

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 subject to local, intermittent interactions aim to minimize a sum of local objective functions subject to a global inequality constraint and a global state constraint set. In contrast to previous work, we do not require that the objective, constraint functions, and state constraint sets to be convex. In order to deal with time-varying network topologies satisfying a standard connectivity assumption, we resort to consensus algorithm techniques and the Lagrangian duality method. We slightly relax the requirement of exact consensus, and propose a distributed approximate dual subgradient algorithm to enable agents to asymptotically converge to a pair of primal-dual solutions to an approximate problem. To guarantee convergence, we assume that the Slater's condition is satisfied and the optimal solution set of the dual limit is singleton. We implement our algorithm over a source localization problem and compare the performance with existing algorithms.

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 [6], [15], [17], [19], [25], [26], coverage control [7], [10], formation control [11], [31] and sensor fusion [34].

M. Zhu is with Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge MA, 02139, (mhzhu@mit.edu). S. Martínez is with Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Dr, La Jolla CA, 92093, (soniamd@ucsd.edu).

The seminal work [3] provides a framework to tackle optimizing a global objective function among different processors where each processor knows the global objective function. In multi-agent environments, 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 [23] [33], distributed source localization [29], and network utility maximization [18]. More recently, consensus techniques have been proposed to address the issues of switching topologies, asynchronous computation and coupling in objective functions; see for instance [16], [21], [22], [30], [36]. More specifically, the paper [21] presents the first analysis of an algorithm that combines average consensus schemes with subgradient methods. Using projection in the algorithm of [21], the authors in [22] further address a more general scenario that takes local state constraint sets into account. Further, in [36] 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 [36]. In order to deal with all aspects of our multi-agent setting, our method integrates Lagrangian dualization, subgradient schemes, and average consensus algorithms. Distributed function computation by a group of anonymous agents interacting intermittently can be done via agreement algorithms [7]. However, agreement algorithms are essentially convex, and so we are led to the investigation of nonconvex optimization solutions via dualization. The techniques of dualization and subgradient schemes have been popular and efficient approaches to solve both convex programs (e.g., in [4], [5]) and nonconvex programs (e.g., in [8], [9]).

Statement of Contributions. Here, we investigate a multi-agent optimization problem where agents desire to agree upon a global decision vector minimizing the sum of local objective functions in the presence of a global inequality constraint and a global state constraint set. Agent interactions are changing with time. The objective, constraint functions, as well as the state-constraint set, can be nonconvex. To deal with both nonconvexity and time-varying interactions, we first define an approximated problem where the exact consensus is slightly relaxed. We then

propose a distributed dual subgradient algorithm to solve it, where the update rule for local dual estimates combines a dual subgradient scheme with average consensus algorithms, and local primal estimates are generated from local dual optimal solution sets. This algorithm is shown to asymptotically converge to a pair of primal-dual solutions to the approximate problem under the following assumptions: firstly, the Slater's condition is satisfied; secondly, the optimal solution set of the dual limit is singleton; thirdly, dynamically changing network topologies satisfy some standard connectivity condition.

A conference version of this manuscript was published in [35]. Main differences are the following: (i) by assuming that the optimal solution set of the dual limit is a singleton, and changing the update rule in the dual estimates, we are able to determine a global solution in contrast to an approximate solution in [35]; (ii) we present a simple criterion to check the new sufficient condition for nonconvex quadratic programming; (iii) we present new simulation results of our algorithm on a source localization example and compare its performance with existing algorithms.

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) pointing from agent j to agent 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 $\mathcal{N}_i(k) = \{j \in V \mid (j, i) \in E(k) \text{ and } j \neq i\}$. 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 network communication graphs:

Assumption 2.1 (Non-degeneracy): There exists a constant $\alpha > 0$ such that $a_j^i(k) \geq \alpha$, 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): ¹It 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 $k \geq 0$.

¹It is also referred to as double stochasticity of the adjacency matrix $A(k)$.

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 to characterize a networked multi-agent system, and has been widely used in the analysis of average consensus algorithms; e.g., see [25], [26], and distributed optimization in [22], [36]. Recently, an algorithm is given in [13] 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):

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

where $z \in \mathbb{R}^n$ is the global decision vector. The function $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are continuous, and the inequality $g(z) \leq 0$ is understood component-wise; i.e., $g_\ell(z) \leq 0$, for all $\ell \in \{1, \dots, m\}$, 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. Specifically, the maximum-consensus algorithm is described as follows:

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 rule of $z_i(k+1) = \max_{j \in \mathcal{N}_i(k) \cup \{i\}} z_j(k)$, where we use the following relation for vectors: for $a, b \in \mathbb{R}^n$, $a < b$ if and only if there is some $\ell \in \{1, \dots, n-1\}$ 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. •

In [36], in order to solve the convex case of the problem (P) (i.e.; f_i and g are convex functions and X is a convex set), we propose two distributed primal-dual subgradient algorithms where primal (resp. dual) estimates move along subgradients (resp. supergradients) and are projected onto convex sets. The absence of convexity impedes the use of the algorithms in [36] 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 nonconvex. In the sequel, we will employ Lagrangian dualization, subgradient methods and average consensus schemes to design a distributed algorithm which is able to find an approximate solution to the problem (P).

Towards this end, we 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}_{cyc} , we will study the following approximate problem of problem (P):

$$\begin{aligned} & \min_{(x_i) \in \mathbb{R}^{nN}} \sum_{i \in V} f_i(x_i), \\ \text{s.t. } & g(x_i) \leq 0, \quad -x_i + x_{i_D} - \Delta \leq 0, \quad x_i - x_{i_D} - \Delta \leq 0, \quad x_i \in X, \quad \forall i \in V, \end{aligned} \quad (2)$$

where $\Delta := \delta \mathbf{1}$, with δ a small positive scalar, and $\mathbf{1}$ is the column vector of n ones. The problem (2) provides an approximation of the problem (P), and will be referred to as problem (P_Δ). In particular, the approximate problem (2) reduces to the problem (P) when $\delta = 0$. 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}_{cyc} can be replaced by any strongly connected graph \mathcal{G} . Given \mathcal{G} , 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 . This set of inequalities implies that any feasible solution $x = (x_i)_{i \in V}$ of problem (P_Δ) satisfies the approximate consensus; i.e., $\max_{i,j \in V} \|x_i - x_j\| \leq N\delta$. For notational simplicity, we will use the cyclic graph \mathcal{G}_{cyc} , which has a minimum number of constraints, as the initial graph. •

A. Dual problems

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

$$\max_{\mu, \lambda, w} Q(\mu, \lambda, w), \quad \text{s.t.} \quad \mu, \lambda, w \geq 0, \quad (3)$$

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 $Q : \Xi \rightarrow \mathbb{R}$ is given as $Q(\xi) \equiv Q(\mu, \lambda, w) := \inf_{x \in X^N} \mathcal{L}(x, \mu, \lambda, w)$, where $\mathcal{L} : \mathbb{R}^{nN} \times \Xi \rightarrow \mathbb{R}$ 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_Δ) by d_Δ^* and the set of dual optimal solutions by D_Δ^* . We endow each agent i with the local Lagrangian function $\mathcal{L}_i : \mathbb{R}^n \times \Xi' \rightarrow \mathbb{R}$ and the local dual function $Q_i : \Xi' \rightarrow \mathbb{R}$ defined by

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

In the approximate problem (P_Δ) , the introduction of $-\Delta \leq x_i - x_{i_D} \leq \Delta$, $i \in V$, 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:

$$Q(\xi) = \inf_{x \in X^N} \sum_{i \in V} (f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle \lambda_i, -x_i + x_{i_D} - \Delta \rangle + \langle w_i, x_i - x_{i_D} - \Delta \rangle).$$

