since $\lim_{k\to+\infty} \alpha(k+1)s(k) = 0$ and $\lim_{k\to+\infty} \alpha(k) = 0$, we know that

$$\lim_{k\to+\infty} \alpha(k+1)M_\mu(k+1) = 0, \quad \lim_{k\to+\infty} \alpha(k+1)M_\lambda(k+1) = 0.$$

Notice that the following estimate on $\mathscr{S}_x^i(k)$ holds:

$$\|\mathscr{S}_x^i(k)\| \le D_F + D_{G^+}M_\mu(k) + D_H M_\lambda(k). \tag{2.27}$$

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 $\lim_{k\to+\infty} \alpha(k)\|\mathscr{S}_x^i(k)\| = 0$ follows. By (2.25), we have

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

It follows from the step-size Assumption 2.2 that $\sum_{k=0}^{+\infty} \alpha(k)^2 M_\mu^2(k) < +\infty$. Similarly, one can show that $\sum_{k=0}^{+\infty} \alpha(k)^2 M_\lambda^2(k) < +\infty$. By using (2.25)–(2.27), we have the following estimate:

$$\sum_{k=0}^{+\infty}\alpha(k)^2\|\mathscr{S}_x^i(k)\|^2 \le \alpha(0)^2(D_F + D_{G^+}M_\mu(0) + D_H M_\lambda(0))^2$$
$$+\sum_{k=1}^{+\infty}\alpha(k)^2(D_F + D_{G^+}(M_\mu(0) + G^+ s(k-1)) + D_H(M_\lambda(0) + Hs(k-1)))^2.$$

Then the summability of $\{\alpha(k)^2\}$, $\{\alpha(k+1)^2 s(k)\}$ and $\{\alpha(k+1)^2 s(k)^2\}$ verifies that of $\{\alpha(k)^2\|\mathscr{S}_x^i(k)\|^2\}$.

(b) Consider the dynamics of $\mu^i(k)$ which is in the same form as the distributed projected subgradient algorithm in [13]. Recall that $\{[g(v_x^i(k))]^+\}$ is uniformly bounded. Then following from Lemma 2.5 in the Appendix 2.5 with $Z = \mathbb{R}^m_{\ge 0}$ 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:

$$\|\hat{\mu}(k) - v_\mu^i(k)\| \le \sum_{j=1}^N a_j^i(k)\|\hat{\mu}(k) - \mu^j(k)\| \le \max_{i\in V}\|\hat{\mu}(k) - \mu^i(k)\|, \tag{2.28}$$

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 1.1 with $Z = X$, $z = v_x^i(k) - \alpha(k)\mathscr{S}_x^i(k)$ and $y = v_x^i(k)$, we have

$$\|x^i(k+1) - v_x^i(k)\|^2 \le \|v_x^i(k) - \alpha(k)\mathscr{S}_x^i(k) - v_x^i(k)\|^2$$
$$- \|x^i(k+1) - (v_x^i(k) - \alpha(k)\mathscr{S}_x^i(k))\|^2,$$

and thus $\|e_x^i(k) + \alpha(k)\mathscr{S}_x^i(k)\| \le \alpha(k)\|\mathscr{S}_x^i(k)\|$. With this relation, from Lemma 2.5 with $Z = X$ and $d_i(k) = \mathscr{S}_x^i(k)$, the following holds for some $\gamma > 0$ and $0 < \beta < 1$:

$$\|x^i(k) - \hat{x}(k)\| \le N\gamma\beta^{k-1}\sum_{i=0}^N\|x^i(0)\| + 2N\gamma\sum_{\tau=0}^{k-1}\beta^{k-\tau}\alpha(\tau)\|\mathscr{S}_x^i(\tau)\|. \tag{2.29}$$

Multiplying both sides of (2.29) by $\alpha(k)M_\mu(k)$ and using (2.27), we obtain

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

Notice that the above inequalities hold for all $i \in V$. Then by employing the relation of $ab \le \frac{1}{2}(a^2 + b^2)$ and regrouping similar terms, we obtain

