An approximate dual subgradient algorithm for multi-agent non-convex optimization

Minghui Zhu and Sonia Martínez

Abstract— **We consider a multi-agent optimization problem where agents aim to cooperatively minimize a sum of local objective functions subject to a global inequality constraint and a global state constraint set. In contrast to existing papers, we do not require that the objective, constraint functions, and state constraint sets are convex. We propose a distributed approximate dual subgradient algorithm to enable agents to asymptotically converge to a pair of approximate primal-dual solutions over dynamically changing network topologies. Convergence can be guaranteed provided that the Slater's condition and strong duality property are satisfied.**

I. Introduction

Recent advances in computation, communication, sensing and actuation have stimulated an intensive research in networked multi-agent systems. In the systems and control community, this has been translated into how to solve global control problems, expressed by global objective functions, by means of local agent actions. More specifically, problems considered include multi-agent consensus or agreement [5], [12], [14], [18], [23], [24], coverage control [6], [8], formation control [9], [28], sensor fusion [32] and game-theoretic control [1], [27].

In the optimization community, a problem of focus is to minimize a sum of local objective functions by a group of agents, where each function depends on a common global decision vector and is only known to a specific agent. This problem is motivated by others in distributed estimation [22] [31], distributed source localization [25], and network utility maximization [15]. More recently, consensus techniques have been proposed to address the issues of switching topologies in networks and non-separability in objective functions; see for instance [13], [20], [21], [26], [33]. More specifically, the paper [20] presents the first analysis of an algorithm that combines average consensus schemes with subgradient methods. Using projection in the algorithm of [20], the authors in [21] further solve a more general setup that includes local state constraint sets. Further, in [33] we develop two distributed primal-dual subgradient algorithms, which are based on saddle-point theorems, to analyze a more general situation that incorporates global inequality and equality constraints. The aforementioned algorithms are extensions of classic (primal or primal-dual) subgradient methods which generalize gradient-based methods to minimize non-smooth functions. This requires the optimization problems under consideration to be convex in order to determine a global optimum.

The focus of the current paper is to relax the convexity assumption in [33]. To achieve this, we will integrate Lagrangian dualization and subgradient schemes to circumvent the non-convexity property, which have been popular and efficient approaches to solve large-scale, structured convex optimization problems, e.g., [3], [4]. In particular, these two techniques have been successfully utilized to design decentralized resource allocation algorithms; see [7], [15], [30], in the networking community. However, subgradient methods do not automatically generate primal solutions for nonsmooth convex optimization problems. Numerous approaches have been designed to construct primal solutions; e.g., by removing the nonsmoothness [29], by employing ascent approaches [16], and the generation of ergodic sequences [17], [19].

Statement of Contributions. Here, we investigate a multi-agent optimization problem where agents are trying to minimize a sum of local objective functions subject to a global inequality constraint and a global state constraint set. The objective and constraint functions as well as the state-constraint set could be non-convex. A distributed approximate dual subgradient algorithm is introduced to find a pair of approximate primal-dual solutions. Specifically, the update rule for dual estimates combines an approximate dual subgradient scheme with average consensus algorithms. To obtain primal solutions from dual estimates, we propose a novel recovery scheme: primal estimates are not updated if the variations induced by dual estimates are smaller than some predetermined threshold; otherwise, primal estimates are set to some solutions in dual optimal solution sets. This algorithm is shown to asymptotically converge to a pair of approximate primal-dual solutions over a class of switching network topologies. Convergence is guaranteed under the Slater's condition and strong duality property.

II. Problem formulation and preliminaries

Consider a networked multi-agent system where agents are labeled by $i \in V := \{1, \dots, N\}$. The multi-agent system operates in a synchronous way at time instants $k \in \mathbb{N} \cup \{0\}$, and its topology will be represented by a directed weighted graph $\mathcal{G}(k) = (V, E(k), A(k))$, for $k \geq 0$. Here, $A(k) := [a_j^i(k)] \in \mathbb{R}^{N\times N}$ is the adjacency matrix, where the scalar $a_j^i(k) \geq 0$ is the weight assigned to the edge (j, i), and $E(k) \subseteq V \times V \setminus \operatorname{diag}(V)$ is the set of edges with non-zero weights. The set of in-neighbors of agent i at time k

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

is denoted by $\mathcal{N}_i(k) = \{j \in V \mid (j, i) \in E(k) \text{ and } j \neq i\}$. Similarly, we define the set of out-neighbors of agent i at time k as $\mathcal{N}_i^{\text{out}}(k) = \{j \in V \mid (i, j) \in E(k) \text{ and } j \neq i\}$. We here make the following assumptions on the network communication graphs:

Assumption 2.1 (Non-degeneracy): There exists a constant $\alpha > 0$ such that $a_i^i(k) \geq \alpha$, and $a_j^i(k)$, for $i \neq j$, satisfies $a_j^i(k) \in \{0\} \cup [\alpha, 1]$, for all $k \geq 0$.

Assumption 2.2 (Balanced Communication): [1]It holds that $\sum_{j \in V} a_j^i(k) = 1$ for all $i \in V$ and $k \geq 0$, and $\sum_{i \in V} a_j^i(k) = 1$ for all $j \in V$ and $k \geq 0$.

Assumption 2.3 (Periodical Strong Connectivity): There is a positive integer B such that, for all $k_0 \geq 0$, the directed graph $(V, \bigcup_{k=0}^{B-1} E(k_0 + k))$ is strongly connected.

The above network model is standard in the analysis of average consensus algorithms; e.g., see [23], [24], and distributed optimization in [21], [33]. Recently, an algorithm is given in [10] which allows agents to construct a balanced graph out of a non-balanced one under certain assumptions.

The objective of the agents is to cooperatively solve the following primal problem (P):

$$\begin{aligned} &\min_{z \in \mathbb{R}^n} \sum_{i \in V} f_i(z), \\ &\text{s.t.} \quad g(z) \leq 0, \quad z \in X, \end{aligned} \tag{1}$$

where $z \in \mathbb{R}^n$ is the global decision vector. The function $f_i : \mathbb{R}^n \to \mathbb{R}$ is only known to agent i, continuous, and referred to as the objective function of agent i. The set $X \subseteq \mathbb{R}^n$, the state constraint set, is compact. The function $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuous, and the inequality $g(z) \leq 0$ is understood component-wise; i.e., $g_\ell(z) \leq 0$, for all $\ell \in \{1, \ldots, m\}$, and represents a global inequality constraint. We will denote $f(z) := \sum_{i \in V} f_i(z)$ and $Y := \{z \in \mathbb{R}^n \mid g(z) \leq 0\}$. We will assume that the set of feasible points is non-empty; i.e., $X \cap Y \neq \emptyset$. Since X is compact and Y is closed, then we can deduce that $X \cap Y$ is compact. The continuity of f follows from that of f_i. In this way, the optimal value p^* of the problem (P) is finite and X^*, the set of primal optimal points, is non-empty. Throughout this paper, we suppose the following Slater's condition holds:

Assumption 2.4 (Slater's Condition): There exists a vector $\bar{z} \in X$ such that $g(\bar{z}) < 0$. Such $\bar{z}$ is referred to as a Slater vector of the problem (P).

Remark 2.1: All the agents can agree upon a common Slater vector $\bar{z}$ through a maximum-consensus scheme:

Initially, each agent i chooses a Slater vector $z_i(0) \in X$ such that $g(z_i(0)) < 0$. At every time $k \geq 0$, each agent i updates its estimates by using the following rule:

$$z_i(k+1) = \max_{j \in \mathcal{N}_i(k) \cup \{i\}} z_j(k). \tag{2}$$

where we use the following relation for vectors: for $a, b \in \mathbb{R}^n$, $a < b$ if and only if there is some $\ell \in \{1, \ldots, n-1\}$ such that $a_\kappa = b_\kappa$ for all $\kappa < \ell$ and $a_\ell < b_\ell$.

The periodical strong connectivity assumption 2.3 ensures that after at most $(N-1)B$ steps, all the agents reach the consensus; i.e., $z_i(k) = \max_{j \in V} z_j(0)$ for all $k \geq (N-1)B$. In the remainder of this paper, we assume that the Slater vector $\bar{z}$ is known to all the agents. •

[1]It is also referred to as double stochasticity.