Notice that in the sum of $\sum_{i \in V} \langle \lambda_i, -x_i + x_{i_D} - \Delta \rangle$, each x_i for any $i \in V$ appears in two terms: one is $\langle \lambda_i, -x_i + x_{i_D} - \Delta \rangle$, and the other is $\langle \lambda_{i_U}, -x_{i_U} + x_i - \Delta \rangle$. With this observation, we regroup the terms in the summation in terms of x_i , and have the following:

$$\begin{aligned} Q(\xi) &= \inf_{x \in X^N} \sum_{i \in V} (f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle -\lambda_i + \lambda_{i_U}, x_i \rangle + \langle w_i - w_{i_U}, x_i \rangle - \langle \lambda_i, \Delta \rangle - \langle w_i, \Delta \rangle) \\ &= \sum_{i \in V} \inf_{x_i \in X} (f_i(x_i) + \langle \mu_i, g(x_i) \rangle + \langle -\lambda_i + \lambda_{i_U}, x_i \rangle + \langle w_i - w_{i_U}, x_i \rangle - \langle \lambda_i, \Delta \rangle - \langle w_i, \Delta \rangle) \\ &= \sum_{i \in V} Q_i(\xi_i). \end{aligned} \quad (4)$$

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., [14], [20]. We will shortly see that the Slater's condition plays the same role in nonconvex optimization. To achieve this, we define the function $\hat{Q}_i : \mathbb{R}_{\geq 0}^m \times \mathbb{R}_{\geq 0}^n \times \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$ as follows:

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

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 [36], 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

$$\max_{\xi \in D_\Delta^*} \|\xi\| \leq N \max_{i \in V} \frac{f_i(\bar{z}) - \hat{Q}_i(\mu_i, \lambda_i, w_i)}{\beta(\bar{z})}, \quad (5)$$

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 (5), and it leads to the following upper bound on D_Δ^* :

$$\max_{\xi \in D_\Delta^*} \|\xi\| \leq N \max_{i \in V} \frac{f_i(\bar{z}) - \hat{Q}_i(0, 0, 0)}{\beta(\bar{z})}, \quad (6)$$

where $\hat{Q}_i(0, 0, 0) = \inf_{x_i \in X} f_i(x_i)$ and it can be computed locally. We denote

$$\gamma_i(\bar{z}) := \frac{f_i(\bar{z}) - \hat{Q}_i(0, 0, 0)}{\beta(\bar{z})}. \quad (7)$$

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 $D_\Delta^* \neq \emptyset$.

Remark 2.3: The requirement of exact agreement on z in the problem P is slightly relaxed in the problem P_Δ by introducing a small positive scalar δ . In this way, on the one hand, the global dual function Q is a sum of the local dual functions Q_i , as in (4); on the other hand, D_Δ^* is non-empty and uniformly bounded. These two properties play important roles in the devise of our subsequent algorithm. •

C. Other notation

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

$$\min_{x_i \in X} \mathcal{L}_i(x_i, \xi_i). \quad (8)$$

Here, Ω_i is referred to as the *marginal map* of agent i . Since X is compact and f_i, g are continuous, then $\Omega_i(\xi_i) \neq \emptyset$ in (8) for any $\xi_i \in \Xi'$. In the algorithm we will develop in next section, each agent is required to obtain *one* (globally) optimal solution and the optimal value the local optimization problem (8) at each iterate. We assume that this can be easily solved, and this is the case for problems of $n = 1$, or f_i and g being smooth (the extremum candidates are the critical points of the objective function and isolated corners of the boundaries of the constraint regions) or having some specific structure which allows the use of global optimization methods such as branch and bound algorithms.

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 $B_{\mathcal{U}}(A, r) := \{u \in \mathcal{U} \mid \text{dist}(u, A) \leq r\}$ and $B_{2^{\mathcal{U}}}(A, r) := \{U \in 2^{\mathcal{U}} \mid \text{dist}(U, A) \leq r\}$ 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 primal-dual solutions to the approximate 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 approximate problem (P_{Δ}) at time $k \geq 0$, $\mu_i(k) \in \mathbb{R}_{\geq 0}^m$ be the estimate of the multiplier on the inequality constraint $g(x_i) \leq 0$, $\lambda^i(k) \in \mathbb{R}_{\geq 0}^{nN}$ (resp. $w^i(k) \in \mathbb{R}_{\geq 0}^{nN}$)² be the estimate of the multiplier associated with the 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 $\xi_i(k) := (\mu_i(k)^T, \lambda^i(k)^T, w^i(k)^T)^T \in \Xi'$, for $i \in V$ to be the collection of dual estimates of agent i . And denote $v_i(k) := (\mu_i(k)^T, v_{\lambda}^i(k)^T, v_w^i(k)^T)^T \in \Xi'$ where $v_{\lambda}^i(k) := \sum_{j \in V} a_j^i(k) \lambda^j(k) \in \mathbb{R}_{\geq 0}^{nN}$ and $v_w^i(k) := \sum_{j \in V} a_j^i(k) w^j(k) \in \mathbb{R}_{\geq 0}^{nN}$ are convex combinations of dual estimates of agent i and its neighbors at time k .

At time k , we associate each agent i a supergradient vector $\mathcal{D}_i(k)$ defined as $\mathcal{D}_i(k) := (\mathcal{D}_{\mu}^i(k)^T, \mathcal{D}_{\lambda}^i(k)^T, \mathcal{D}_w^i(k)^T)^T$, where $\mathcal{D}_{\mu}^i(k) := g(x_i(k)) \in \mathbb{R}^m$, $\mathcal{D}_{\lambda}^i(k)$ has components $\mathcal{D}_{\lambda}^i(k)_i := -\Delta - x_i(k) \in \mathbb{R}^n$, $\mathcal{D}_{\lambda}^i(k)_{i_U} := x_i(k) \in \mathbb{R}^n$, and $\mathcal{D}_{\lambda}^i(k)_j = 0 \in \mathbb{R}^n$ for $j \in V \setminus \{i, i_U\}$, while the components of $\mathcal{D}_w^i(k)$ are given by: $\mathcal{D}_w^i(k)_i := -\Delta + x_i(k) \in \mathbb{R}^n$, $\mathcal{D}_w^i(k)_{i_U} :=$

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

$-x_i(k) \in \mathbb{R}^n$, and $\mathcal{D}_w^i(k)_j = 0 \in \mathbb{R}^n$, for $j \in V \setminus \{i, i_U\}$. For each agent i , we define the set $M_i := \{\xi_i \in \Xi' \mid \|\xi_i\| \leq \gamma + \theta_i\}$ for some $\theta_i > 0$ where $\gamma := N \max_{i \in V} \gamma_i(\bar{z})$. Let P_{M_i} to be the projection onto the set M_i . It is easy to check that M_i is closed and convex, and thus the projection map P_{M_i} is well-defined.

The *Distributed Approximate Dual Subgradient* (DADS, for short) Algorithm is described in Table 1.

Algorithm 1 The Distributed Approximate Dual Subgradient Algorithm

Initialization: Initially, all the agents agree upon some $\delta > 0$ in the approximate problem (P_Δ) . Each agent i chooses a common Slater vector \bar{z} , computes $\gamma_i(\bar{z})$ and obtains $\gamma = N \max_{i \in V} \gamma_i(\bar{z})$ through a max-consensus algorithm where $\gamma_i(\bar{z})$ is given in (7). After that, each agent i chooses initial states $x_i(0) \in X$ and $\xi_i(0) \in \Xi'$.

Iteration: At each time k , each agent i executes the following steps:

- 1: For each $k \geq 1$, given $v_i(k)$, solve the local optimization problem (8), obtain a solution $x_i(k) \in \Omega_i(v_i(k))$ and the dual optimal value $Q_i(v_i(k))$.
- 2: For each $k \geq 0$, generate the dual estimate $\xi_i(k+1)$ according to the following rule:

$$\xi_i(k+1) = P_{M_i}[v_i(k) + \alpha(k)\mathcal{D}_i(k)], \quad (9)$$

where the scalar $\alpha(k) \geq 0$ is a step-size.

- 3: Repeat for $k = k + 1$.
-

Remark 3.1: The DADS algorithm is an extension of the classical dual algorithm, e.g., in [28] and [4] to the multi-agent setting and nonconvex case. In the initialization of the DADS algorithm, the value γ serves as an upper bound on D_Δ^* . In Step 1, *one solution* in $\Omega_i(v_i(k))$ is needed, and it is unnecessary to compute the whole set $\Omega_i(v_i(k))$. •

In order to assure the primal convergence, we will assume that the dual estimates converge to the set where each has a single optimal solution.

