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 the objective, constraint functions, and state constraint sets to be 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], [27], sensor fusion [31] and game-theoretic control [1].

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] [30], 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], [29], 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 [28], 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 $i \in V := \{1, \ldots, N\}$. 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 $G(k) = (V, E(k), A(k))$, for $k \geq 0$. Here, $A(k) := [a_{ij}^k(k)] \in \mathbb{R}^{N \times N}$ is the adjacency matrix, where the scalar $a_{ij}^k(k) \geq 0$ is the weight assigned to the edge $(j, i)$, and $E(k) \subseteq V \times V \setminus \text{diag}(V)$ is the set of edges with non-zero weights. The set of in-neighbors of agent $i$ at time $k$ is denoted by $N_i(k) = \{j \in V \mid (j, i) \in E(k)\}$ and $j \neq i$. Similarly, we define the set of out-neighbors of agent $i$ at time $k$ as $N'^{\text{out}}_i(k) = \{j \in V \mid (i, j) \in E(k)\}$ 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 $a_i^k(k) \geq \alpha$, and $a_i^j(k)$, for $i \neq j$, satisfies $a_i^j(k) \in \{0\} \cup [\alpha, 1]$, for all $k \geq 0$.

Assumption 2.2 (Balanced Communication): It holds that $\sum_{i \in V} a_i^j(k) = 1$ for all $i \in V$ and $k \geq 0$, and $\sum_{i \in V} a_i^j(k) = 1$ 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 $k \geq 0$, the directed graph $(V, \bigcup_{k=0}^{B-1} E(k_0 + k))$ is strongly connected.

The above network model is in 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 $(P)$:

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

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 $g : \mathbb{R}^n \to \mathbb{R}^m$ is continuous, and the inequality $g(z) \leq 0$ is understood component-wise; i.e., $g_j(z) \leq 0$, for all $j \in \{1, \ldots, m\}$, 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. This can be easily implemented as part of an initialization step, and thus the assumption that the Slater vector is known to all agents does not limit the applicability of our algorithm.

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:

\[
z_i(k + 1) = \max_{j \in X_\{i\}(k)} z_j(k), \tag{2}
\]

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 consensus; i.e., $z_i(k) = \max_{j \in V} z_j(k)$ for all $k \geq (N - 1)B$.

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. subgradients) 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, motivated by the well-known fact that dual problems are always convex, we will employ Lagrangian dualization to circumvent the challenges caused by non-convexity.

We first construct a directed cyclic graph $G_{cyc} := (V, E_{cyc})$ where $|E_{cyc}| = N$. 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 $G_{cyc}$, we will study the following approximate problem of problem $(P)$:

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

where $\Delta := \delta 1$, with $\delta$ a small positive scalar, and $1$ 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 $(P_\Delta)$. Its optimal value and the set of optimal solutions will be denoted by $p^*_\Delta$ and $X^*_\Delta$, respectively. Similarly to the problem $(P)$, $p^*_\Delta$ is finite and $X^*_\Delta \neq \emptyset$.

Remark 2.2: The cyclic graph $G_{cyc}$ 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 $-x_i + x_{i_D} - \Delta \leq 0$, for each out-neighbor $j$. For notational simplicity, we will use the cyclic graph $G_{cyc}$, which has a minimum number of constraints, as the initial graph.

A. Dual problems

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

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

where $\mu := (\mu_i) \in \mathbb{R}^{mN}_\geq$, $\lambda := (\lambda_i) \in \mathbb{R}^{nN}_\geq$, and $w := (w_i) \in \mathbb{R}^{nN}_\leq$. 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

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

We denote the dual optimal value of the problem $(D_{\Delta})$ by $p^*_\Delta$ and the set of dual optimal solutions by $D^*_\Delta$. In what follows we will assume that the duality gap is zero.
Assumption 2.5 (Strong duality): For the introduced problems \((P_{\Delta})\) and \((D_{\Delta})\), it holds that \(p^*_{\Delta} = d^*_{\Delta}\).