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

Part (a) gives that $\{\alpha(k)^2M_\mu^2(k)\}$ is summable. Combining this fact with the property of $\sum_{\tau=0}^{k-1}\beta^{k-\tau} \le \sum_{k=0}^{+\infty}\beta^k = \frac{1}{1-\beta}$, 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

$$\lim_{k\to+\infty}\alpha(k)^2(D_F + D_{G^+}M_\mu^2(k) + D_H M_\lambda^2(k)) = 0,$$

and thus $\{\alpha(k)^2(D_F+D_{G^+}M_\mu^2(k)+D_H M_\lambda^2(k))\}$ is summable. Then Lemma 7 in [13] with $\gamma_\ell = N\gamma\alpha(\ell)^2(D_F + D_{G^+}M_\mu^2(\ell) + D_H M_\lambda^2(\ell))$ 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 (2.28), one can show the summability of $\{\alpha(k)M_\mu(k)\|v_x^i(k) - \hat{x}(k)\|\}$. Following analogous arguments, we have that $\{\alpha(k)M_\lambda(k)\|v_x^i(k) - \hat{x}(k)\|\}$ and $\{\alpha(k)\|v_x^i(k) - \hat{x}(k)\|\}$ are summable. ■

Remark 2.5 In Lemma 2.11, the assumption of all local constraint sets being identical is utilized to find an upper bound of the convergence rate of $\|\hat{x}(k) - v_x^i(k)\|$ to zero. This property is crucial to establish the summability of expansions pertaining to $\|\hat{x}(k) - v_x^i(k)\|$ in part (b). •

The following is a basic iteration relation of the DISTRIBUTED PENALTY PRIMAL-DUAL SUBGRADIENT ALGORITHM.

Lemma 2.12 (Basic iteration relation) *The following estimates hold for any $x \in X$ and $(\mu, \lambda) \in \mathbb{R}^m_{\ge 0} \times \mathbb{R}^\nu_{\ge 0}$:*

$$\sum_{i=1}^{N}\|e_x^i(k) + \alpha(k)\mathscr{S}_x^i(k)\|^2 \le \sum_{i=1}^{N}\alpha(k)^2\|\mathscr{S}_x^i(k)\|^2$$
$$-\sum_{i=1}^{N}2\alpha(k)(\mathscr{H}_i(v_x^i(k), v_\mu^i(k), v_\lambda^i(k)) - \mathscr{H}_i(x, v_\mu^i(k), v_\lambda^i(k)))$$
$$+\sum_{i=1}^{N}(\|x^i(k) - x\|^2 - \|x^i(k+1) - x\|^2), \tag{2.30}$$

$$
\begin{aligned}
0 \le & \sum_{i=1}^{N} (\|\mu^i(k) - \mu\|^2 - \|\mu^i(k+1) - \mu\|^2) \\
& + \sum_{i=1}^{N} (\|\lambda^i(k) - \lambda\|^2 - \|\lambda^i(k+1) - \lambda\|^2) + \\
& \sum_{i=1}^{N} 2\alpha(k)(\mathscr{H}_i(v_x^i(k), v_\mu^i(k), v_\lambda^i(k)) - \mathscr{H}_i(v_x^i(k), \mu, \lambda)) \\
& + \sum_{i=1}^{N} \alpha(k)^2(\|[g(v_x^i(k))]^+\|^2 + \|h(v_x^i(k))\|^2).
\end{aligned} \tag{2.31}
$$

Proof One can finish the proof by following analogous arguments in Lemma 2.8. ∎

Lemma 2.13 (Achieving consensus) *Let us suppose that the nondegeneracy Assumption 1.1, the double stochasticity Assumption 1.2, and the periodical strong connectivity Assumption 1.3 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 (2.16), we have

$$
\begin{aligned}
& \sum_{i=1}^{N} \|x^i(k+1) - x\|^2 \le \sum_{i=1}^{N} \|x^i(k) - x\|^2 \\
& + \sum_{i=1}^{N} \alpha(k)^2 \|\mathscr{S}_x^i(k)\|^2 + \sum_{i=1}^{N} 2\alpha(k)\|\mathscr{S}_x^i(k)\|\|v_x^i(k) - x\|.
\end{aligned}
$$

Since $\lim_{k\to+\infty} \alpha(k)\|\mathscr{S}_x^i(k)\| = 0$, the proofs of $\lim_{k\to+\infty} \|x^i(k) - \tilde{x}\| = 0$ for all $i \in V$ are analogous to those in Lemma 2.9. The remainder of the proofs can be finished by Theorem 1.4 with the properties of $\lim_{k\to+\infty} u_\mu^i(k) = 0$, $\lim_{k\to+\infty} u_\lambda^i(k) = 0$ and $\lim_{k\to+\infty} u_y^i(k) = 0$ (due to $\lim_{k\to+\infty} x^i(k) = \tilde{x}$ and f_i is continuous). ∎

We now proceed to show Theorem 2.3 based on five claims.