Definition 3.1 (Singleton optimal dual solution set): The set of $D_s^* \subseteq \mathbb{R}^{(m+2n)N}$ is the collection of ξ such that the set $\Omega_i(\xi_i)$ is a singleton, where $\xi_i = (\mu_i, \lambda, w)$ for each $i \in V$. •

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

state this in the following theorem:

Theorem 3.1 (Convergence properties of the DADS algorithm): Consider the problem (P) and the corresponding approximate problem (P_Δ) with some $\delta > 0$. We 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 holds for the problem (P). 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 $\{\alpha(k)\}$ satisfying $\lim_{k \rightarrow +\infty} \alpha(k) = 0$,

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

- 1) (Dual estimate convergence) There exists a dual solution $\xi^* \in D_\Delta^*$ where $\xi^* := (\mu^*, \lambda^*, w^*)$ and $\mu^* := (\mu_i^*)$ such that the following holds for all $i \in V$:

$$\lim_{k \rightarrow +\infty} \|\mu_i(k) - \mu_i^*\| = 0, \quad \lim_{k \rightarrow +\infty} \|\lambda^i(k) - \lambda^*\| = 0, \quad \lim_{k \rightarrow +\infty} \|w^i(k) - w^*\| = 0.$$

- 2) (Primal estimate convergence) If the dual solution verifies $\xi^* \in D_s^*$, i.e. $\Omega_i(\xi_i^*)$ is a singleton for all $i \in V$, then there is $x^* \in X_\Delta^*$ with $x^* := (x_i^*)$ such that the following holds for all $i \in V$:

$$\lim_{k \rightarrow +\infty} \|x_i(k) - x_i^*\| = 0.$$

IV. DISCUSSION

Before proceeding with the technical proofs for Theorem 3.1, we would like to make the following observations. First, our methodology is motivated by the need of solving a nonconvex problem in a distributed way by a group of agents whose interactions change with time. This places a number of restrictions on the type of solutions that one can find. Time-varying interactions of anonymous agents can be currently solved via agreement algorithms; however these are inherently convex operations, which does not work well in nonconvex settings. To overcome this, one can resort to dualization. Admittedly, zero duality gap does not hold in general for nonconvex problems. A possibility would be to resort to nonlinear augmented Lagrangians, for which strong duality holds in a broad class of programs [8], [9], [32]. However, we find here another problem, as a distributed solution using agreement requires separability, as the one ensured by the linear Lagrangians we use here. Thus, we have looked for alternative assumptions that can be easier to check and allow the dualization approach to work.

More precisely, Theorem 3.1 shows that dual estimates always converge to a dual optimal solution. The convergence of primal estimates requires an additional assumption that the dual limit has a single optimal solution. Let us refer to this assumption as *the singleton dual optimal solution set* (SD for short). This assumption may not be easy to check *a priori*, however it is of similar nature as existing algorithms for nonconvex optimization. In [8] and [9], subgradient methods are defined in terms of (nonlinear) augmented Lagrangians, and it is shown that every accumulation point of the primal sequence is a primal solution provided that the dual function is required to be differentiable at the dual limit. An open question is how to resolve the above issues imposed by the multi-agent setting with less stringent conditions on the nature of the nonconvex optimization problem.

In the following, we study a class of nonconvex quadratic programs for which a sufficient condition guarantees that the SD assumption holds. Nonconvex quadratic programs hold great importance from both theoretic and practical aspects. In general, nonconvex quadratic programs are NP-hard, and please refer to [27] for detailed discussion. The aforementioned sufficient condition only requires checking the positive definiteness of a matrix.

Consider the following nonconvex quadratic program:

$$\begin{aligned} \min_{z \in \bigcap_{i \in V} X_i} f(z) &= \sum_{i \in V} f_i(z) = \sum_{i \in V} (\|z\|_{P_i}^2 + 2\langle q_i, z \rangle), \\ \text{s.t.} \quad &\|z\|_{A_{i,\ell_i}}^2 + 2\langle b_{i,\ell_i}, z \rangle + c_{i,\ell_i} \leq 0, \quad \ell_i = 1, \dots, m_i, \end{aligned} \quad (10)$$

where $\|z\|_{A_{i,\ell_i}}^2 \triangleq z^T A_{i,\ell_i} z$ and A_{i,ℓ_i} are real and symmetric matrices. The approximate problem of P_Δ is given by

$$\begin{aligned} \min_{x \in \mathbb{R}^{2N}} \sum_{i \in V} f_i(x_i) &= \sum_{i \in V} (\|x_i\|_{P_i}^2 + 2\langle q_i, x_i \rangle), \\ \text{s.t.} \quad &\|x_i\|_{A_{i,\ell_i}}^2 + 2\langle b_{i,\ell_i}, x_i \rangle + c_{i,\ell_i} \leq 0, \quad \ell_i = 1, \dots, m_i, \\ &-x_i + x_{i_D} - \Delta \leq 0, \quad x_i - x_{i_D} - \Delta \leq 0, \quad x_i \in X_i, \quad i \in V. \end{aligned} \quad (11)$$

We introduce the dual multipliers (μ, λ, w) as before. The local Lagrangian function \mathcal{L}_i can be written as follows:

$$\mathcal{L}_i(x_i, \xi_i) \triangleq \|x_i\|_{P_i + \sum_{\ell_i=1}^{m_i} \mu_{i,\ell_i} A_{i,\ell_i}}^2 + \langle \zeta_i, x_i \rangle,$$

where the term independent of x_i is dropped and ζ_i is a linear function of $\xi_i = (\mu_i, \lambda, w)$. The dual function and dual problem can be defined as before. Consider any dual optimal solution ξ^* . If for all $i \in V$:

(P1) $P_i + \sum_{\ell_i=1}^{m_i} \mu_{i,\ell_i}^* A_{i,\ell_i}$ is positive definite;

(P2) $x_i^* = (P_i + \sum_{\ell_i=1}^{m_i} \mu_{i,\ell_i}^* A_{i,\ell_i})^{-1} \zeta_i^* \in X_i$;

then the SD assumption holds. The properties (P1) and (P2) are easy to verify in a distributed way once obtaining a dual solution ξ^* . We would like to remark that (P1) is used in [12] to determine the unique global optimal solution via canonical duality when the constraint set X is absent.

V. CONVERGENCE ANALYSIS

This section provides the complete analysis of Theorem 3.1. Recall that g is continuous and X is compact. Then there are $G, H > 0$ such that $\|g(x)\| \leq G$ and $\|x\| \leq H$ for all $x \in X$. We start our analysis from the computation of supergradients of Q_i .

Lemma 5.1 (Supergradient computation): 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 supergradient of Q_i at $\bar{\xi}_i$; i.e., the following holds for any $\xi_i \in \Xi'$:

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

Proof: The proof is based on the computation of dual subgradients, e.g., in [4], [5]. ■

A direct result of Lemma 5.1 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 a supergradient of Q_i at $v_i(k)$; i.e., the following supergradient inequality holds for any $\xi_i \in \Xi'$:

$$\begin{aligned} Q_i(\xi_i) - Q_i(v_i(k)) &\leq \langle g(x_i(k)), \mu_i - \mu_i(k) \rangle + \langle -\Delta - x_i(k), \lambda_i - v_\lambda^i(k)_i \rangle \\ &\quad + \langle x_i(k), \lambda_{i_U} - v_\lambda^i(k)_{i_U} \rangle + \langle x_i(k) - \Delta, w_i - v_w^i(k)_i \rangle + \langle -x_i(k), w_{i_U} - v_w^i(k)_{i_U} \rangle. \end{aligned} \quad (13)$$

Now we can see that the update rule (9) of dual estimates in the DADS algorithm is a combination of a dual subgradient scheme and average consensus algorithms. The following establishes that Q_i is Lipschitz continuous with some Lipschitz constant L .

Lemma 5.2 (Lipschitz continuity of Q_i): There is a constant $L > 0$ such that for any $\xi_i, \bar{\xi}_i \in \Xi'$, it holds that $\|Q_i(\xi_i) - Q_i(\bar{\xi}_i)\| \leq L\|\xi_i - \bar{\xi}_i\|$.