In [33], in order to solve the convex case of the problem (P), we propose two distributed primal-dual subgradient algorithms where primal (resp. dual) estimates move along subgradients (resp. supgradients) and are projected onto convex sets. The absence of convexity impedes the use of the algorithms in [33] since, on the one hand, (primal) gradient-based algorithms are easily trapped in local minima.; on the other hand, projection maps may not be well-defined when (primal) state constraint sets are non-convex. In this paper, we will employ Lagrangian dualization to circumvent the challenges caused by non-convexity.

We first construct a directed cyclic graph $\mathcal{G}_{\text{cyc}} := (V, E_{\text{cyc}})$ where $|E_{\text{cyc}}| = N$. We assume that each agent has a unique in-neighbor (and out-neighbor). The out-neighbor (resp. in-neighbor) of agent i is denoted by i_D (resp. i_U). With the graph $\mathcal{G}_{\text{cyc}}$, we will study the following approximate problem of problem (P):

$$\begin{aligned} &\min_{(x_i) \in \mathbb{R}^{nN}} \sum_{i \in V} f_i(x_i), \\ &\text{s.t.} \quad g(x_i) \leq 0, \quad , -x_i + x_{i_D} - \Delta \leq 0 \\ &x_i - x_{i_D} - \Delta \leq 0, \quad x_i \in X, \quad \forall i \in V, \end{aligned} \tag{3}$$

where $\Delta := \delta \mathbf{1}$, with δ a small positive scalar, and $\mathbf{1}$ is the column vector of n ones. The problem (3) reduces to the problem (P) when $\delta = 0$, and will be referred to as problem (P_Δ). Its optimal value and the set of optimal solutions will be denoted by p_Δ^* and X_Δ^*, respectively. Similarly to the problem (P), p_Δ^* is finite and $X_\Delta^* \neq \emptyset$.

Remark 2.2: The cyclic graph $\mathcal{G}_{\text{cyc}}$ can be replaced by any strongly connected graph. Each agent i is endowed with two inequality constraints: $x_i - x_j - \Delta \leq 0$ and $-x_i + x_j - \Delta \leq 0$, for each out-neighbor j. For notational simplicity, we will use the cyclic graph $\mathcal{G}_{\text{cyc}}$, which has a minimum number of constraints, as the initial graph. •

A. Dual problems

Before introducing dual problems, let us denote by $\Xi_i := \mathbb{R}^m_{\geq 0} \times \mathbb{R}^{nN}_{\geq 0} \times \mathbb{R}^{nN}_{\geq 0}$, $\Xi := \mathbb{R}^{mN}_{\geq 0} \times \mathbb{R}^{nN}_{\geq 0} \times \mathbb{R}^{nN}_{\geq 0}$, $\xi_i := (\mu_i, \lambda, w) \in \Xi_i$, $\xi := (\mu, \lambda, w) \in \Xi$ and $x := (x_i) \in X^N$. The dual problem (D_Δ) associated with (P_Δ) is given by

$$\max_{\mu, \lambda, w} Q(\mu, \lambda, w), \quad \text{s.t.} \quad \mu, \lambda, w \geq 0, \tag{4}$$

where $\mu := (\mu_i) \in \mathbb{R}^{mN}$, $\lambda := (\lambda_i) \in \mathbb{R}^{nN}$ and $w := (w_i) \in \mathbb{R}^{nN}$. Here, the dual function $Q : \Xi \to \mathbb{R}$ is given as

$$Q(\xi) \equiv Q(\mu, \lambda, w) := \inf_{x \in X^N} \mathcal{L}(x, \mu, \lambda, w),$$

where $\mathcal{L} : \mathbb{R}^{nN} \times \Xi \to \mathbb{R}$ is the Lagrangian function

$$\begin{aligned} \mathcal{L}(x, \xi) \equiv \mathcal{L}(x, \mu, \lambda, w) := \sum_{i \in V} \big(f_i(x_i) + \langle \mu_i, g(x_i) \rangle \\ + \langle \lambda_i, -x_i + x_{i_D} - \Delta \rangle + \langle w_i, x_i - x_{i_D} - \Delta \rangle \big). \end{aligned}$$

We denote the dual optimal value of the problem (D_Δ) by d^*_Δ and the set of dual optimal solutions by D^*_Δ. In what follows we will assume that the duality gap is zero.

Assumption 2.5 (Strong duality): For the introduced problems (P_Δ) and (D_Δ), it holds that $p^*_\Delta = d^*_\Delta$.

We endow each agent i with the local Lagrangian function $\mathcal{L}_i : \mathbb{R}^n \times \Xi_i \to \mathbb{R}$ and the local dual function $Q_i : \Xi_i \to \mathbb{R}$ defined by

$$\begin{aligned}\mathcal{L}_i(x_i, \xi_i) &:= f_i(x_i) + \langle \mu_i, g(x_i)\rangle + \langle -\lambda_i + \lambda_{i_U}, x_i\rangle \\ &\quad + \langle w_i - w_{i_U}, x_i\rangle - \langle \lambda_i, \Delta\rangle - \langle w_i, \Delta\rangle, \\ Q_i(\xi_i) &:= \inf_{x_i\in X} \mathcal{L}_i(x_i, \xi_i).\end{aligned}$$

In the problem (P_Δ), the introduction of approximate consensus constraints $-\Delta \le x_i - x_{i_D} \le \Delta$, $i \in V$, renders the f_i and g separable. As a result, the global dual function Q can be decomposed into a simple sum of the local dual functions Q_i. More precisely, the following holds:

$$\begin{aligned}Q(\xi) &= \inf_{x\in X^N} \sum_{i\in V} \big(f_i(x_i) + \langle \mu_i, g(x_i)\rangle \\ &+ \langle \lambda_i, -x_i + x_{i_D} - \Delta\rangle + \langle w_i, x_i - x_{i_D} - \Delta\rangle\big) \\ &= \inf_{x\in X^N} \sum_{i\in V} \big(f_i(x_i) + \langle \mu_i, g(x_i)\rangle \\ &+ \langle -\lambda_i + \lambda_{i_U}, x_i\rangle + \langle w_i - w_{i_U}, x_i\rangle - \langle \lambda_i, \Delta\rangle - \langle w_i, \Delta\rangle\big) \\ &= \sum_{i\in V} \inf_{x_i\in X} \big(f_i(x_i) + \langle \mu_i, g(x_i)\rangle \\ &+ \langle -\lambda_i + \lambda_{i_U}, x_i\rangle + \langle w_i - w_{i_U}, x_i\rangle - \langle \lambda_i, \Delta\rangle - \langle w_i, \Delta\rangle\big) \\ &= \sum_{i\in V} Q_i(\xi_i). \end{aligned} \tag{5}$$

It is worth mentioning that $\sum_{i\in V} Q_i(\xi_i)$ is not separable since Q_i depends upon neighbor's multipliers λ_{i_U} and w_{i_U}.