We endow each agent \(i\) with the local Lagrangian function \(L_i : \mathbb{R}^n \times \Xi_i \to \mathbb{R}\) and the local dual function \(Q_i : \Xi_i \to \mathbb{R}\) defined by

\[
L_i(x_i, \xi_i) := f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle -\lambda_i + \lambda_{iu}, x_i \rangle + \langle w_i - w_{iu}, x_i \rangle - \langle \lambda_i, \Delta \rangle - \langle w_i, \Delta \rangle,
\]

\[
Q_i(\xi_i) := \inf_{x_i \in X} L_i(x_i, \xi_i).
\]

In the problem \((P_{\Delta})\), the introduction of approximate consensus constraints \(-\Delta \leq x_i - x_{i_{\Delta}} \leq \Delta, i \in V\), 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:

\[
Q(\xi) = \inf_{x \in X^N} \sum_{i \in V} \left( f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle -\lambda_i + \lambda_{iu}, x_i \rangle + \langle w_i - w_{iu}, x_i \rangle - \langle \lambda_i, \Delta \rangle - \langle w_i, \Delta \rangle \right)
\]

\[
= \sum_{i \in V} \inf_{x_i \in X} \left( f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle -\lambda_i + \lambda_{iu}, x_i \rangle + \langle w_i - w_{iu}, x_i \rangle - \langle \lambda_i, \Delta \rangle - \langle w_i, \Delta \rangle \right)
\]

\[
= \sum_{i \in V} Q_i(\xi_i).
\]

It is worth mentioning that \(\sum_{i \in V} Q_i(\xi_i)\) is not separable since \(Q_i\) depends upon neighbor’s multipliers \(\lambda_{iu}\) and \(w_{iu}\).

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 \(\hat{Q}_i : \mathbb{R}^m_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}\) as follows:

\[
\hat{Q}_i(\mu_i, \lambda_i, w_i) = \inf_{x_i \in X, x_{i_{\Delta}} \in X} \left( f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle -\lambda_i + \lambda_{iu}, x_i \rangle + \langle w_i - w_{iu}, x_i \rangle - \langle \lambda_i, \Delta \rangle - \langle w_i, \Delta \rangle \right).
\]

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 \((P_{\Delta})\). Similarly to (3) and (4) in [33], which make use of Lemma 3.2 in the same paper, we have that for any \(\mu_i, \lambda_i, w_i \geq 0\) it holds that

\[
\max_{\xi \in \Omega_{\Delta}^i} ||\xi|| \leq N \max_{i \in V} \frac{f_i(\bar{z}) - \hat{Q}_i(\mu_i, \lambda_i, w_i)}{\beta(\bar{z})},
\]

where \(\beta(\bar{z}) := \min\{\min_{i \in \{1, \ldots, m\}} -g_i(\bar{z}), \delta\}\). Let \(\mu_i, \lambda_i\) and \(w_i\) be zero in (6), and it leads to the following upper bound on \(D_{\Delta}\):

\[
\max_{\xi \in D_{\Delta}} ||\xi|| \leq N \max_{i \in V} \frac{f_i(\bar{z}) - \hat{Q}_i(0, 0, 0)}{\beta(\bar{z})},
\]

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 \(D_{\Delta}\) is also bounded, then we have that \(D_{\Delta} \neq \emptyset\).

Remark 2.3: From the above analysis of \(D_{\Delta}\), it can be seen that if \(\delta = 0\), which corresponds to the exact consensus, then \(D_{\Delta}\) could be unbounded and empty.