Proof: Similarly to Lemma 5.1, 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 supergradient of Q_i at $\bar{\xi}_i$; i.e., the following holds for any $\xi_i \in \Xi'$:

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

Since $\|g(\bar{x}_i)\| \leq G$ and $\|\bar{x}_i\| \leq H$, there is $L > 0$ such that $Q_i(\xi_i) - Q_i(\bar{\xi}_i) \leq L\|\xi_i - \bar{\xi}_i\|$. Similarly, $Q_i(\bar{\xi}_i) - Q_i(\xi_i) \leq L\|\xi_i - \bar{\xi}_i\|$. We then reach the desired result. \blacksquare

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 5.3 (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 $k \geq 0$:

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

Proof: Recall that M_i is closed and convex. The proof is a combination of the nonexpansion property of projection operators in [5] and the property of balanced graphs. \blacksquare

The lemma below shows the asymptotic convergence of dual estimates.

Lemma 5.4 (Dual estimate convergence): Under the assumptions in Theorem 3.1, there exists a dual optimal solution $\xi^* := ((\mu_i^*), \lambda^*, w^*) \in D_\Delta^*$ such that $\lim_{k \rightarrow +\infty} \|\mu_i(k) - \mu_i^*\| = 0$, $\lim_{k \rightarrow +\infty} \|\lambda^i(k) - \lambda^*\| = 0$, and $\lim_{k \rightarrow +\infty} \|w^i(k) - w^*\| = 0$.

Proof: By the dual decomposition property (4) and the boundedness of dual optimal solution sets, the dual problem (D_Δ) is equivalent to the following:

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

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 convex and compact. The rest of the proofs can be finished by following similar lines in [36], and thus omitted. ■

The remainder of this section is dedicated to characterizing the convergence properties of primal estimates. Toward this end, we present some properties of Ω_i .

Lemma 5.5 (Properties of marginal maps): The set-valued marginal map Ω_i is closed. In addition, it is upper semicontinuous at $\xi_i \in \Xi'$; i.e., for any $\epsilon' > 0$, there is $\delta' > 0$ such that for any $\tilde{\xi}_i \in B_{\Xi'}(\xi_i, \delta')$, it holds that $\Omega_i(\tilde{\xi}_i) \subset B_{2^X}(\Omega_i(\xi_i), \epsilon')$.

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

$$\mathcal{L}_i(\bar{x}_i, \bar{\xi}_i) = \lim_{k \rightarrow +\infty} \mathcal{L}_i(x_i(k), \xi_i(k)) \leq \lim_{k \rightarrow +\infty} (Q_i(\xi_i(k))) = Q_i(\bar{\xi}_i),$$

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

Note that $\Omega_i(\xi_i) = \Omega_i(\xi_i) \cap X$. Recall that Ω_i is closed and X is compact. Then it is a result of Proposition 1.4.9 in [2] that $\Omega_i(\xi_i)$ is upper semicontinuous at $\xi_i \in \Xi'$; i.e, for any neighborhood \mathcal{U} in 2^X of $\Omega_i(\xi_i)$, there is $\delta' > 0$ such that $\forall \tilde{\xi}_i \in B_{\Xi'}(\xi_i, \delta')$, it holds that $\Omega_i(\tilde{\xi}_i) \subset \mathcal{U}$. Let $\mathcal{U} = B_{2^X}(\Omega_i(\xi_i), \epsilon')$, and we obtain upper semicontinuity at ξ_i . ■

With the above results, we are ready to show the convergence of primal estimates.

Lemma 5.6 (Primal estimate convergence): Under the assumptions in Theorem 3.1, for each $i \in V$, it holds that $\lim_{k \rightarrow +\infty} x_i(k) = \tilde{x}_i$ where $\tilde{x}_i = \Omega_i(\xi_i^*)$.

Proof: The combination of upper semicontinuity of Ω_i in Lemma 5.6 and $\lim_{k \rightarrow +\infty} \xi_i(k) = \xi_i^*$ with ξ_i^* given in Lemma 5.4 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. Since $\Omega_i(\xi_i^*)$ is singleton, then $\tilde{x}_i = \Omega_i(\xi_i^*)$. We arrive in the desired result. ■

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

Proof for Theorem 3.1:

Claim 1: $\langle -\Delta - \tilde{x}_i + \tilde{x}_{i_D}, \lambda_i^* \rangle = 0$, $\langle -\Delta + \tilde{x}_i - \tilde{x}_{i_D}, w_i^* \rangle = 0$ and $\langle g(\tilde{x}_i), \mu_i^* \rangle = 0$.

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

$$\begin{aligned}
& - \sum_{i \in V} 2\alpha(k) (\langle -\Delta - x_i(k), v_\lambda^i(k)_i - \lambda_i \rangle + \langle x_{i_D}(k), v_\lambda^{i_D}(k)_i - \lambda_i \rangle) \\
& \leq \alpha(k)^2 \sum_{i \in V} \|\mathcal{D}_i(k)\|^2 + \sum_{i \in V} (\|\xi_i(k) - \xi_i\|^2 - \|\xi_i(k+1) - \xi_i\|^2) \\
& + 2\alpha(k) \sum_{i \in V} \{ \langle -x_i(k), v_w^i(k)_{i_U} - w_{i_U} \rangle + \langle x_i(k) - \Delta, v_w^i(k)_i - w_i \rangle + \langle g(x_i(k)), \mu_i(k) - \mu_i \rangle \}.
\end{aligned} \tag{16}$$

Sum (16) over $[0, K]$, divide by $s(K) := \sum_{k=0}^K \alpha(k)$, and we have

$$\begin{aligned}
& \frac{1}{s(K)} \sum_{k=0}^K \alpha(k) \sum_{i \in V} 2(\langle \Delta + x_i(k), v_\lambda^i(k)_i - \lambda_i \rangle + \langle -x_{i_D}(k), v_\lambda^{i_D}(k)_i - \lambda_i \rangle) \\
& \leq \frac{1}{s(K)} \sum_{k=0}^K \alpha(k)^2 \sum_{i \in V} \|\mathcal{D}_i(k)\|^2 + \frac{1}{s(K)} \left\{ \sum_{i \in V} (\|\xi_i(0) - \xi_i\|^2 - \|\xi_i(K+1) - \xi_i\|^2) \right. \\
& \left. + \sum_{k=0}^K 2\alpha(k) \sum_{i \in V} (\langle g(x_i(k)), \mu_i(k) - \mu_i \rangle + \langle x_i(k) - \Delta, v_w^i(k)_i - w_i \rangle + \langle -x_i(k), v_w^i(k)_{i_U} - w_{i_U} \rangle) \right\}.
\end{aligned} \tag{17}$$

We now proceed to show $\langle -\Delta - \tilde{x}_i + \tilde{x}_{i_D}, \lambda_i^* \rangle \geq 0$ for each $i \in V$. Notice that we have shown that $\lim_{k \rightarrow +\infty} \|x_i(k) - \tilde{x}_i\| = 0$ for all $i \in V$, and it also holds that $\lim_{k \rightarrow +\infty} \|\xi_i(k) - \xi_i^*\| = 0$ for all $i \in V$. Let $\lambda_i = \frac{1}{2}\lambda_i^*$, $\lambda_j = \lambda_j^*$ for $j \neq i$ and $\mu_i = \mu_i^*$, $w = w^*$ in (17). Recall that $\{\alpha(k)\}$ is not summable but square summable, and $\{\mathcal{D}_i(k)\}$ is uniformly bounded. Take $K \rightarrow +\infty$, and then it follows from Lemma 5.1 in [36] that:

$$\langle \Delta + \tilde{x}_i - \tilde{x}_{i_D}, \lambda_i^* \rangle \leq 0. \tag{18}$$

On the other hand, since $\xi^* \in D_\Delta^*$, we have $\|\xi^*\| \leq \gamma$ given the fact that γ is an upper bound of D_Δ^* . Let $\xi := (\mu, \lambda, w)$ where $\xi_i := (\mu_i, \lambda, w)$. Then we could choose a sufficiently small $\delta' > 0$ and $\xi \in \Xi$ in (17) such that $\|\xi_i\| \leq \gamma + \theta_i$ where θ_i is given in the definition of M_i and ξ is given by: $\lambda_i = (1 + \delta')\lambda_i^*$, $\lambda_j = \lambda_j^*$ for $j \neq i$, $w = w^*$, $\mu = \mu^*$. Following the same lines toward (18), it gives that $-\delta \langle \Delta + \tilde{x}_i - \tilde{x}_{i_D}, \lambda_i^* \rangle \leq 0$. Hence, it holds that $\langle -\Delta - \tilde{x}_i + \tilde{x}_{i_D}, \lambda_i^* \rangle = 0$. The rest of the proof is analogous and thus omitted. \blacksquare