B. Dual solution sets

The Slater's condition ensures the boundedness of dual solution sets for convex optimization; e.g., [11], [19]. We will shortly see that the Slater's condition plays the same role in non-convex optimization. To achieve this, we define the function $\hat{Q}_i : \mathbb{R}^m_{\ge 0} \times \mathbb{R}^n_{\ge 0} \times \mathbb{R}^n_{\ge 0} \to \mathbb{R}$ as follows:

$$\begin{aligned}\hat{Q}_i(\mu_i, \lambda_i, w_i) &= \inf_{x_i\in X, x_{i_D}\in X} \big(f_i(x_i) + \langle \mu_i, g(x_i)\rangle \\ &+ \langle \lambda_i, -x_i + x_{i_D} - \Delta\rangle + \langle w_i, x_i - x_{i_D} - \Delta\rangle\big).\end{aligned}$$

Let $\bar{z}$ be a Slater vector for problem (P). Then $\bar{x} = (\bar{x}_i) \in X^N$ with $\bar{x}_i = \bar{z}$ is a Slater vector of the problem (P_Δ). Similarly to (3) and (4) in [33], which make use of Lemma 3.2 in the same paper, we have that for any $\mu_i, \lambda_i, w_i \ge 0$, it holds that

$$\max_{\xi\in D^*_\Delta} \|\xi\| \le N \max_{i\in V} \frac{f_i(\bar{z}) - \hat{Q}_i(\mu_i, \lambda_i, w_i)}{\beta(\bar{z})}, \tag{6}$$

where $\beta(\bar{z}) := \min\{\min_{\ell\in\{1,\dots,m\}} -g_\ell(\bar{z}), \delta\}$. Let μ_i, λ_i and w_i be zero in (6), and it leads to the following upper bound on D^*_Δ:

$$\max_{\xi\in D^*_\Delta} \|\xi\| \le N \max_{i\in V} \frac{f_i(\bar{z}) - \hat{Q}_i(0, 0, 0)}{\beta(\bar{z})}, \tag{7}$$

where $\hat{Q}_i(0, 0, 0)$ can be computed locally. Since f_i and g are continuous and X is compact, it is known that Q_i is continuous; e.g., see Theorem 1.4.16 in [2]. Similarly, Q is continuous. Since D^*_Δ is also bounded, then we have that $D^*_\Delta \ne \emptyset$.

Remark 2.3: From the above analysis of D^*_Δ, it can be seen that if $\delta = 0$, which corresponds to the exact consensus, then D^*_Δ could be unbounded and empty. •

Denote by $D^\epsilon_\Delta := \{\xi \in \Xi \mid Q(\xi) \ge d^*_\Delta - N\epsilon\}$. Similar to (7), we have the following upper bound on D^ϵ_Δ:

$$\max_{\xi\in D^\epsilon_\Delta} \|\xi\| \le N \max_{i\in V} \frac{f_i(\bar{z}) - \hat{Q}_i(0, 0, 0) + \epsilon}{\beta(\bar{z})}. \tag{8}$$

In the algorithm we will present in the following section, agents will compute $\gamma_i(\bar{z}) := \frac{f_i(\bar{z}) - \hat{Q}_i(0,0,0)+\epsilon}{\beta(\bar{z})}$.

C. Other notation

Define the set-valued map $\Omega_i : \Xi_i \to 2^X$ in the following way $\Omega_i(\xi_i) := \mathrm{argmin}_{x_i\in X}\mathcal{L}_i(x_i, \xi_i)$; i.e., given ξ_i, the set $\Omega_i(\xi_i)$ is the collection of solutions to the following local optimization problem:

$$\min_{x_i\in X} \mathcal{L}_i(x_i, \xi_i). \tag{9}$$

Here, Ω_i is referred to as the *marginal map* of agent i. Since X is compact and f_i, g are continuous, then $\Omega_i(\xi_i) \ne \emptyset$ in (9) for any $\xi_i \in \Xi_i$. In the algorithm we will develop in next section, each agent is required to solve the local optimization problem (9) at each iterate. We assume that this problem (9) can be easily solved. This is the case for problems where f_i and g are smooth. For some $\epsilon > 0$, we define the set-valued map $\Omega^\epsilon_i : \Xi_i \to 2^X$ as follows:

$$\Omega^\epsilon_i(\xi_i) := \{x_i \in X \mid \mathcal{L}_i(x_i, \xi_i) \le Q_i(\xi_i) + \epsilon\},$$

which is referred to as the *approximate marginal map* of agent $i \in V$.

In the space $\mathbb{R}^n$, we define the distance between a point $z \in \mathbb{R}^n$ to a set $A \subset \mathbb{R}^n$ as $\mathrm{dist}(z, A) := \inf_{y\in A} \|z - y\|$, and the Hausdorff distance between two sets $A, B \subset \mathbb{R}^n$ as $\mathrm{dist}(A, B) := \max\{\sup_{z\in A} \mathrm{dist}(z, B), \sup_{y\in B} \mathrm{dist}(A, y)\}$. We denote by $B_{\mathcal{U}}(A, r) := \{u \in \mathcal{U} \mid \mathrm{dist}(u, A) \le r\}$ and $B_{2^{\mathcal{U}}}(A, r) := \{U \in 2^{\mathcal{U}} \mid \mathrm{dist}(U, A) \le r\}$ where $\mathcal{U} \subset \mathbb{R}^n$.

III. Distributed approximate dual subgradient algorithm

In this section, we devise a distributed approximate dual subgradient algorithm which aims to find a pair of approximate primal-dual solutions to the problem (P_Δ). Its convergence properties are also summarized.

For each agent i, let $x_i(k) \in \mathbb{R}^n$ be the estimate of the primal solution x_i to the problem (P_Δ) at time $k \ge 0$, $\mu_i(k) \in \mathbb{R}^m_{\ge 0}$ be the estimate of the multiplier on the inequality constraint $g(x_i) \le 0$, $\lambda^i(k) \in \mathbb{R}^{nN}_{\ge 0}$ (resp. $w^i(k) \in \mathbb{R}^{nN}_{\ge 0}$)[2] be the estimate of the multiplier associated with the collection of the local inequality constraints $-x_j + x_{j_D} -$

[2]We will use the superscript i to indicate that $\lambda^i(k)$ and $w^i(k)$ are estimates of some global variables.

$\Delta \leq 0$ (resp. $x_j - x_{j_D} - \Delta \leq 0$), for all $j \in V$. We let $\xi_i(k) := (\mu_i(k)^T, \lambda^i(k)^T, w^i(k)^T)^T$, for $i \in V$, and $v_i(k) := (\mu_i(k)^T, v^i_\lambda(k)^T, v^i_w(k)^T)^T$ where $v^i_\lambda(k) := \sum_{j\in V} a^i_j(k)\lambda^j(k)$ and $v^i_w(k) := \sum_{j\in V} a^i_j(k)w^j(k)$.

The *Distributed Approximate Dual Subgradient* (DADS, for short) Algorithm is described as follows.

Initially, each agent i chooses a common Slater vector $\bar{z}$ and computes $\gamma := N\max_{i\in V}\gamma_i(\bar{z})$ through a max-consensus algorithm. After that, each agent i chooses initial states $x_i(0) \in X$ and $\xi_i(0) \in \Xi_i$.

Agent i updates $x_i(k)$ and $\xi_i(k)$ as follows:

Step 1. For each $k \geq 1$, given $v_i(k)$, solve the local optimization problem (9), obtain the dual solution set $\Omega_i(v_i(k))$ and the dual optimal value $Q_i(v_i(k))$. Produce the primal estimate $x_i(k)$ as follows: if $x_i(k-1) \in \Omega^\epsilon_i(v_i(k))$, then $x_i(k) = x_i(k-1)$; otherwise, choose $x_i(k) \in \Omega_i(v_i(k))$.

Step 2. For each $k \geq 0$, generate the dual estimate $\xi_i(k+1)$ according to the following rule:

$$\xi_i(k+1) = P_{M_i}[v_i(k) + \alpha(k)\mathcal{D}_i(k)], \qquad (10)$$

where the scalar $\alpha(k)$ is a step-size. The supgradient vector of agent i is defined as $\mathcal{D}_i(k) := (\mathcal{D}^i_\mu(k)^T, \mathcal{D}^i_\lambda(k)^T, \mathcal{D}^i_w(k)^T)^T$, where $\mathcal{D}^i_\mu(k) := g(x_i(k)) \in \mathbb{R}^m$, $\mathcal{D}^i_\lambda(k)$ has components $\mathcal{D}^i_\lambda(k)_i := -\Delta - x_i(k) \in \mathbb{R}^n$, $\mathcal{D}^i_\lambda(k)_{i_U} := x_i(k) \in \mathbb{R}^n$, and $\mathcal{D}^i_\lambda(k)_j = 0 \in \mathbb{R}^n$ for $j \in V\setminus\{i, i_U\}$, while the components of $\mathcal{D}^i_w(k)$ are given by: $\mathcal{D}^i_w(k)_i := -\Delta + x_i(k) \in \mathbb{R}^n$, $\mathcal{D}^i_w(k)_{i_U} := -x_i(k) \in \mathbb{R}^n$, and $\mathcal{D}^i_w(k)_j = 0 \in \mathbb{R}^n$, for $j \in V\setminus\{i, i_U\}$. The set M_i in the projection map, P_{M_i}, above is defined as $M_i := \{\xi_i \in \Xi_i \mid \|\xi_i\| \leq \gamma + \theta\}$ for some $\theta > 0$.