Proof of Theorem 2.3:

Claim 2.3 *For any* $x^* \in X^*$ *and* $(\mu^*, \lambda^*) \in D_P^*$*, the following sequences are summable:*

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

Proof Observe that

$$\begin{aligned}
&\|\mathscr{H}_i(x^*, v_\mu^i(k), v_\lambda^i(k)) - \mathscr{H}_i(x^*, \hat{\mu}(k), \hat{\lambda}(k))\|\\
&\le \|v_\mu^i(k) - \hat{\mu}(k)\|\|[g(x^*)]^+\| + \|v_\lambda^i(k) - \hat{\lambda}(k)\|\|h(x^*)\|\\
&\le G^+\|v_\mu^i(k) - \hat{\mu}(k)\| + H\|v_\lambda^i(k) - \hat{\lambda}(k)\|.
\end{aligned} \tag{2.32}$$

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 2.11, we have that the following are summable:

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

Similarly, the following estimates hold:

$$\begin{aligned}
&\|\mathscr{H}_i(v_x^i(k), \mu^*, \lambda^*) - \mathscr{H}_i(\hat{x}(k), \mu^*, \lambda^*)\|\\
&\le \|f_i(v_x^i(k)) - f_i(\hat{x}(k))\| + \|(\mu^*)^T([g(v_x^i(k))]^+ - [g(\hat{x}(k))]^+)\|\\
&+ \|(\lambda^*)^T(|h(v_x^i(k))| - |h(\hat{x}(k))|)\|\\
&\le (D_F + D_{G^+}\|\mu^*\| + D_H\|\lambda^*\|)\|v_x^i(k) - \hat{x}(k)\|.
\end{aligned}$$

Then the property of $\sum_{k=0}^{+\infty}\alpha(k)\|\hat{x}(k) - v_x^i(k)\| < +\infty$ in Part (b) of Lemma 2.11 implies the summability of the following sequences:

$$\left\{\alpha(k)\sum_{i=1}^{N}\|\mathscr{H}_i(v_x^i(k), \mu^*, \lambda^*) - \mathscr{H}_i(\hat{x}(k), \mu^*, \lambda^*)\|\right\},$$
$$\left\{\alpha(k)\sum_{i=1}^{N}\left(\mathscr{H}_i(v_x^i(k), \mu^*, \lambda^*) - \mathscr{H}_i(\hat{x}(k), \mu^*, \lambda^*)\right)\right\}.$$

■

Claim 2.4 *Denote the weighted version of $\mathscr{H}_i$ as*

$$\hat{\mathscr{H}}_i(k) \triangleq \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) \mathscr{H}_i(v_x^i(\ell), v_\mu^i(\ell), v_\lambda^i(\ell)).$$

The following property holds: $\lim_{k \to +\infty} \sum_{i=1}^{N} \hat{\mathscr{H}}_i(k) = p^*$.

Proof Summing (2.30) over $[0, k-1]$ and replacing x by $x^* \in X^*$ leads to

$$\sum_{\ell=0}^{k-1} \alpha(\ell) \sum_{i=1}^{N} (\mathscr{H}_i(v_x^i(\ell), v_\mu^i(\ell), v_\lambda^i(\ell)) - \mathscr{H}_i(x^*, v_\mu^i(\ell), v_\lambda^i(\ell)))$$
$$\leq \sum_{i=1}^{N} \|x^i(0) - x^*\|^2 + \sum_{\ell=0}^{k-1} \sum_{i=1}^{N} \alpha(\ell)^2 \|\mathscr{S}_x^i(\ell)\|^2. \tag{2.33}$$