Claim 2: \tilde{x} is primal feasible to the approximate 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 $\|\xi^*\| \leq \gamma$, we could choose a sufficiently small $\delta' > 0$ and $\xi := (\mu, \lambda, w)$ where $\xi_i := (\mu_i, \lambda, w)$ and $\|\xi_i\| \leq \gamma + \theta_i$ in (17) as follows: if $(-\Delta - \tilde{x}_i + \tilde{x}_{i_D})_\ell > 0$, then $(\lambda_i)_\ell = (\lambda_i^*)_\ell + \delta'$; otherwise, $(\lambda_i)_\ell = (\lambda_i^*)_\ell$, and $w = w^*$, $\mu = \mu^*$. 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. We conclude that \tilde{x} is primal feasible to the approximate problem (P_Δ) . ■

Claim 3: \tilde{x} is a primal solution to the problem (P_Δ) .

Proof: Since \tilde{x} is primal feasible to the approximate problem (P_Δ) , then $\sum_{i \in V} f_i(\tilde{x}_i) \geq p_\Delta^*$. On the other hand, it follows from Claim 1 that

$$\sum_{i \in V} f_i(\tilde{x}_i) = \sum_{i \in V} \mathcal{L}_i(\tilde{x}_i, \xi_i^*) \leq \sum_{i \in V} Q_i(\xi_i^*) \leq p_\Delta^*.$$

We then conclude that $\sum_{i \in V} f_i(\tilde{x}_i) = p_\Delta^*$. In conjunction with the feasibility of \tilde{x} , this further ensures that \tilde{x} is primal optimal to the problem (P_Δ^*) . This completes the proofs for Theorem 3.1. ■

VI. SIMULATIONS

In this section, we examine several numerical examples to illustrate the performance of our algorithm.

A. Robust source localization

We consider a robust source localization problem where the objective function is adopted from [1], [24]. In particular, consider a network of four agents $V \triangleq \{1, \dots, 4\}$. The objective functions of agents are piecewise linear and given by $f_i(z) = |||z - a_i|| - r|$. The local inequality functions are given by:

$$g_1(z) = \begin{bmatrix} z_1 - 8 \\ -z_1 - 8 \\ z_2 - 8 \\ -z_2 - 8 \end{bmatrix}, \quad g_2(z) = \begin{bmatrix} z_1 - 9 \\ -z_1 - 9 \\ z_2 - 9 \\ -z_2 - 9 \end{bmatrix}, \quad g_3(z) = \begin{bmatrix} z_1 - 8.5 \\ -z_1 - 8.5 \\ z_2 - 8.5 \\ -z_2 - 8.5 \end{bmatrix}, \quad g_4(z) = \begin{bmatrix} z_1 - 9.5 \\ -z_1 - 9.5 \\ z_2 - 9.5 \\ -z_2 - 9.5 \end{bmatrix},$$

and, the local constraint sets are given by

$$\begin{aligned} X_1 &= \{z \in \mathbb{R}^2 \mid -10 \leq z_1 \leq 10, -10 \leq z_2 \leq 10\}, \\ X_2 &= \{z \in \mathbb{R}^2 \mid -10.5 \leq z_1 \leq 10.5, -10.5 \leq z_2 \leq 10.5\}, \\ X_3 &= \{z \in \mathbb{R}^2 \mid -9 \leq z_1 \leq 9, -10 \leq z_2 \leq 10\}, \\ X_4 &= \{z \in \mathbb{R}^2 \mid -11 \leq z_1 \leq 11, -9 \leq z_2 \leq 9\}. \end{aligned}$$

In the simulation, we choose the parameter $\delta = 0.1$. The local Lagrangian function can be written as $\mathcal{L}_i(x_i, \xi_i) = f_i(x_i) + \langle \zeta_i, x_i \rangle$ by dropping the terms independent of x_i and ζ_i is linear in ξ_i . Figure 3 shows the sectional plot of $f(z) \triangleq \sum_{i \in V} f_i(z)$ along z_1 -axle, demonstrating that f is nonconvex and has local minima.

The inter-agent topologies $\mathcal{G}(k)$ are given by: $\mathcal{G}(k)$ is $1 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 4$ when k is odd, and $\mathcal{G}(k)$ is $1 \rightarrow 2 \leftrightarrow 3 \leftarrow 4 \rightarrow 1$ when k is even. It is easy to see that $\mathcal{G}(k)$ satisfies the periodical strong connectivity assumption 2.3.

1) *Simulation 1; the assumption of SD is satisfied:* For this numerical simulation, we consider the set of parameters $r = 0.75$, $a_1 = [0 \ 0]^T$, $a_2 = [0 \ 1]^T$, $a_3 = [1 \ 0]^T$ and $a_4 = [1 \ 1]^T$. Figure 1 shows the surface of the global objective function $f(x) = \sum_{i \in V} f_i(x)$. The contour, Figure 2, indicates that the set of optimal solutions is a region around $[0.5 \ 0.5]^T$. Figure 4 is the sectional plot of f_1 along z_1 -axle.

From Figures 5–8, one can see that $\zeta_i(k)$ converges to some point $(0, 0.05] \times (0, 0.05]$. Hence, $\xi^* \in D_s^*$; i.e., the assumption of SD is satisfied.

The simulation results are shown in Figures 9 to 10. In particular, Figure 9 (resp. Figure 10) shows the evolution of primal estimates of the primal solution $x^*(1)$ (resp. $x^*(2)$). After about 25 iterates, the primal estimates oscillate within a very small region and eventually agree upon the point $[0.4697 \ 0.472]^T$ which coincides with a global optimal solution.

2) *Simulation 2; the assumption of SD is violated:* Consider the same problem as Simulation 1 with $r = 0.75$ and $a_i = [0 \ 0]^T$ for $i \in V$. From Figures 11 and 12, one can see that $\zeta_i(k)$ converges to $[0 \ 0]^T$. Hence, $\xi^* \notin D_s^*$ and the assumption of SD is not satisfied. Figures 13 and 14 confirms that primal estimates fail to converge in this case.

3) *Simulation 3; comparison with gradient-based algorithms:* Consider the same set of parameters as in Simulation 1 without including the inequality constraints. The multi-agent interaction

topologies are the same. We implement the diffusion gradient algorithm in [22] for this problem. Figures 15 and 16 show that the primal estimates reach the consensus value of $[-0.65 \ -0.38]^T$ after 40000 iterates. From Figure 2, it is clear that $[-0.65 \ -0.38]^T$ is not a global optimum. By comparing Figures 9, 10, 15 and 16, one can see that our algorithm is much faster than the diffusion gradient method at the expense of solving a global optimization problem at each iterate.

We also implement the incremental gradient algorithm in [33] for the same set of parameters in Simulation 1 without including inequality constraints. Figure 17 demonstrates that the performance of the incremental gradient method is analogous to the diffusion gradient algorithm; i.e., the estimates are trapped in some local minimum, and the convergence rate is slower than our algorithm.