Remark 3.1: In the initialization of the DADS algorithm, the quantity γ is an upper bound on D^ϵ_Δ. Note that in Step 1, the check $x_i(k-1) \in \Omega^\epsilon_i(v_i(k))$ reduces to verifying that $\mathcal{L}_i(x_i(k-1), v_i(k)) \leq Q_i(v_i(k)) + \epsilon$. Then, only if $\mathcal{L}_i(x_i(k-1), v_i(k)) > Q_i(v_i(k)) + \epsilon$, it is necessary to find *one solution* in $\Omega_i(v_i(k))$. That is, it is unnecessary to compute all the set $\Omega_i(v_i(k))$. In Step 2, since M_i is closed and convex, the projection map P_{M_i} is well-defined. •

The primal and dual estimates in the DADS algorithm will be shown to asymptotically converge to a pair of approximate primal-dual solutions to the problem (P_Δ). We formally state this in the following.

Theorem 3.1: Consider the problem (P_Δ) and let the non-degeneracy assumption 2.1, the balanced communication assumption 2.2 and the periodic strong connectivity assumption 2.3 hold. In addition, suppose the Slater's condition 2.4 and the strong duality assumption 2.5 hold. Consider the dual sequences of $\{\mu_i(k)\}$, $\{\lambda^i(k)\}$, $\{w^i(k)\}$ and the primal sequence of $\{x_i(k)\}$ of the distributed approximate 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 exists a feasible dual pair $\tilde{\xi} := (\tilde{\mu}, \tilde{\lambda}, \tilde{w})$ such that $\lim_{k\to+\infty}\|\mu_i(k) - \tilde{\mu}_i\| = 0$, $\lim_{k\to+\infty}\|\lambda^i(k) - \tilde{\lambda}\| = 0$, and $\lim_{k\to+\infty}\|w^i(k) - \tilde{w}\| = 0$, for all $i \in V$. Moreover, there is a feasible primal vector $\tilde{x} := (\tilde{x}_i) \in X^N$ such that $\lim_{k\to+\infty}\|x_i(k) - \tilde{x}_i\| = 0$, for all $i \in V$. In addition, $(\tilde{x}, \tilde{\xi})$ is a pair of approximate primal-dual solutions in the sense that $d^*_\Delta - N\epsilon \leq Q(\tilde{\xi}) \leq d^*_\Delta = p^*_\Delta \leq \sum_{i\in V} f_i(\tilde{x}_i) \leq p^*_\Delta + N\epsilon$.

The analysis of Theorem 3.1 will be provided in next section. Before doing that, we would like to discuss several possible extensions of Theorem 3.1.

Firstly, the step-size scheme in the DADS algorithm can be slightly generalized to the following: $\lim_{k\to+\infty}\alpha_i(k) = 0$, $\sum_{k=0}^{+\infty}\alpha_i(k) = +\infty$, $\sum_{k=0}^{+\infty}\alpha_i(k)^2 < +\infty$, $\min_{i\in V}\alpha_i(k) \geq C_\alpha \max_{i\in V}\alpha_i(k)$, where $\alpha_i(k)$ is the step-size of agent i at time k and $C_\alpha \in (0, 1]$.

Secondly, the periodic strong connectivity assumption 2.3 can be weakened into the eventual strong connectivity assumption, e.g. Assumption 6.1 in [33], if $\mathcal{G}(k)$ is undirected.

Thirdly, each agent can use a different ϵ_i in Step 1 of the DADS algorithm, which would lead to replacing $N\epsilon$ in the approximate solution by $\sum_{i\in V}\epsilon_i$.

Lastly, each agent i could have different constraint functions g_i and constraint sets X_i if a Slater vector is known to all the agents. For example, consider the case that g is convex, X_i is convex and potentially different, and there is a Slater vector $\bar{z} \in \cap_{i\in V} X_i$. Then the solution $\tilde{z}$ to the following problem is such that $g(\tilde{z}) \leq g(\bar{z}) < 0$:

$$\min_{z\in\mathbb{R}^n} Ng(z), \quad \text{s.t.} \quad z \in X_i, \ \ \forall i \in V \qquad (11)$$

Through implementing the distributed primal subgradient algorithm in [33], agents can solve the problem (11) in a distributed fashion and agree upon the minimizer $\tilde{z}$ which coincides with a Slater vector. In such a way, Theorem 3.1 still holds and the corresponding proof is a slight variation of those in next section.

IV. Convergence analysis

Recall that g is continuous and X is compact. Then there are $G, H > 0$ such that $\|g(z)\| \leq G$ and $\|z\| \leq H$ for all $z \in X$. We start our analysis of the DADS algorithm from the computation of supgradients of Q_i.

Lemma 4.1 (Approximate supgradient): If $\bar{x}_i \in \Omega^\epsilon_i(\bar{\xi}_i)$, then $\big(g(\bar{x}_i)^T, (-\Delta - \bar{x}_i)^T, \bar{x}_i^T, (\bar{x}_i - \Delta)^T, -\bar{x}_i^T\big)^T$ is an approximate supgradient of Q_i at $\bar{\xi}_i$; i.e., the following holds for any $\xi_i \in \Xi_i$:

$$\begin{aligned}&Q_i(\xi_i) - Q_i(\bar{\xi}_i) \leq \langle g(\bar{x}_i), \mu_i - \bar{\mu}_i\rangle + \langle -\Delta - \bar{x}_i, \lambda_i - \bar{\lambda}_i\rangle\\ &+ \langle \bar{x}_i, \lambda_{i_U} - \bar{\lambda}_{i_U}\rangle + \langle \bar{x}_i - \Delta, w_i - \bar{w}_i\rangle\\ &+ \langle -\bar{x}_i, w_{i_U} - \bar{w}_{i_U}\rangle + \epsilon. \qquad (12)\end{aligned}$$

Proof: The proof is analogous to the computation of dual subgradients, e.g., in [3], [4], and omitted here due to the space limitation. ■

Since $\Omega_i(v_i(k)) \subseteq \Omega^\epsilon_i(v_i(k))$, it is clear that $x_i(k) \in \Omega^\epsilon_i(v_i(k))$ for all $k \geq 0$. A direct result of Lemma 4.1 is that the vector $(g(x_i(k))^T, (-\Delta - x_i(k))^T, x_i(k)^T, (x_i(k) - \Delta)^T, -x_i(k)^T)$ is an approximate supgradient of Q_i at $v_i(k)$;

i.e., the following approximate supgradient inequality holds for any $\xi_i \in \Xi_i$:

$$
\begin{aligned}
&Q_i(\xi_i) - Q_i(v_i(k)) \le \langle g(x_i(k)), \mu_i - \mu_i(k)\rangle \\
&+ \langle -\Delta - x_i(k), \lambda_i - v^i_\lambda(k)_i\rangle \\
&+ \langle x_i(k), \lambda_{i_U} - v^i_\lambda(k)_{i_U}\rangle + \langle x_i(k) - \Delta, w_i - v^i_w(k)_i\rangle \\
&+ \langle -x_i(k), w_{i_U} - v^i_w(k)_{i_U}\rangle + \epsilon. \qquad (13)
\end{aligned}
$$

Now we can see that the update rule of dual estimates in the DADS algorithm is a combination of an approximate dual subgradient scheme and average consensus algorithms. The following establishes that Q_i is Lipschitz continuous with some Lipschitz constant L.