The summability of $\{\alpha(k)^2 \|\mathscr{S}_x^i(k)\|^2\}$ in Part (b) of Lemma 2.11 implies that the right-hand side of (2.33) is finite as $k \to +\infty$, and thus

$$\limsup_{k \to \infty} \frac{1}{s(k-1)} \sum_{\ell=0}^{k-1} \alpha(\ell) \left[\sum_{i=1}^{N} \left(\mathscr{H}_i(v_x^i(\ell), v_\mu^i(\ell), v_\lambda^i(\ell)) - \mathscr{H}_i(x^*, v_\mu^i(\ell), v_\lambda^i(\ell)) \right) \right] \leq 0. \tag{2.34}$$

Pick any $(\mu^*, \lambda^*) \in D_P^*$. It follows from Theorem 2.2 that (x^*, μ^*, λ^*) is a saddle point of $\mathscr{H}$ over $X \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$. Since $(\hat{\mu}(k), \hat{\lambda}(k)) \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^\nu_{\geq 0}$, then we have $\mathscr{H}(x^*, \hat{\mu}(k), \hat{\lambda}(k)) \leq \mathscr{H}(x^*, \mu^*, \lambda^*) = p^*$. Combining this relation, Claim 2.3 and (2.34) renders that

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

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

On the other hand, $\hat{x}(k) \in X$ (due to the fact that X is convex) implies that $\mathscr{H}(\hat{x}(k), \mu^*, \lambda^*) \geq \mathscr{H}(x^*, \mu^*, \lambda^*) = p^*$. Along similar lines, by using (2.31) with $\mu = \mu^*$, $\lambda = \lambda^*$, and Claim 2.3, we have $\liminf_{k \to +\infty} \sum_{i=1}^{N} \hat{\mathscr{H}}_i(k) \geq p^*$. Then we have the desired relation. ∎

Claim 2.5 *Denote by* $\pi(k) \triangleq \sum_{i=1}^{N} \mathscr{H}_i(v_x^i(k), v_\mu^i(k), v_\lambda^i(k)) - \mathscr{H}(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k))$. *And we denote the weighted version of* $\mathscr{H}$ *as*

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

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

Proof Notice that

$$\begin{aligned}
\pi(k) = &\sum_{i=1}^{N} (f_i(v_x^i(k)) - f_i(\hat{x}(k))) + \sum_{i=1}^{N} \left(v_\mu^i(k)^T [g(v_x^i(k))]^+ - v_\mu^i(k)^T [g(\hat{x}(k))]^+\right) \\
&+ \sum_{i=1}^{N} \left(v_\mu^i(k)^T [g(\hat{x}(k))]^+ - \hat{\mu}(k)^T [g(\hat{x}(k))]^+\right) \\
&+ \sum_{i=1}^{N} \left(v_\lambda^i(k)^T |h(v_x^i(k))| - v_\lambda^i(k)^T |h(\hat{x}(k))|\right) \\
&+ \sum_{i=1}^{N} \left(v_\lambda^i(k)^T |h(\hat{x}(k))| - \hat{\lambda}(k)^T |h(\hat{x}(k))|\right).
\end{aligned} \tag{2.35}$$

By using the boundedness of subdifferentials and the primal estimates, it follows from (2.35) that

$$\begin{aligned}
\|\pi(k)\| \le &(D_F + D_{G^+} M_\mu(k) + D_H M_\lambda(k)) \times \sum_{i=1}^{N} \|v_x^i(k) - \hat{x}(k)\| \\
&+ G^+ \sum_{i=1}^{N} \|v_\mu^i(k) - \hat{\mu}(k)\| + H \sum_{i=1}^{N} \|v_\lambda^i(k) - \hat{\lambda}(k)\|.
\end{aligned} \tag{2.36}$$

Then it follows from (b) in Lemma 2.11 that $\{\alpha(k)\|\pi(k)\|\}$ is summable. Notice that $\|\Gamma(k) - \sum_{i=1}^{N} \hat{\mathscr{H}}_i(k)\| \le \frac{\sum_{\ell=0}^{k-1} \alpha(\ell)\|\pi(\ell)\|}{s(k-1)}$, and thus $\lim_{k \to +\infty} \|\Gamma(k) - \sum_{i=1}^{N} \hat{\mathscr{H}}_i(k)\| = 0$. The desired result immediately follows from Claim 2.4. ∎