B. Nonconvex quadratic programming

Consider a network of four agents where the topologies are the same as before. The local objective function is $f_i(z) = \|z\|_{P_i}^2 + \langle q_i, z \rangle$ and the local constraint function is $g_i(z) = \|z\|_{A_i}^2 + \langle b_i, z \rangle + c_i \leq 0$. In particular, we use the following parameters:

$$\begin{aligned}
 P_1 = P_2 = P_3 &= \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, & P_4 &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\
 A_1 = A_4 &= \begin{bmatrix} 18 & 0 \\ 0 & 8 \end{bmatrix}, & b_1 = b_4 &= [2 \ 0], & c_1 = c_4 &= -1, \\
 A_2 &= \begin{bmatrix} 13 & -2 \\ -2 & 8 \end{bmatrix}, & b_2 &= [0 \ 4], & c_2 &= -1, \\
 A_3 &= \begin{bmatrix} 5 & -5 \\ -5 & 5 \end{bmatrix}, & b_3 &= [10 \ 10], & c_3 &= -1.
 \end{aligned}$$

And the local constraint sets are given by

$$\begin{aligned}
 X_1 &= \{z \in \mathbb{R}^2 \mid -10 \leq z_1 \leq 10, \ -10 \leq z_2 \leq 10\}, \\
 X_2 &= \{z \in \mathbb{R}^2 \mid -10.5 \leq z_1 \leq 10.5, \ -10.5 \leq z_2 \leq 10.5\}, \\
 X_3 &= \{z \in \mathbb{R}^2 \mid -9 \leq z_1 \leq 9, \ -10 \leq z_2 \leq 10\}, \\
 X_4 &= \{z \in \mathbb{R}^2 \mid -11 \leq z_1 \leq 11, \ -9 \leq z_2 \leq 9\}.
 \end{aligned}$$

One can see that the sum of P_i is

$$P = \sum_{i=1}^4 P_i = \begin{bmatrix} 0 & 4 \\ 4 & 3 \end{bmatrix}$$

which is indefinite. We choose $\delta = 0.3$ for the simulation.

The dual estimates associated with the inequality constraints converge to $\mu_1^* = 0.5027$, $\mu_2^* = 3.1061$, $\mu_3^* = 1.8792$ and $\mu_4^* = 2.2910$ in Figure 20. One can verify that properties P1 and P2 hold in this case:

$$P_i + \mu_i^* A_i > 0, \quad (P_i + \mu_i^* A_i)^{-1} \zeta_i^* \in X_i, \quad i \in V.$$

The primal estimates converge to $[-0.1933 \quad -0.3005]^T$, $[-0.2621 \quad -0.5360]^T$, $[-0.1013 \quad -0.0116]^T$ and $[-0.2144 \quad -0.2667]^T$ in Figures 18 and 19, and the collection of these points consists of a global optimal solution to the approximate problem.

VII. CONCLUSIONS

We have studied a distributed dual algorithm for a class of multi-agent nonconvex optimization problems. The convergence of the algorithm has been proven under the assumptions that (i) the Slater's condition holds; (ii) the optimal solution set of the dual limit is singleton; (iii) the network topologies are strongly connected over any given bounded period. An open question is how to address the shortcomings imposed by nonconvexity and multi-agent interactions settings.

REFERENCES

- [1] A. Abramo, F. Blanchini, L. Geretti, and C. Savorgnan. A mixed convex/nonconvex distributed localization approach for the deployment of indoor positioning services. *IEEE Transactions Mobile Computing*, 7(11):1325–1337, 2008.
- [2] J.P. Aubin and H. Frankowska. *Set-valued analysis*. Birkhäuser, 1990.
- [3] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [4] D.P. Bertsekas. *Convex optimization theory*. Anthena Scietific, 2009.
- [5] D.P. Bertsekas, A. Nedic, and A. Ozdaglar. *Convex analysis and optimization*. Anthena Scietific, 2003.
- [6] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, 2006.
- [7] 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>.
- [8] R.S. Burachik. On primal convergence for augmented Lagrangian duality. *Optimization*, 60(8):979–990, 2011.
- [9] R.S. Burachik and C.Y. Kaya. An update rule and a convergence result for a penalty function method. *Journal of Industrial Management and Optimization*, 3(2):381–398, 2007.

- [10] 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.
- [11] 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.
- [12] D.Y. Gao, N. Ruan, and H. Sherali. Solutions and optimality criteria for nonconvex constrained global optimization problems with connections between canonical and Lagrangian duality. *Journal of Global Optimization*, 45(3):474–497, 2009.
- [13] B. Gharesifard and J. Cortés. Distributed strategies for generating weight-balanced and doubly stochastic digraphs. *European Journal of Control*, 2012. To appear.
- [14] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*. Springer, 1996.
- [15] 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.
- [16] 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.
- [17] A. Kashyap, T. Başar, and R. Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.
- [18] 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.
- [19] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.
- [20] A. Nedic and A. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization*, 19(4):1757–1780, 2009.
- [21] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transaction on Automatic Control*, 54(1):48–61, 2009.
- [22] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.
- [23] R. D. Nowak. Distributed EM algorithms for density estimation and clustering in sensor networks. *IEEE Transactions on Signal Processing*, 51:2245–2253, 2003.
- [24] P. Oguz-Ekim, J. Gomes, J. Xavier, and P. Oliveira. A convex relaxation for approximate maximum-likelihood 2D source localization from range measurements. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 2698–2701, 2010.
- [25] 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.
- [26] A. Olshevsky and J. N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009.
- [27] P.M. Pardalos and S.A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1(1):15–22, 1991.
- [28] B.T. Polyak. A general method for solving extremum problems. *Soviet Mathematics Doklady*, 3(8):593–597, 1967.

- [29] M. G. Rabbat and R. D. Nowak. Decentralized source localization and tracking. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 921–924, May 2004.
- [30] M. G. Rabbat and R. D. Nowak. Quantized incremental algorithms for distributed optimization. *IEEE Journal on Select Areas in Communications*, 23:798–808, 2005.
- [31] W. Ren and R. W. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control*. Communications and Control Engineering. Springer, 2008.
- [32] R.T. Rockafellar and R.J.-B Wets. *Variational analysis*. Springer, 1998.
- [33] 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*, 55(2):488–492, 2010.
- [34] 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.
- [35] M. Zhu and S. Martínez. An approximate dual subgradient algorithm for multi-agent non-convex optimization. In *IEEE International Conference on Decision and Control*, pages 7487–7492, Atlanta, GA, USA, December 2010.
- [36] M. Zhu and S. Martínez. On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods. *IEEE Transactions on Automatic Control*, 57:151–164, 2012.