Lemma 4.2 (Lipschitz continuity of Q_i): There is a constant $L > 0$ such that for any $\xi_i, \bar{\xi}_i \in \Xi_i$, it holds that

$$\|Q_i(\xi_i) - Q_i(\bar{\xi}_i)\| \le L\|\xi_i - \bar{\xi}_i\|.$$

Proof: Similarly to Lemma 4.1, one can show that if $\bar{x}_i \in \Omega_i(\bar{\xi}_i)$, then $(g(\bar{x}_i)^T, (-\Delta - \bar{x}_i)^T, \bar{x}_i^T, (\bar{x}_i - \Delta)^T, -\bar{x}_i^T)^T$ is a supgradient of Q_i at $\bar{\xi}_i$; i.e., the following holds for any $\xi_i \in \Xi_i$:

$$
\begin{aligned}
&Q_i(\xi_i) - Q_i(\bar{\xi}_i) \le \langle g(\bar{x}_i), \mu_i - \bar{\mu}_i\rangle + \langle -\Delta - \bar{x}_i, \lambda_i - \bar{\lambda}_i\rangle \\
&+ \langle \bar{x}_i, \lambda_{i_U} - \bar{\lambda}_{i_U}\rangle + \langle \bar{x}_i - \Delta, w_i - \bar{w}_i\rangle \\
&+ \langle -\bar{x}_i, w_{i_U} - \bar{w}_{i_U}\rangle.
\end{aligned}
$$

Since $\|g(\bar{x}_i)\| \le G$ and $\|\bar{x}_i\| \le H$, there is $L > 0$ such that $Q_i(\xi_i) - Q_i(\bar{\xi}_i) \le L\|\xi_i - \bar{\xi}_i\|$. Similarly, $Q_i(\bar{\xi}_i) - Q_i(\xi_i) \le L\|\xi_i - \bar{\xi}_i\|$. The combination of these two relations renders the desired result. ■

In the DADS algorithm, the error induced by the projection map P_{M_i} is given by:

$$e_i(k) := P_{M_i}[v_i(k) + \alpha(k)\mathcal{D}_i(k)] - v_i(k).$$

We next provide a basic iterate relation of dual estimates in the DADS algorithm.

Lemma 4.3 (Basic iterate relation): Under the assumptions in Theorem 3.1, for any $((\mu_i), \lambda, w) \in \Xi$ with $(\mu_i, \lambda, w) \in M_i$ for all $i \in V$, the following estimate holds for all $k \ge 0$:

$$
\begin{aligned}
&\sum_{i\in V}\|e_i(k) - \alpha(k)\mathcal{D}_i(k)\|^2 \le \alpha(k)^2\sum_{i\in V}\|\mathcal{D}_i(k)\|^2 \\
&+ \sum_{i\in V}(\|\xi_i(k) - \xi_i\|^2 - \|\xi_i(k+1) - \xi_i\|^2) \\
&+ 2\alpha(k)\sum_{i\in V}\{\langle g(x_i(k)), \mu_i(k) - \mu_i\rangle \\
&+ \langle -\Delta - x_i(k), v^i_\lambda(k)_i - \lambda_i\rangle \\
&+ \langle x_i(k), v^i_\lambda(k)_{i_U} - \lambda_{i_U}\rangle + \langle x_i(k) - \Delta, v^i_w(k)_i - w_i\rangle \\
&+ \langle -x_i(k), v^i_w(k)_{i_U} - w_{i_U}\rangle\}. \qquad (14)
\end{aligned}
$$

Proof: Recall that M_i is closed and convex. The proof is an application of Lemma 6.1 in the Appendix. ■

The lemma below shows that dual estimates asymptotically converge to some approximate dual optimal solution.

Lemma 4.4 (Dual estimate convergence): Under the assumptions in Theorem 3.1, there exist a feasible dual pair $\tilde{\xi} := ((\tilde{\mu}_i), \tilde{\lambda}, \tilde{w})$ such that $\lim_{k\to+\infty}\|\mu_i(k) - \tilde{\mu}_i\| = 0$, $\lim_{k\to+\infty}\|\lambda^i(k) - \tilde{\lambda}\| = 0$, and $\lim_{k\to+\infty}\|w^i(k) - \tilde{w}\| = 0$. Furthermore, the vector $\tilde{\xi}$ is an approximate dual solution to the problem (D_Δ) in the sense that $d^*_\Delta - N\epsilon \le Q(\tilde{\xi}) \le d^*_\Delta$.

Proof: By the dual decomposition property (5) and the boundedness of dual optimal solution sets, the dual problem (D_Δ) is equivalent to the following:

$$\max_{(\xi_i)}\sum_{i\in V}Q_i(\xi_i), \quad \text{s.t.} \quad \xi_i \in M_i. \qquad (15)$$

Note that Q_i is affine and M_i is convex, implying that the problem (15) is a constrained convex programming where the global objective function is a simple sum of local ones and the local state constraints are compact.

Since X and M_i are compact, there is some $J > 0$ which is an upper bound of the norm of the last sum on the right-hand side of (14). In this way, inequality (14) leads to:

$$
\begin{aligned}
&\sum_{i\in V}\|\xi_i(K) - \xi_i\|^2 \le \sum_{i\in V}\|\xi_i(K') - \xi_i\|^2 \\
&+ \alpha(K')^2\sum_{i\in V}\|\mathcal{D}_i(K')\|^2 + 2\alpha(K')J, \qquad (16)
\end{aligned}
$$

where $K = K' + 1$. It is not difficult to see that the sequence of $\{\mathcal{D}_i(k)\}$ is uniformly bounded. Since $\lim_{k\to+\infty}\alpha(k) = 0$, then we take the limits on K, and K' in (16), and it renders that $\limsup_{K\to+\infty}\sum_{i\in V}\|\xi_i(K) - \xi_i\|^2 \le \liminf_{K'\to+\infty}\sum_{i\in V}\|\xi_i(K') - \xi_i\|^2$. Therefore, we have $\lim_{k\to+\infty}\sum_{i\in V}\|\xi_i(k) - \xi_i\|^2$ exists.

By using this property and taking the limit on both sides of (14), we then have $\lim_{k\to+\infty}\|e_i(k)\| = 0$. By using Proposition 6.1 in the Appendix, we conclude that the consensus on λ and w is asymptotically achieved; i.e., $\lim_{k\to+\infty}\|\lambda^i(k) - \lambda^j(k)\| = 0$ and $\lim_{k\to+\infty}\|w^i(k) - w^j(k)\| = 0$ for any $i, j \in V$. Combining these with the convergence of $\{\sum_{i\in V}\|\xi_i(k) - \xi_i\|^2\}$ and the closedness of M_i, we can deduce that there exist a feasible dual pair $\tilde{\xi} := ((\tilde{\mu}_i), \tilde{\lambda}, \tilde{w})$ such that $\lim_{k\to+\infty}\|\mu_i(k) - \tilde{\mu}_i\| = 0$, $\lim_{k\to+\infty}\|\lambda^i(k) - \tilde{\lambda}\| = 0$, and $\lim_{k\to+\infty}\|w^i(k) - \tilde{w}\| = 0$, for all $i \in V$. Furthermore, we have $Q(\tilde{\xi}) \le d^*_\Delta$.