Denote by \(D_{\Delta}' := \{\xi \in \Xi | Q(\xi) \geq d_{\Delta}^* - N\xi\}\). Similar to (7), we have the following upper bound on \(D_{\Delta}'\):

\[
\max_{\xi \in D_{\Delta}'} ||\xi|| \leq N \max_{i \in V} f_i(\bar{z}) - \hat{Q}_i(0, 0, 0) + \epsilon \beta(\bar{z}).
\]

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

C. Other notation

Define the set-valued map \(\Omega_i : \Xi_i \to 2^X\) in the following way \(\Omega_i(\xi_i) := \arg\min_{x_i \in X} L_i(x_i, \xi_i)\); i.e., given \(\xi_i\), the set \(\Omega_i(\xi_i)\) is the collection of solutions to the following local optimization problem:

\[
\min_{x_i \in X} L_i(x_i, \xi_i).
\]

Here, \(\Omega_i\) is referred to as the marginal map of agent \(i\). Since \(X\) is compact and \(f_i, g\) are continuous, then \(\Omega_i(\xi_i) \neq \emptyset\) 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 (the extremum candidates are the critical points of the objective function and isolated corners of the boundaries of the constraint regions) or have some specific structure allowing the use of global optimization methods such as branch and bound algorithms. For some \(\epsilon > 0\), we define the set-value map \(\Omega^\epsilon_i : \Xi_i \to 2^X\) as follows:

\[
\Omega^\epsilon_i(\xi_i) := \{x_i \in X | L_i(x_i, \xi_i) \leq Q_i(\xi_i) + \epsilon\},
\]

which we call 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 \(\text{dist}(z, A) := \inf_{y \in A} ||z - y||\), and the Hausdorff distance between two sets \(A, B \subset \mathbb{R}^n\) as \(\text{dist}(A, B) := \max\{\sup_{z \in A} \text{dist}(z, B), \sup_{y \in B} \text{dist}(A, y)\}\). We denote by 1 and 2 norms as \(\text{dist}(U, A, r) := \{u \in U | \text{dist}(u, A) \leq r\}\) and \(\text{dist}(U, A, r) := \{U \in 2^U | \text{dist}(U, A) \leq r\}\) where \(U \subset \mathbb{R}^n\).

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 \((P_{\Delta})\). 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 \((P_{\Delta})\) at time \(k \geq 0\), \(\mu_i(k) \in \mathbb{R}^+\) be the estimate of the multiplier on the inequality constraint \(g(x_i) \leq 0\), \(\lambda^i(k) \in \mathbb{R}^n_+\) (resp. \(w^i(k) \in \mathbb{R}^n_+\)) be the estimate of the multiplier associated with the

1We will use the superscript \(i\) to indicate that \(\lambda^i(k)\) and \(w^i(k)\) are estimates of some global variables.
collection of the local inequality constraints \(-x_j + x_{jD} - \Delta \leq 0\) (resp. \(x_j - x_{jD} - \Delta \leq 0\)), for all \(j \in V\).

We let \(\xi_i(k) := (\mu_i(k)^T, \lambda_i^T(k)^T)^T\), for \(i \in V\), and \(v_i(k) := (\mu_i(k)^T, \lambda_i^T(k)^T, x_i(k)^T)^T\) where \(\lambda_i^T(k)^T := \sum_{j \in V} a_j^i(k) x_j^T(k)\) and \(v_i(k) := \sum_{j \in V} a_j^i(k) w_j^i(k)\).

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

Initially, each agent \(i\) chooses a common Slater vector \(\tilde{z}\) and computes \(\gamma := \max_{i \in V} \gamma_i(\tilde{z})\) 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:

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 \(Q_i(v_i(k))\).

2. Produce the primal estimate \(x_i(k)\) as follows: if \(x_i(k-1) \in \Omega_i(v_i(k))\), 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:

\[
\xi_i(k+1) = P_{M_i}[v_i(k) + \alpha(k)D_i(k)],
\]

where the scalar \(\alpha(k)\) is a step-size. The supgradient vector of agent \(i\) is defined as \(D_i(k) := (D_{\mu i}(k)^T, D_{\lambda i}(k)^T, D_{w i}(k)^T)^T\), where \(D_{\mu i}(k) := g(x_i(k)) \in \mathbb{R}^n\), \(D_{\lambda i}(k)\) has components \(D_{\lambda i}(k)_{ji} := -\Delta - x_i(k) \in \mathbb{R}^n\), \(D_{\lambda i}(k)_{j} := x_i(k) \in \mathbb{R}^n\), and \(D_{\lambda i}(k)_{j} = 0 \in \mathbb{R}^n\) for \(j \in V \setminus \{i, i_d\}\), while the components of \(D_{w i}(k)\) are given by:

\[
D_{\mu i}(k)_{j} := -\Delta + x_i(k) \in \mathbb{R}^n, \quad D_{\lambda i}(k)_{j} := -x_i(k) \in \mathbb{R}^n, \quad D_{w i}(k)_{j} := 0 \in \mathbb{R}^n, \quad j \in V \setminus \{i, i_d\}.
\]

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 \(\gamma\) is an upper bound on \(D_{\lambda i}(k)\). Note that in Step 1, the check \(x_i(k-1) \in \Omega_i(v_i(k))\) reduces to verifying that \(L_i(x_i(k-1), v_i(k)) \leq Q_i(v_i(k)) + \epsilon\). Then, only if \(L_i(x_i(k-1), v_i(k)) > Q_i(v_i(k)) + \epsilon\), 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 \((P_{\Delta})\). We formally state this in the following.

**Theorem 3.1:** Consider the problem \((P_{\Delta})\) 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 \(\{\mu_i(k)\}, \{\lambda_i^T(k)\}, \{w_i(k)\}\) and the primal sequence of \(\{x_i(k)\}\) of the distributed approximate dual subgradient algorithm with the step-sizes \(\{\alpha(k)\}\) satisfying

\[
\lim_{k \to +\infty} \alpha(k) = 0, \quad \sum_{k=0}^{\infty} \alpha(k) = +\infty, \quad \text{and} \quad \sum_{k=0}^{\infty} \alpha(k)^2 < +\infty.
\]

Then, there exists a feasible dual pair \(\hat{\xi} := (\hat{\mu}, \hat{\lambda}, \hat{w})\) such that

\[
\lim_{k \to +\infty} \|\mu_i(k) - \hat{\mu}\| = 0, \quad \lim_{k \to +\infty} \|\lambda_i^T(k) - \hat{\lambda}\| = 0, \quad \text{and} \quad \lim_{k \to +\infty} \|w_i(k) - \hat{w}\| = 0.
\]

**Remark 3.2:** If \(g(x_i(k)) = 0\) and \(x_i(k) \in \Omega_i(v_i(k))\) for all \(k \geq 0\), then \(x_i(k)\) is a feasible primal vector. Theorem 3.1 is that the vector \(g(x_i(k)) T\), \(-\Delta + x_i(k) T\), \(x_i(k) T\), \(x_i(k) T\), \(-\lambda_i^T(k) T\) is an approximate dual subgradient of \(Q_i\) at \(v_i(k)\);
i.e., the following approximate supgradient inequality holds for any \( \xi_i \in \Xi_i \):

\[
Q_i(\xi_i) - Q_i(v_i(k)) \leq \langle g(x_i(k)), \mu_i - \mu_i(k) \rangle + \ldots \text{uniformly bounded.}
\]

Take \( K \to +\infty \), and then it follows from Lemma 6.1 in the Appendix that:

\[
\langle \Delta + \tilde{x}_i - \tilde{x}_i D, \tilde{\lambda}_i \rangle \leq 0. \tag{17}
\]

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 \( \xi_i, \tilde{\xi}_i \in \Xi_i \), it holds that

\[
\| Q_i(\xi_i) - Q_i(\tilde{\xi}_i) \| \leq L \| \xi_i - \tilde{\xi}_i \|. 
\]

In the DADS algorithm, the error induced by the projection map \( P_M \), is given by:

\[
e_i(k) := P_M[\nu_i(k) + \alpha(k)D_i(k)] - \nu_i(k).
\]