Claim 2.6 *The limit point* $\tilde{x}$ *in Lemma 2.13 is a primal optimal solution.*

Proof Let $\hat{\mu}(k) = (\hat{\mu}_1(k), \dots, \hat{\mu}_m(k))^T \in \mathbb{R}_{\ge 0}^m$. By the double stochasticity Assumption 1.2, we obtain

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

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

This implies that the sequence $\{\hat{\mu}_\ell(k)\}$ is nondecreasing in $\mathbb{R}_{\geq 0}$. Observe that $\{\hat{\mu}_\ell(k)\}$ is lower bounded by zero. In this way, we distinguish the following two cases:

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 $\lim_{k \to +\infty} \|\mu_\ell^i(k) - \mu_\ell^*\| = 0$ 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 the property of $\sum_{k=0}^{+\infty} \alpha(k) \sum_{i=1}^N [g_\ell(v_x^i(k))]^+ < +\infty$, further implying the property of $\liminf_{k \to +\infty} [g_\ell(v_x^i(k))]^+ = 0$. Since $\lim_{k \to +\infty} \|x^i(k) - \tilde{x}\| = 0$ for all $i \in V$, then it holds that $\lim_{k \to +\infty} \|v_x^i(k) - \tilde{x}\| = 0$, implying $[g_\ell(\tilde{x})]^+ = 0$.

Case 2: The sequence $\{\hat{\mu}_\ell(k)\}$ is not upper bounded. Since $\{\hat{\mu}_\ell(k)\}$ is nondecreasing, then $\hat{\mu}_\ell(k) \to +\infty$. It follows from Claim 2.5 and (a) in Lemma 2.6 that it is impossible that $\mathscr{H}(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) \to +\infty$. Assume that $[g_\ell(\tilde{x})]^+ > 0$. Then we have

$$\begin{aligned} \mathscr{H}(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) &= f(\hat{x}(k)) + N\hat{\mu}(k)^T [g(\hat{x}(k))]^+ + N\lambda(k)^T |h(\hat{x}(k))| \\ &\geq f(\hat{x}(k)) + \hat{\mu}_\ell(k) [g_\ell(\hat{x}(k))]^+. \end{aligned} \tag{2.37}$$

Taking limits on both sides of (2.37) and we obtain:

$$\liminf_{k \to +\infty} \mathscr{H}(\hat{x}(k), \hat{\mu}(k), \hat{\lambda}(k)) \geq \limsup_{k \to +\infty} (f(\hat{x}(k)) + \hat{\mu}_\ell(k) [g_\ell(\hat{x}(k))]^+) = +\infty.$$

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

In both cases, we have $[g_\ell(\tilde{x})]^+ = 0$ for any $1 \leq \ell \leq m$. By utilizing similar arguments, we can further prove that $|h(\tilde{x})| = 0$. Since $\tilde{x} \in X$, then $\tilde{x}$ is feasible and thus $f(\tilde{x}) \geq p^*$. 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 $\lim_{k \to +\infty} \hat{x}(k) = \tilde{x}$, then Claim 2.5 and (b) in Lemma 2.6 implies that

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

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

Claim 2.7 *It holds that* $\lim_{k \to +\infty} \|y^i(k) - p^*\| = 0$.

Proof The proof follows the same lines in Claim 2.2 of Theorem 2.1 and thus is omitted here. ■

2.6 Notes

Distributed optimization traces back to 1970s. In [14], the classic dual decomposition approach is proposed to the class of distributed optimization problems characterized by separable component functions. This approach has been successfully applied to handle network utility maximization (NUM) in; e.g., [4, 15, 16]. In [17, 18], the authors develop a general framework for parallel and distributed computation over a set of processors.

Recently, diffusion consensus algorithms have been integrated into distributed optimization to address the nonseparability in component functions and dynamic changes of network topologies. In particular, distributed projected subgradient algorithms are proposed in [13] to address non-smooth multi-agent optimization with constraint sets. The paper [19] comes up with Distributed Lagrangian Primal-Dual Subgradient Algorithm and Distributed Penalty Primal-Dual Subgradient Algorithm to further address inequality and equality constraints. The results developed in [19] are extended to solve a class of distributed nonconvex optimization problems in [20]. All the algorithms aforementioned are discrete-time. The continuous-time counterparts are investigated in [21–23]. In [24], a distributed continuous-time algorithm with discrete-time communication is proposed. Random network and state-dependent topologies are investigated in [25, 26] respectively.