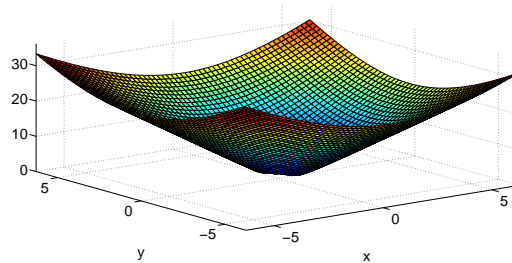


Fig. 1. The 3 – D plot of the global objective function

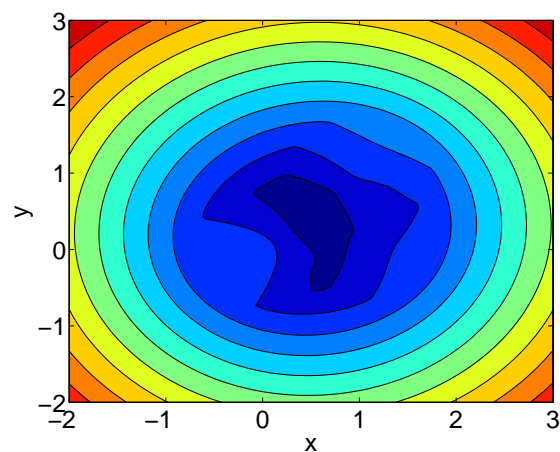


Fig. 2. The contour of the global objective function

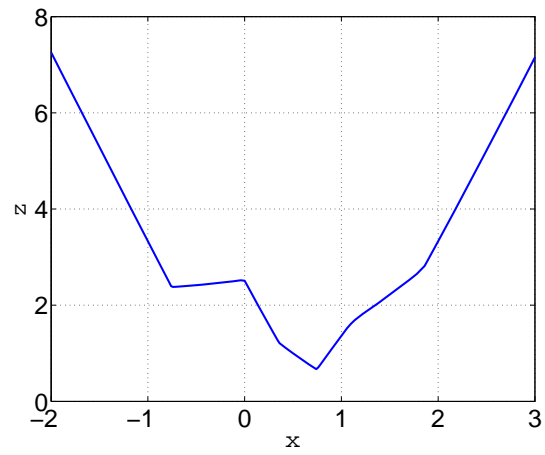


Fig. 3. The sectional plot of the global objective function along z_1 -axle

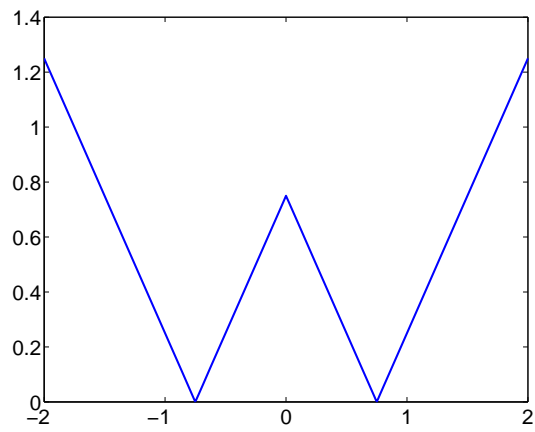


Fig. 4. The sectional plot of f_1 along z_1 -axle

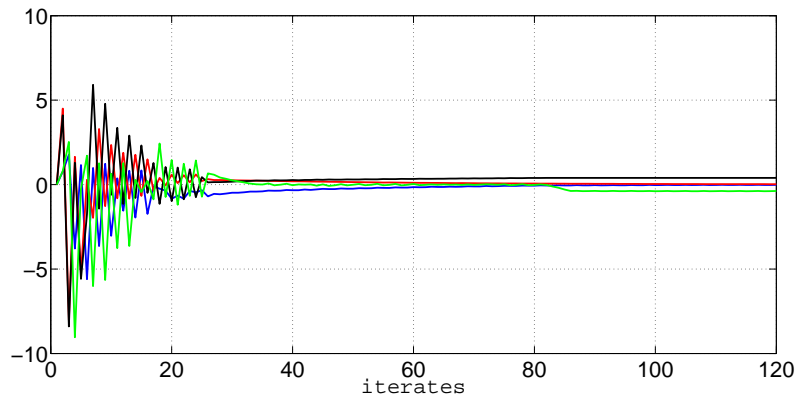


Fig. 5. The evolution of $\zeta_{i,1}$

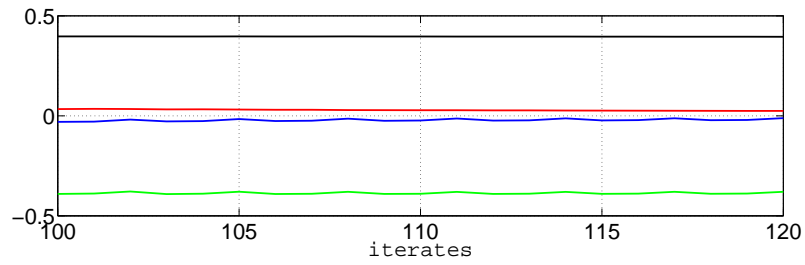


Fig. 6. A portion of the evolution of $\zeta_{i,1}$

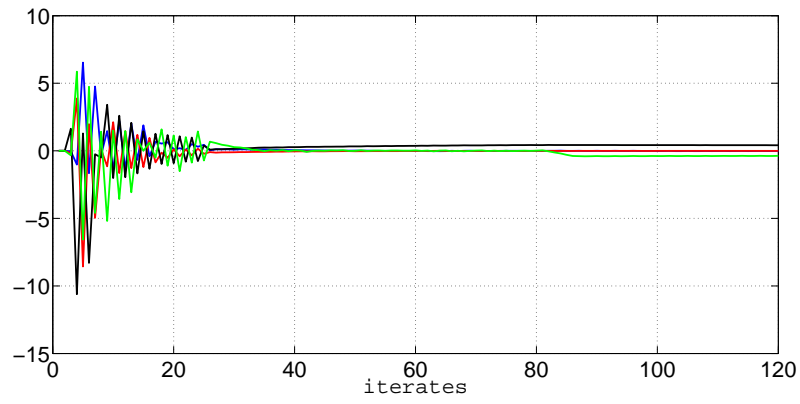


Fig. 7. The evolution of $\zeta_{i,2}$

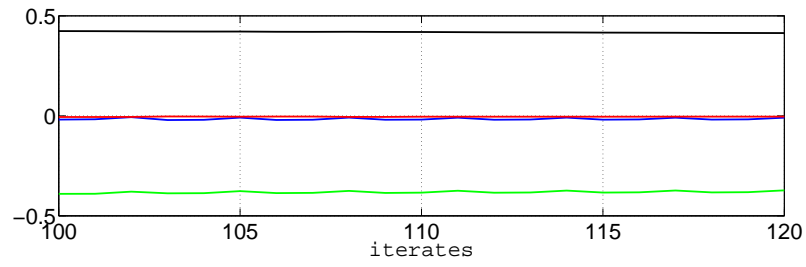


Fig. 8. A portion of the evolution of $\zeta_{i,2}$

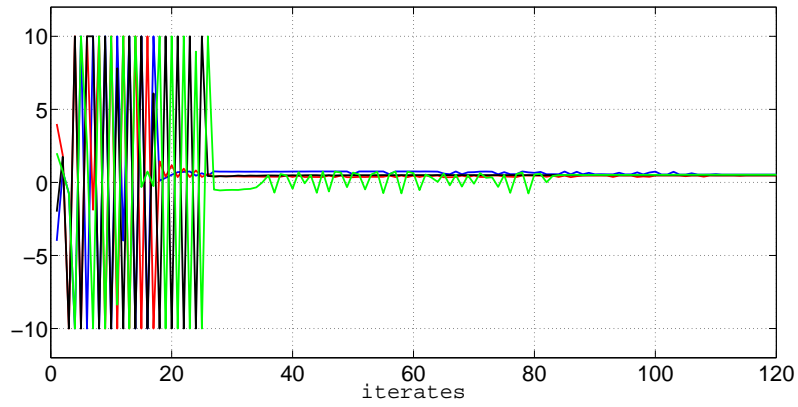


Fig. 9. The primal estimates of $x^*(1)$

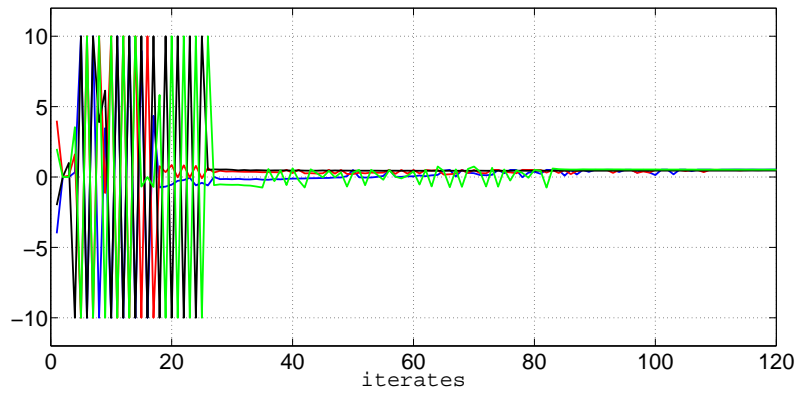


Fig. 10. The primal estimates of $x^*(2)$

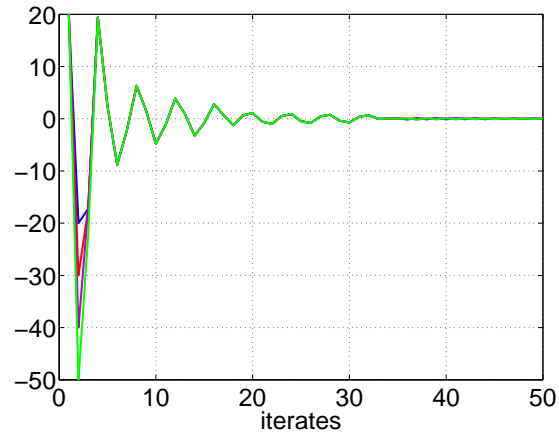