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 \( k \geq 0 \):

\[
\begin{align*}
\sum_{i \in V} \| e_i(k) - \alpha(k)D_i(k) \|^2 &\leq \alpha(k)^2 \sum_{i \in V} \| D_i(k) \|^2 \\
+ \sum_{i \in V} \| (\xi_i(k) - \xi_i) \|^2 - \| \xi_i(k+1) - \xi_i \|^2 \] + 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^k(k)i - \lambda_i \rangle \\
+ \langle x_i(k), v_i^k(k)i - \lambda_i \rangle \} + \langle x_i(k) - \Delta, w_i^k(k)i - w_i \rangle
\end{align*}
\]

(14)

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 \( \xi := ((\mu_i), \lambda, w) \) such that \( \lim_{k \to +\infty} \| \mu_i(k) - \mu_i \| = 0 \), \( \lim_{k \to +\infty} \| \lambda^*(k) - \lambda \| = 0 \), and \( \lim_{k \to +\infty} \| w^*(k) - w \| = 0 \).

Furthermore, the vector \( \xi \) is an approximate dual solution to the problem \( (D_\Delta) \) in the sense that \( d_{\Delta}^* - N\epsilon \leq Q(\xi) \leq d_{\Delta}^* \).

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 \( \Omega' \).

**Lemma 4.5 (Properties of \( \Omega' \)):** The approximate set-valued marginal map \( \Omega'_i \) 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 \( \Omega'_i(\tilde{\xi}_i) \subset B_{\Xi_i}(\Omega'_i(\xi_i), \epsilon') \).

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

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

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: \( \langle -\Delta - \tilde{x}_i + \tilde{x}_i D, \tilde{\lambda}_i \rangle = 0, \langle -\Delta + \tilde{x}_i - \tilde{x}_i D, \tilde{\mu}_i \rangle = 0 \) and \( \langle g(\tilde{x}_i), \mu_i \rangle = 0 \).

Proof: Rearranging the terms related to \( \lambda \) 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 \( i \in V \):

\[
\begin{align*}
- \sum_{i \in V} 2\alpha(k)(\langle -\Delta - x_i(k), v_i^k(k)i - \lambda_i \rangle \\
+ \langle x_i(k), v_i^k(k)i - \lambda_i \rangle)
\leq \alpha(k)^2 \sum_{i \in V} \| D_i(k) \|^2 \\
+ \sum_{i \in V} \| (\xi_i(k) - \xi_i) \|^2 - \| \xi_i(k+1) - \xi_i \|^2 \] + 2\alpha(k) \sum_{i \in V} \{ \langle x_i(k), v_i^k(k)i - w_i \rangle + \langle x_i(k) - \Delta, v_i^k(k)i - w_i \rangle
\end{align*}
\]

(15)

Sum (15) over \( [0, K] \), divide by \( s(K) := \sum_{k=0}^K \alpha(k) \), and we have

\[
\frac{1}{s(K)} \sum_{k=0}^K \alpha(k) \sum_{i \in V} \{ \langle g(x_i(k)), \mu_i(k) - \mu_i \rangle + \langle x_i(k) - \Delta, v_i^k(k)i - w_i \rangle
\]

(16)

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 \( \lim_{k \to +\infty} \| x_i(k) - \tilde{x}_i \| = 0 \) for all \( i \in V \). Let \( \lambda_i = \frac{1}{2} \tilde{\lambda}_i, \lambda_j = \frac{1}{2} \tilde{\lambda}_j \) for \( j \neq i \) and \( \mu_i = \tilde{\mu}_i, w = \tilde{w} \) in (16). Recall that \( \{ \alpha(k) \} \) is not summable but square summable, and \( \{ D_i(k) \} \) is uniformly bounded. Take \( K \to +\infty \), and then it follows from Lemma 6.1 in the Appendix that:

\[
\langle \Delta + \tilde{x}_i - \tilde{x}_i D, \tilde{\lambda}_i \rangle \leq 0. \tag{17}
\]
On the other hand, since \( \bar{\xi} \in D_{\epsilon} \cap \Delta \), we have \( \|\bar{\xi}\| \leq \gamma \) by (8). Then we could choose a sufficiently small \( \delta' > 0 \) and \( \xi \in \Xi \) via primal-dual subgradient methods. IEEE Transactions on Automatic Control, 2009. Provisionally accepted.