Substitute the approximate supgradient inequality (13) into (14), rearrange terms, and we have

$$
\begin{aligned}
&2\alpha(k)\sum_{i\in V}(Q_i(\xi_i) - Q_i(v_i(k)) - \epsilon) \le \sum_{i\in V}\alpha(k)^2\|\mathcal{D}_i(k)\|^2 \\
&+ \sum_{i\in V}(\|\xi_i(k) - \xi_i\|^2 - \|\xi_i(k+1) - \xi_i\|^2). \qquad (17)
\end{aligned}
$$

Let $\hat{\lambda}(k) := \frac{1}{N}\sum_{i\in V}\lambda^i(k)$ and $\hat{w}(k) := \frac{1}{N}\sum_{i\in V} w^i(k)$. By Lipschitz continuity of Q_i, it follows from (17) that

$$\begin{aligned}&\sum_{i\in V} 2\alpha(k)(Q_i(\xi_i) - Q_i(\mu_i(k), \hat{\lambda}(k), \hat{w}(k)) - \epsilon)\\&\le \sum_{i\in V}\alpha(k)^2\|\mathcal{D}_i(k)\|^2\\&+\sum_{i\in V}(\|\xi_i(k)-\xi_i\|^2 - \|\xi_i(k+1)-\xi_i\|^2)\\&+\sum_{i\in V}2\alpha(k)L(\|v^i_\lambda(k)-\hat{\lambda}(k)\| + \|v^i_w(k)-\hat{w}(k)\|).\end{aligned} \quad (18)$$

Now we follow a contradiction argument, and state $\tilde{\xi}$ is not approximate dual optimal. That is, assume that $\sum_{i\in V} Q_i(\tilde{\mu}_i, \tilde{\lambda}, \tilde{w}) < d^*_\Delta - N\epsilon$. Then $\rho := -\sum_{i\in V} Q_i(\tilde{\mu}_i, \tilde{\lambda}, \tilde{w}) + d^*_\Delta - N\epsilon > 0$. Let ξ_i in (18) be some dual optimal solution. Since $\lim_{k\to+\infty}\|v^i_\lambda(k) - \hat{\lambda}(k)\| = 0$ and $\lim_{k\to+\infty}\|v^i_w(k) - \hat{w}(k)\| = 0$, there is $K' \ge 0$ such that for all $k \ge K'$, there holds

$$\begin{aligned}\frac{1}{2}\rho\alpha(k) &\le \sum_{i\in V}\alpha(k)^2\|\mathcal{D}_i(k)\|^2\\&+\sum_{i\in V}(\|\xi_i(k)-\xi_i\|^2 - \|\xi_i(k+1)-\xi_i\|^2)\end{aligned} \quad (19)$$

Sum (19) over $[K', K]$ and rearrange it. It gives that

$$\begin{aligned}\sum_{i\in V}\|\xi_i(K+1)-\xi_i\|^2 &\le \sum_{k=K'}^{K}\sum_{i\in V}\alpha(k)^2\|\mathcal{D}_i(k)\|^2\\&-\frac{1}{2}\rho\sum_{k=K'}^{K}\alpha(k) + \sum_{i\in V}\|\xi_i(K')-\xi_i\|^2\end{aligned}$$

Since $\{\xi_i(k)\}$ converges, it is uniformly bounded. Recall that $\{\alpha(k)\}$ is not summable but square summable. When K is sufficiently large, the above inequality leads to a contradiction. Hence, it must be that $d^*_\Delta - N\epsilon \le Q(\tilde{\xi})$. ■

The remainder of this section is dedicated to characterizing the convergence properties of primal estimates. Toward this end, we present the closedness and upper semicontinuity properties of Ω^ϵ_i.

Lemma 4.5 (Properties of Ω^ϵ_i): The approximate set-valued marginal map Ω^ϵ_i is closed. In addition, it is upper semicontinuous at $\xi_i \in \Xi_i$; i.e., for any $\epsilon' > 0$, there is $\delta > 0$ such that for any $\tilde{\xi}_i \in B_{\Xi_i}(\xi_i, \delta)$, it holds that $\Omega^\epsilon_i(\tilde{\xi}_i) \subset B_{2^X}(\Omega^\epsilon_i(\xi_i), \epsilon')$.

Proof: Consider sequences $\{x_i(k)\}$ and $\{\xi_i(k)\}$ satisfying $\lim_{k\to+\infty}\xi_i(k) = \bar{\xi}_i$, $x_i(k) \in \Omega^\epsilon_i(\xi_i(k))$ and $\lim_{k\to+\infty} x_i(k) = \bar{x}_i$. Since $\mathcal{L}_i$ is continuous, then we have

$$\begin{aligned}\mathcal{L}_i(\bar{x}_i, \bar{\xi}_i) &= \lim_{k\to+\infty}\mathcal{L}_i(x_i(k), \xi_i(k))\\&\le \lim_{k\to+\infty}(Q_i(\xi_i(k)) + \epsilon) = Q_i(\bar{\xi}_i) + \epsilon,\end{aligned}$$

where in the inequality we use the property of $x_i(k) \in \Omega^\epsilon_i(\xi_i(k))$, and in the last equality we use the continuity of Q_i. Then $\bar{x}_i \in \Omega^\epsilon_i(\bar{\xi}_i)$ and the closedness of Ω^ϵ_i follows.

Note that $\Omega^\epsilon_i(\xi_i) = \Omega^\epsilon_i(\xi_i) \cap X$. Recall that Ω^ϵ_i is closed and X is compact. Then it is a result of Theorem 6.1 in the Appendix that $\Omega^\epsilon_i(\xi_i)$ is upper semicontinuous at $\xi_i \in \Xi_i$; i.e, for any neighborhood $\mathcal{U}$ in 2^X of $\Omega^\epsilon_i(\xi_i)$, there is $\delta > 0$ such that $\forall\tilde{\xi}_i \in B_{\Xi_i}(\xi_i, \delta)$, it holds that $\Omega^\epsilon_i(\tilde{\xi}_i) \subset \mathcal{U}$. Let $\mathcal{U} = B_{2^X}(\Omega^\epsilon_i(\xi_i), \epsilon')$, and we obtain the property of upper semicontinuity at ξ_i. ■

Upper semicontinuity of Ω^ϵ_i ensures that each accumulation point of $\{x_i(k)\}$ is a point in the set $\Omega^\epsilon_i(\tilde{\xi}_i)$; i.e., the convergence of $\{x_i(k)\}$ to the set $\Omega^\epsilon_i(\tilde{\xi}_i)$ can be guaranteed. In what follows, we further characterize the convergence of $\{x_i(k)\}$ to a point in $\Omega^\epsilon_i(\tilde{\xi}_i)$ within a finite time.

Lemma 4.6 (Primal estimate convergence): For each $i \in V$, there are a finite time $T_i \ge 0$ and $\tilde{x}_i \in \Omega^\epsilon_i(\tilde{\xi}_i)$ such that $x_i(k) = \tilde{x}_i$ for all $k \ge T_i + 1$.

Proof: Choose $\bar{\epsilon} > 0$ and $\hat{\epsilon} > 0$ such that $2(G + 4H + 2\sqrt{m}\delta)\bar{\epsilon} + 2\hat{\epsilon} \le \epsilon$. Since Q_i is continuous and $\lim_{k\to+\infty}\|v_i(k) - \tilde{\xi}_i\| = 0$, then there is $K_i \ge 0$ such that for all $k \ge K_i$, it holds that

$$\|\tilde{\xi}_i - v_i(k)\| \le \bar{\epsilon}, \quad \|Q_i(\tilde{\xi}_i) - Q_i(v_i(k))\| \le \hat{\epsilon}. \quad (20)$$

The time instant $T_i \ge 0$ is defined as follows: if there is some finite time $k \ge K_i + 1$ such that $\mathcal{L}_i(x_i(k), v_i(k+1)) > Q_i(v_i(k+1)) + \epsilon$, then T_i is the smallest one among such k; otherwise, $T_i = K_i + 1$. In what follows we prove that T_i is the time in the statement of the lemma.