Fig. 11. The evolution of $\zeta_{i,1}$

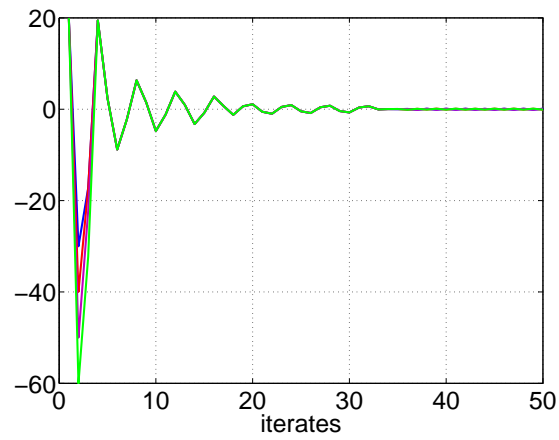


Fig. 12. The evolution of $\zeta_{i,2}$

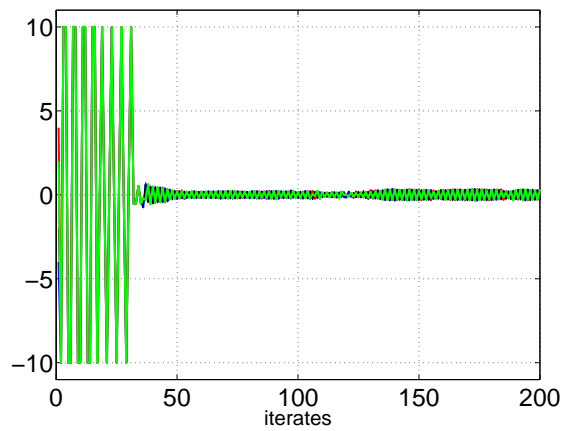


Fig. 13. The primal estimates of $x^*(1)$

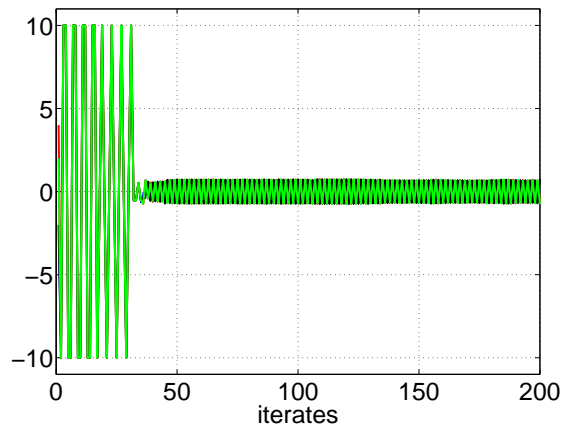


Fig. 14. The primal estimates of $x^*(2)$

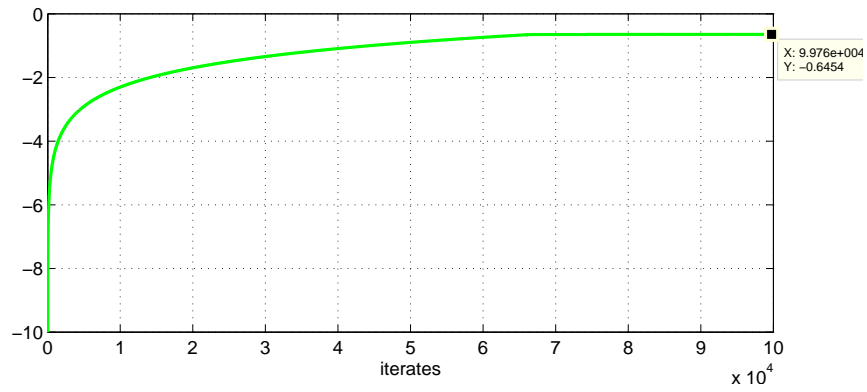


Fig. 15. The primal estimates of $x^*(1)$ of the diffusion gradient method

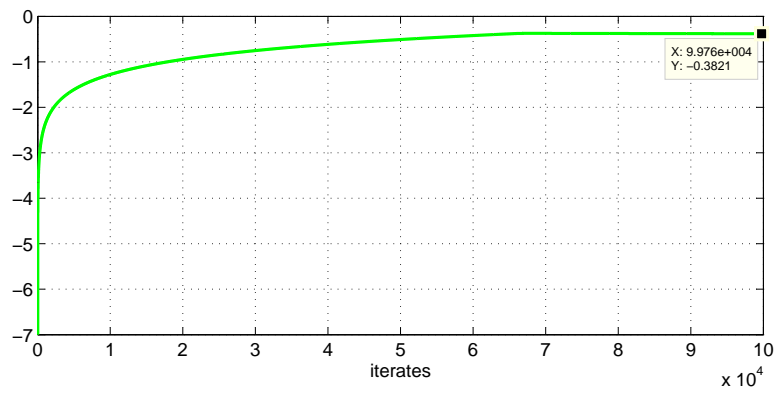


Fig. 16. The primal estimates of $x^*(2)$ of the diffusion gradient method

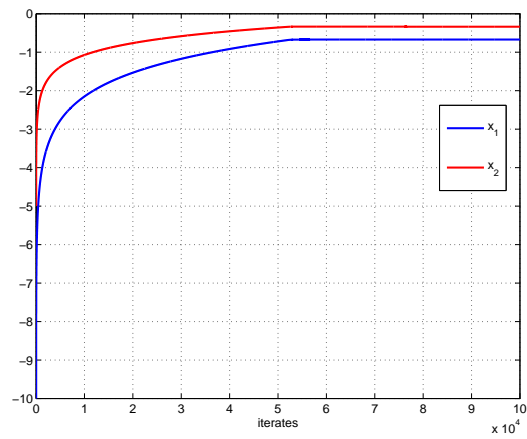


Fig. 17. The primal estimates of $x^*(1)$ of the incremental gradient method

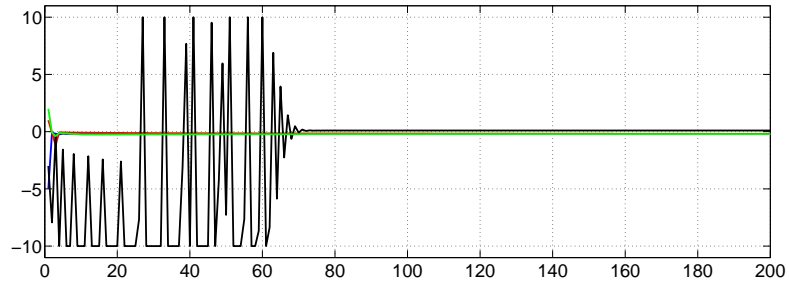


Fig. 18. The primal estimates of $x^*(1)$ of quadratic programming

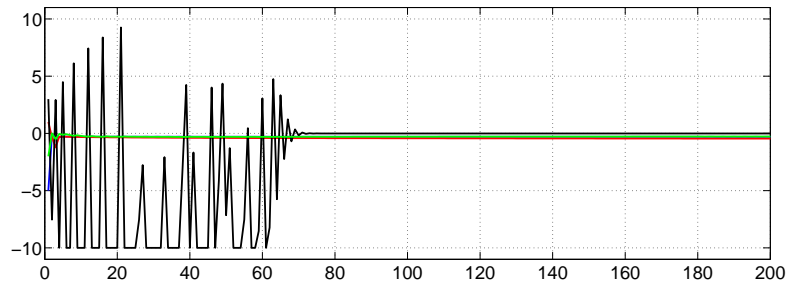


Fig. 19. The primal estimates of $x^*(2)$ of quadratic programming

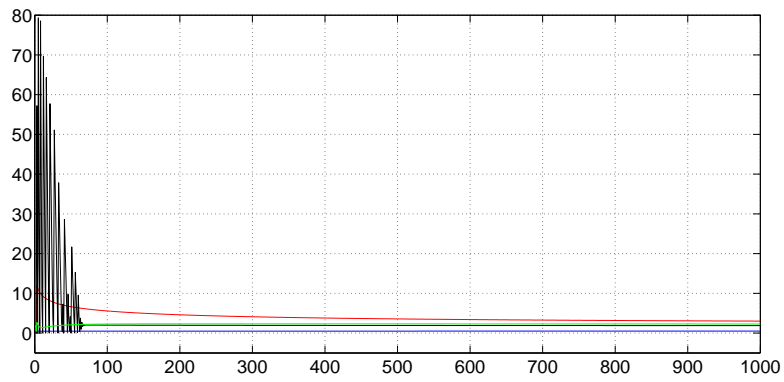


Fig. 20. The dual estimates of μ_i^* of quadratic programming