Claim 2: \( \hat{z} \) is primal feasible to the problem \( (P_{\Delta}) \).

\textit{Proof:} We have known that \( \hat{x}_i \in X \). We proceed to show \( -\Delta - \hat{x}_i + \hat{x}_{iD} \leq 0 \) by contradiction. Since \( \|\bar{\xi}\| \leq \gamma \), we could choose a sufficiently small \( \delta' > 0 \) and \( \xi \in \Xi \) such that \( \|\xi\| \leq \gamma + \delta' \) in (16) as follows: if \( (-\Delta - \hat{x}_i + \hat{x}_{iD}) \xi > 0 \), then \( (\lambda_i) \xi = (\lambda_i) + \delta' \); otherwise, \( (\lambda_i) \xi = (\lambda_i) \xi \), and \( w = \bar{w}, \mu = \bar{\mu} \). The rest of the proofs is analogous to Claim 1.

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

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

\textit{Proof:} Since \( \hat{x} \) is primal feasible, then \( \sum_{i \in V} f_i(\hat{x}_i) \geq \sum_{i \in V} Q_i(\hat{x}_i) + N \epsilon \leq p_{\Delta}^* + N \epsilon \).

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. Here, we have focused on presenting the distributed approximate dual subgradient algorithm and reporting its convergence properties under the assumption that the Slater's condition and strong duality property are satisfied. The extended version [32] will include a more detailed numerical study of a case study.

VI. APPENDIX

Lemma 6.1: [33] Consider the sequence \( \{\delta(k)\} \) of \( \delta(k) := \frac{\sum_{t=0}^{k} \alpha(t) \rho(t)}{\sum_{t=0}^{k} \alpha(t)} \) with \( \alpha(k) > 0 \) and \( \sum_{t=k}^{\infty} \alpha(k) = +\infty \). If \( \lim_{k \to +\infty} \rho(k) = \rho^* \), then \( \lim_{k \to +\infty} \delta(k) = \rho^* \).

REFERENCES
[1] G. Arslan, J. R. Marden, and J. S. Shamma. Autonomous vehicle-target assignment: A game theoretic formulation. ASME Journal on Dynamic Systems, Measurements, 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 Scitific, 2009.
[4] D.P. Bertsekas, A. Nedic, and A. Ozdaglar. Convex analysis and optimization. Anthena Scitific, 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] L. G. Hiriart-Urruty and C. Lemaréchal. Convex analysis and minimization algorithms. Springer, 1996.
[12] A. Jadabaei, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Transactions on Automatic Control, 48(6):1001–1011, 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 S. 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, 23(3):327–342, 1998.
[16] K.C. Kiwiel. Approximations in bundle methods and decomposition of convex programs. Journal of Optimization Theory and Applications, 84:529–548, 1995.
[17] T. Larsson, M. Patríciosson, and A. Strömberg. Ergodic primal convergence in dual subgradient schemes for convex programming. Mathematical Programming, 80:283–312, 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. Olshovsky 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] W. Ren and R. W. Beard. Distributed Consensus in Multi-vehicle Cooperative Control. Communications and Control Engineering, Springer, 2008.
[28] R. T. Rockafella. Augmented lagrangian and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research, 1:97–116, 1976.
[29] S. Sridharan and S. Mohanty. Distributed source localization. In ICASSP, 2008.
[30] R. Srikant. Mathematics of Internet Congestion Control. Systems & Control: Foundations & Applications. Birkhäuser, 2005.
[31] 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.
[32] M. Zhu and Martínez. An approximate dual subgradient algorithm for multi-agent non-convex optimization. International Journal of Robust and Nonlinear Control, pages 1–15, 2010.
[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.