Consider the first case of T_i. In this case, $\mathcal{L}_i(x_i(T_i), v_i(T_i + 1)) > Q_i(v_i(T_i + 1)) + \epsilon$; i.e., $x_i(T_i) \notin \Omega^\epsilon_i(v_i(T_i + 1))$. Then $x_i(T_i + 1) \in \Omega_i(v_i(T_i + 1))$; i.e., $\mathcal{L}_i(x_i(T_i + 1), v_i(T_i + 1)) = Q_i(v_i(T_i + 1))$. By using this property, we have that for any $k \ge T_i + 1$, it holds that

$$\begin{aligned}&\|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - Q_i(\tilde{\xi}_i)\|\\&\le \|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - Q_i(v_i(T_i+1))\|\\&+\|Q_i(v_i(T_i+1)) - Q_i(\tilde{\xi}_i)\|\\&= \|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - \mathcal{L}_i(x_i(T_i+1), v_i(T_i+1))\|\\&+\|Q_i(v_i(T_i+1)) - Q_i(\tilde{\xi}_i)\|.\end{aligned} \quad (21)$$

Notice that the term $\|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - \mathcal{L}_i(x_i(T_i+1), v_i(T_i+1))\|$ can be upper bounded in the following way:

$$\begin{aligned}&\|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - \mathcal{L}_i(x_i(T_i+1), v_i(T_i+1))\|\\&\le \|\langle\mu_i(k) - \mu_i(T_i+1), g(x_i(T_i+1))\rangle\\&+\langle -v^i_\lambda(k)_i + v^i_\lambda(k)_{i_U} + v^i_\lambda(T_i+1)_i - v^i_\lambda(T_i+1)_{i_U},\\&x_i(T_i+1)\rangle + \langle v^i_w(k)_i - v^i_w(k)_{i_U}\\&-v^i_w(T_i+1)_i + v^i_w(T_i+1)_{i_U}, x_i(T_i+1)\rangle\\&-\langle v^i_\lambda(k)_i - v^i_\lambda(T_i+1)_i, \Delta\rangle - \langle v^i_w(k)_i - v^i_w(T_i+1)_i, \Delta\rangle\|\\&\le 2(G + 4H + 2\sqrt{m}\delta)\bar{\epsilon}.\end{aligned} \quad (22)$$

Substituting (20) and (22) into (21) gives that

$$\begin{aligned}&\|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - Q_i(\tilde{\xi}_i)\|\\&\le 2(G + 4H + 2\sqrt{m}\delta)\bar{\epsilon} + \hat{\epsilon}.\end{aligned} \quad (23)$$

This implies that for any $k \ge T_i + 1$, it holds that

$$\begin{aligned}0 &\le \mathcal{L}_i(x_i(T_i+1), v_i(k)) - Q_i(v_i(k))\\&\le \|\mathcal{L}_i(x_i(T_i+1), v_i(k)) - Q_i(\tilde{\xi}_i)\|\\&+\|Q_i(\tilde{\xi}_i) - Q_i(v_i(k))\|\\&\le 2(G + 4H + 2\sqrt{m}\delta)\bar{\epsilon} + 2\hat{\epsilon} \le \epsilon.\end{aligned}$$

Hence, we conclude that $x_i(T_i+1) \in \Omega_i^\epsilon(v_i(k))$ for all $k \geq T_i+1$, and thus $x_i(k) = x_i(T_i+1)$ for all $k \geq T_i+1$.

We now consider the second possibility for T_i. In this case, $\mathcal{L}_i(x_i(k), v_i(k+1)) \leq Q_i(v_i(k+1)) + \epsilon$ for all $k \geq T_i = K_i+1$. Therefore, we have $x_i(T_i+1) \in \Omega_i^\epsilon(v_i(k))$ and then $x_i(k) = x_i(T_i+1)$ for all $k \geq T_i+1$.

In both cases, the chosen finite $T_i \geq 0$ guarantees that for all $k \geq T_i+1$, $x_i(k) = x_i(T_i+1)$ and $x_i(k) \in \Omega_i^\epsilon(v_i(T_i+1))$. Upper semicontinuity of Ω_i^ϵ ensures $x_i(T_i+1) \in \Omega_i^\epsilon(\tilde{\xi}_i)$. ■

Now we are ready to show the main result of this paper, Theorem 3.1. In particular, we will show the property of complementary slackness, primal feasibility of $\tilde{x}$, and characterize its primal suboptimality.

Proof for Theorem 3.1:

Claim 1: $\langle -\Delta - \tilde{x}_i + \tilde{x}_{i_D}, \tilde{\lambda}_i\rangle = 0$, $\langle -\Delta + \tilde{x}_i - \tilde{x}_{i_D}, \tilde{w}_i\rangle = 0$ and $\langle g(\tilde{x}_i), \tilde{\mu}_i\rangle = 0$.

Proof: Rearranging the terms related to λ in (14) leads to the following inequality holding for any $((\mu_i), \lambda, w) \in \Xi$ with $(\mu_i, \lambda, w) \in M_i$ for all $i \in V$:

$$\begin{aligned}
&-\sum_{i\in V} 2\alpha(k)(\langle -\Delta - x_i(k), v_\lambda^i(k)_i - \lambda_i\rangle \\
&+ \langle x_{i_D}(k), v_\lambda^{i_D}(k)_i - \lambda_i\rangle) \leq \alpha(k)^2 \sum_{i\in V} \|\mathcal{D}_i(k)\|^2 \\
&+ \sum_{i\in V}(\|\xi_i(k) - \xi_i\|^2 - \|\xi_i(k+1) - \xi_i\|^2) \\
&+ 2\alpha(k)\sum_{i\in V}\{\langle -x_i(k), v_w^i(k)_{i_U} - w_{i_U}\rangle + \langle x_i(k) - \Delta, \\
&v_w^i(k)_i - w_i\rangle + \langle g(x_i(k)), \mu_i(k) - \mu_i\rangle\}. \qquad (24)
\end{aligned}$$

Sum (24) over $[0, K]$, divide by $s(K) := \sum_{k=0}^K \alpha(k)$, and we have

$$\begin{aligned}
&\frac{1}{s(K)}\sum_{k=0}^K \alpha(k)\sum_{i\in V} 2(\langle \Delta + x_i(k), v_\lambda^i(k)_i - \lambda_i\rangle \\
&+ \langle -x_{i_D}(k), v_\lambda^{i_D}(k)_i - \lambda_i\rangle) \leq \frac{1}{s(K)}\sum_{k=0}^K \alpha(k)^2 \sum_{i\in V}\|\mathcal{D}_i(k)\|^2 \\
&+ \frac{1}{s(K)}\{\sum_{i\in V}(\|\xi_i(0) - \xi_i\|^2 - \|\xi_i(K+1) - \xi_i\|^2) \\
&+ \sum_{k=0}^K 2\alpha(k)\sum_{i\in V}(\langle g(x_i(k)), \mu_i(k) - \mu_i\rangle + \langle x_i(k) - \Delta, \\
&v_w^i(k)_i - w_i\rangle + \langle -x_i(k), v_w^i(k)_{i_U} - w_{i_U}\rangle)\}. \qquad (25)
\end{aligned}$$

We now proceed to show $\langle -\Delta - \tilde{x}_i + \tilde{x}_{i_D}, \tilde{\lambda}_i\rangle \geq 0$ for each $i \in V$. Notice that we have shown that $\lim_{k\to+\infty} \|x_i(k) - \tilde{x}_i\| = 0$ for all $i \in V$, and it also holds that $\lim_{k\to+\infty}\|\xi_i(k) - \tilde{\xi}_i\| = 0$ for all $i \in V$. Let $\lambda_i = \frac{1}{2}\tilde{\lambda}_i$, $\lambda_j = \tilde{\lambda}_j$ for $j \neq i$ and $\mu_i = \tilde{\mu}_i$, $w = \tilde{w}$ in (25). Recall that $\{\alpha(k)\}$ is not summable but square summable, and $\{\mathcal{D}_i(k)\}$ is uniformly bounded. Take $K \to +\infty$, and then it follows from Lemma 6.2 in the Appendix that:

$$\langle \Delta + \tilde{x}_i - \tilde{x}_{i_D}, \tilde{\lambda}_i\rangle \leq 0. \qquad (26)$$

On the other hand, since $\tilde{\xi} \in D_\Delta^\epsilon$, we have $\|\tilde{\xi}\| \leq \gamma$ by (8). Then we could choose a sufficiently small $\delta' > 0$ and $\xi \in \Xi$ in (25) such that $\|\xi\| \leq \gamma + \theta$ where θ is given in the definition of M_i and ξ is given by: $\lambda_i = (1+\delta')\tilde{\lambda}_i$, $\lambda_j = \tilde{\lambda}_j$ for $j \neq i$, $w = \tilde{w}$, $\mu = \tilde{\mu}$. Following the same lines toward (26), it gives that $-\delta\langle \Delta + \tilde{x}_i - \tilde{x}_{i_D}, \tilde{\lambda}_i\rangle \leq 0$. Hence, it holds that $\langle -\Delta - \tilde{x}_i + \tilde{x}_{i_D}, \tilde{\lambda}_i\rangle = 0$. The rest of the proof is analogous and thus omitted. ■

Claim 2: $\tilde{x}$ is primal feasible to the problem (P_Δ).