There have been a number of other interesting algorithms for distributed optimization. The authors in [27, 28] apply the second-order Newton method to distributed optimization. The paper [29] studies the dual averaging algorithm and the papers [30, 31] investigate the algorithm of Alternating Direction Method of Multipliers. Distributed Nesterov gradient algorithms are developed in [32] to accelerate the convergence. In [33], the authors aim to minimize a sequence of dynamically changing convex functions. In [34, 35], the authors investigate the robustness of distributed algorithms against external disturbances. In [36], game design is utilized to address distributed optimization. In [37], the authors propose a distributed algorithm to compute Pareto optimal solutions of multiobjective optimization problems.

Distributed Lagrangian Primal-Dual Subgradient Algorithm and Distributed Penalty Primal-Dual Subgradient Algorithm presented in this chapter are built on saddle point dynamics. For a convex–concave function, continuous-time saddle point dynamics is proved in [8] to converge globally towards a saddle point. Recently, [9] presents (discrete-time) primal-dual subgradient methods which relax the differentiability of [8] and further incorporate state constraints. The method in [8] is adopted by [38, 39] to study a distributed optimization problem on fixed graphs where objective functions are separable.

References

1. R.D. Nowak, Distributed EM algorithms for density estimation and clustering in sensor networks. IEEE Trans. Signal Process. **51**(8), 2245–2253 (2003)
2. S.S. Ram, A. Nedić, V.V. Veeravalli, Distributed and recursive parameter estimation in parametrized linear state-space models. IEEE Trans. Autom. Control **55**(2), 488–492 (2010)
3. M.G. Rabbat, R.D. Nowak, Decentralized source localization and tracking, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 921–924, May 2004
4. F.P. Kelly, A. Maulloo, D. Tan, Rate control in communication networks: shadow prices, proportional fairness and stability. J. Oper. Res. Soc. **49**(3), 237–252 (1998)
5. A. Oliveira, S. Soares, L. Nepomuceno, Optimal active power dispatch combining network flow and interior point approaches. IEEE Trans. Power Syst. **18**(4), 1235–1240 (2003)
6. H. Wei, H. Sasaki, J. Kubokawa, R. Yokoyama, An interior point nonlinear programming for optimal power flow problems with a novel data structure. IEEE Trans. Power Syst. **13**(3), 870–877 (1998)
7. J. Derenick, J. Spletzer, Convex optimization strategies for coordinating large-scale robot formations. IEEE Trans. Robot. **23**(6), 1252–1259 (2007)
8. K.J. Arrow, L. Hurwicz, H. Uzawa, *Studies in Linear and Nonlinear Programming* (Stanford University Press, Stanford, 1958)
9. A. Nedić, A. Ozdaglar, Subgradient methods for saddle-point problems. J. Optim. Theory Appl. **142**(1), 205–228 (2009)
10. A. Nedić, A. Ozdaglar, Approximate primal solutions and rate analysis for dual subgradient methods. SIAM J. Optim. **19**(4), 1757–1780 (2009)
11. D.P. Bertsekas, *Convex Optimization Theory* (Athena Scientific, Belmont, 2009)
12. D.P. Bertsekas, A. Nedić, A. Ozdaglar, *Convex Analysis and Optimization* (Athena Scientific, Belmont, 2003)
13. A. Nedić, A. Ozdaglar, P.A. Parrilo, Constrained consensus and optimization in multi-agent networks. IEEE Trans. Autom. Control **55**(4), 922–938 (2010)
14. L.S. Lasdon, *Optimization Theory for Large Systems* (Macmillian, New York, 1970)
15. M. Chiang, S.H. Low, A.R. Calderbank, J.C. Doyle, Layering as optimization decomposition: a mathematical theory of network architectures. Proc. IEEE **95**(1), 255–312 (2007)
16. S.H. Low, D.E. Lapsey, Optimization flow control I: basic algorithm and convergence. IEEE/ACM Trans. Netw. **7**(6), 861–874 (1999)
17. D.P. Bertsekas, J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods* (Athena Scientific, Belmont, 1997)
18. J.N. Tsitsiklis, Problems in decentralized decision making and computation. Ph.D. thesis, Massachusetts Institute of Technology. Technical Report LIDS-TH-1424. http://web.mit.edu/jnt/www/Papers/PhD-84-jnt.pdf (1984)
19. M. Zhu, S. Martínez, On distributed convex optimization under inequality and equality constraints. IEEE Trans. Autom. Control **57**(1), 151–164 (2012)
20. M. Zhu, S. Martínez, An approximate dual subgradient algorithm for multi-agent non-convex optimization. IEEE Trans. Autom. Control **58**(6), 1534–1539 (2013)
21. B. Gharesifard, J. Cortés, Distributed continuous-time convex optimization on weight-balanced digraphs. IEEE Trans. Autom. Control **59**(3), 781–786 (2014)
22. J. Lu, C.Y. Tang, Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case. IEEE Trans. Autom. Control **57**(9), 2348–2354 (2012)
23. J. Wang, N. Elia, Control approach to distributed optimization, in *Allerton Conference on Communications, Control and Computing*, Illinois, USA, October 2010
24. S.S. Kia, J. Cortés, S. Martínez, Distributed convex optimization via continuous time coordination algorithms with discrete-time communication. Automatica (2013). Submitted
25. I. Matei, J. Baras, Performance evaluation of the consensus-based distributed subgradient method under random communication topologies. IEEE J. Sel. Top. Signal Process. **5**(4), 754–771 (2011)

26. I. Lobel, A. Ozdaglar, D. Feijer, Distributed multi-agent optimization with state-dependent communication. Math. Program. **129**(2), 255–284 (2011)
27. E. Wei, A. Ozdaglar, A. Jadbabaie, A distributed Newton method for network utility maximization, I: algorithm. IEEE Trans. Autom. Control **58**(9), 2162–2175 (2013)
28. E. Wei, A. Ozdaglar, A. Jadbabaie, A distributed Newton method for network utility maximization, II: convergence. IEEE Trans. Autom. Control **58**(9), 2176–2188 (2013)
29. J. Duchi, A. Agarwal, M. Wainwright, Dual averaging for distributed optimization: convergence analysis and network scaling. IEEE Trans. Autom. Control **57**(3), 592–606 (2012)
30. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. IEEE Trans. Autom. Control **3**(1), 1–122 (2011)
31. E. Wei, A. Ozdaglar, On the O(1/k) convergence of asynchronous distributed alternating direction method of multipliers. Math. Program. (2013). Submitted
32. D. Jakovetic, J. Xavier, J. Moura, Fast distributed gradient methods. IEEE Trans. Autom. Control **59**(5), 1131–1146 (2014)
33. R. Cavalcante, A. Rogers, N. Jennings, I. Yamada, Distributed asymptotic minimization of sequences of convex functions by a broadcast adaptive subgradient method. IEEE J. Sel. Top. Signal Process. **5**(4), 739–753 (2011)
34. D. Richert, J. Cortés, Integral input-to-state stable saddle-point dynamics for distributed linear programming, in *IEEE International Conference on Decision and Control* (Florence, Italy, 2013), pp. 7480–7485
35. D. Mateos-Núñez, J. Cortés, Noise-to-state stable distributed convex optimization on weight-balanced digraphs, in *IEEE International Conference on Decision and Control* (Florence, Italy, 2013), pp. 2781–2786
36. N. Li, J. Marden, Designing games for distributed optimization. IEEE J. Sel. Top. Signal Process. **7**(2), 230–242 (2013)
37. J. Chen, A. Sayed, Distributed Pareto optimization via diffusion strategies. IEEE J. Sel. Top. Signal Process. **7**(2), 205–220 (2013)
38. P. Martin, M. Egerstedt, Optimization of multi-agent motion programs with applications to robotic marionettes. Hybrid Syst. Comput. Control (2009)
39. A. Rantzer, Using game theory for distributed control engineering, in *World Congress of the Game Theory Society* (2008)

http://www.springer.com/978-3-319-19071-6

Distributed Optimization-Based Control of Multi-Agent Networks in Complex Environments
Zhu, M.; Martínez, S.
2015, XIII, 124 p. 24 illus., 23 illus. in color., Softcover
ISBN: 978-3-319-19071-6