Proof: We have known that $\tilde{x}_i \in X$. We proceed to show $-\Delta - \tilde{x}_i + \tilde{x}_{i_D} \leq 0$ by contradiction. Since $\|\tilde{\xi}\| \leq \gamma$, we could choose a sufficiently small $\delta' > 0$ and ξ with $\|\xi\| \leq \gamma+\theta$ in (25) as follows: if $(-\Delta - \tilde{x}_i + \tilde{x}_{i_D})_\ell > 0$, then $(\lambda_i)_\ell = (\tilde{\lambda}_i)_\ell + \delta'$; otherwise, $(\lambda_i)_\ell = (\tilde{\lambda}_i)_\ell$, and $w = \tilde{w}$, $\mu = \tilde{\mu}$. The rest of the proofs is analogous to Claim 1.

Similarly, one can show $g(\tilde{x}_i) \leq 0$ and $-\Delta + \tilde{x}_i - \tilde{x}_{i_D} \leq 0$ by applying analogous arguments. ■

Claim 3: It holds that $p_\Delta^* \leq \sum_{i\in V} f_i(\tilde{x}_i) \leq p_\Delta^* + N\epsilon$.

Proof: Since $\tilde{x}$ is primal feasible, then $\sum_{i\in V} f_i(\tilde{x}_i) \geq p_\Delta^*$. On the other hand, $\sum_{i\in V} f_i(\tilde{x}_i) = \sum_{i\in V}\mathcal{L}_i(\tilde{x}_i, \tilde{\xi}_i) \leq \sum_{i\in V} Q_i(\tilde{\xi}_i) + N\epsilon \leq p_\Delta^* + N\epsilon$. ■

V. Conclusion

We have studied a multi-agent optimization problem where the goal of agents is to minimize a sum of local objective functions in the presence of a global inequality constraint and a global state constraint set. Objective and constraint functions as well as constraint sets are not necessarily convex. We have presented the distributed approximate dual subgradient algorithm which allow agents to asymptotically converge to a pair of approximate primal-dual solutions provided that the Slater's condition and strong duality property are satisfied.

VI. Appendix

A. Nonexpansion property of projection operators

Lemma 6.1: [4] Let Z be a non-empty, closed and convex set in $\mathbb{R}^n$. For any $z \in \mathbb{R}^n$, the following holds for any $y \in Z$: $\|P_Z[z] - y\|^2 \leq \|z - y\|^2 - \|P_Z[z] - z\|^2$.

B. A property of weighted sequence

Lemma 6.2: [33] Consider the sequence $\{\delta(k)\}$ defined by $\delta(k) := \frac{\sum_{\tau=0}^{k-1}\alpha(\tau)\rho(\tau)}{\sum_{\tau=0}^{k-1}\alpha(\tau)}$, where $\rho(k) \in \mathbb{R}^n$, $\alpha(k) > 0$, and $\sum_{k=0}^{+\infty}\alpha(k) = +\infty$. If $\lim_{k\to+\infty}\rho(k) = \rho^*$, then $\lim_{k\to+\infty}\delta(k) = \rho^*$.

C. Background on set-valued maps

We let $\mathbb{X}$ and $\mathbb{Y}$ denote Hausdorff topological spaces. A set-valued map $\Omega : \mathbb{X} \to \mathbb{Y}$ is a map that associates with any $x \in \mathbb{X}$ a subset $\Omega(x)$ of $\mathbb{Y}$. The following definitions and theorem are adopted from [2].

Definition 6.1: The set-valued map Ω is closed at a point $x \in \mathbb{X}$ if $\{x(k)\} \subset \mathbb{X}$, $\lim_{k\to+\infty}\mathrm{dist}(x(k), x) = 0$, $y(k) \in \Omega(x(k))$, and $\lim_{k\to+\infty}\mathrm{dist}(y(k), y) = 0$ implies that $y \in \Omega(x)$.

Definition 6.2: The set-valued map Ω is called upper semicontinuous at $x \in \mathbb{X}$ if and only if any neighborhood

$\mathcal{U}$ of $\Omega(x)$, there is $\eta > 0$ such that $\forall x' \in B(x, \eta)$, it holds that $\Omega(x') \subset \mathcal{U}$.

Theorem 6.1: Let Ω and Π be two set-valued maps from $\mathbb{X}$ to $\mathbb{Y}$. Assume that Ω is closed, $\Pi(x)$ is compact and Π is upper semicontinuous at $x \in \mathbb{X}$. Then $\Omega \cap \Pi$ is upper semicontinuous at x.

D. Dynamic average consensus algorithms

The following is the vector version of the first-order dynamic average consensus algorithm proposed in [34]:

$$x^i(k+1) = \sum_{j=1}^{N} a^i_j(k) x^j(k) + \eta^i(k), \tag{27}$$

where $x^i(k), \eta^i(k) \in \mathbb{R}^n$. Denote $\Delta\eta_\ell(k) := \max_{i\in V} \eta^i_\ell(k) - \min_{i\in V} \eta^i_\ell(k)$ for $1 \leq \ell \leq n$.

Proposition 6.1: [34] Let the periodic strong connectivity assumption 2.3, the non-degeneracy assumption 2.1 and the balanced communication assumption 2.2 hold. Assume that $\lim_{k\to+\infty} \Delta\eta_\ell(k) = 0$ for all $1 \leq \ell \leq n$ and all $k \geq 0$. Then the implementation of Algorithm (27) achieves consensus, i.e., $\lim_{k\to+\infty} \|x^i(k) - x^j(k)\| = 0$ for all $i, j \in V$.

References

[1] G. Arslan, J. R. Marden, and J. S. Shamma. Autonomous vehicle-target assignment: A game theoretic formulation. *ASME Journal on Dynamic Systems, Measurement, and Control*, 129(5):584–596, 2007.

[2] J.P. Aubin and H. Frankowska. *Set-valued analysis*. Birkhäuser, 1990.

[3] D.P. Bertsekas. *Convex optimization theory*. Anthena Scietific, 2009.

[4] D.P. Bertsekas, A. Nedic, and A. Ozdaglar. *Convex analysis and optimization*. Anthena Scietific, 2003.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.

[6] F. Bullo, J. Cortés, and S. Martínez. *Distributed Control of Robotic Networks*. Applied Mathematics Series. Princeton University Press, 2009. Available at http://www.coordinationbook.info.

[7] M. Chiang, S.H. Low, A.R. Calderbank, and J.C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of IEEE*, 95(1):255–312, January 2007.

[8] J. Cortés, S. Martínez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, 2004.

[9] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, 2004.

[10] B. Gharesifard and J. Cortés. Distributed strategies for generating weight-balanced and doubly stochastic digraphs. *SIAM Journal on Control and Optimization*, October 2009. Submitted.

[11] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms*. Springer, 1996.

[12] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

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

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

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

[16] K.C. Kiwiel. Approximations in bundle methods and decompostion of convex programs. *Journal of Optimization Theory and Applications*, 84:529–548, 1995.

[17] T. Larsson, M. Patriksson, and A. Strömberg. Ergodic primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming*, 86:283–312, 1999.

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

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

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

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

[22] R. D. Nowak. Distributed em algorithms for density estimation and clustering in sensor networks. *IEEE Transactions on Signal Processing*, 51:2245–2253, 2003.

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

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

[25] M. G. Rabbat and R. D. Nowak. Decentralized source localization and tracking. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 921–924, May 2004.

[26] M. G. Rabbat and R. D. Nowak. Quantized incremental algorithms for distributed optimization. *IEEE Journal on Select Areas in Communications*, 23:798–808, 2005.

[27] A. Rantzer. Using game theory for distributed control engineering. In *Games 2008, 3rd World Congress of the Game Theory Society*, 2008.

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

[29] R. T. Rockafella. Augmented lagrangian and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

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

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

[32] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Symposium on Information Processing of Sensor Networks*, pages 63–70, Los Angeles, CA, April 2005.

[33] M. Zhu and S. Martínez. On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods. *IEEE Transactions on Automatic Control*, 2009. Provisionally accepted.

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