On the Rate of Convergence of Distributed Subgradient Methods for Multi-agent Optimization

Angelia Nedić
Department of Industrial and
Enterprise Systems Engineering
University of Illinois
Urbana-Champaign, IL 61801
Email: angelia@uiuc.edu

Asuman Ozdaglar
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02142
Email: asuman@mit.edu

***Abstract*— We study a distributed computation model for optimizing the sum of convex (nonsmooth) objective functions of multiple agents. We provide convergence results and estimates for convergence rate. Our analysis explicitly characterizes the tradeoff between the accuracy of the approximate optimal solutions generated and the number of iterations needed to achieve the given accuracy.**

I. INTRODUCTION

There has been considerable recent interest in the analysis of large-scale networks, such as the Internet, which consist of multiple agents with different objectives. For such networks, it is essential to design network control methods that can operate in a decentralized manner with limited local information and converge to an approximately optimal operating point fairly rapidly. Recent literature has adopted the utility-maximization framework of economics to design distributed algorithms that captures the multiple objectives of different agents, represented by different utility functions (see Kelly et al. [7], Low and Lapsley [8], and Srikant [13]). This framework builds on convex optimization duality and is limited to applications where the utility function of an agent depends only on the resource allocated to that agent. In many applications however, individual agent's utility depends on the entire resource allocation vector.

In this paper, we study a simple distributed computation model that captures these interactions and provide convergence rate analysis for the resulting algorithm. In particular, we consider a network consisting of $V = \{1, \ldots, m\}$ nodes (or agents) that cooperatively minimize a common additive cost. More formally, the agents want to cooperatively solve the following unconstrained optimization problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m f_i(x) \\ \text{subject to} & x \in \mathbb{R}^n, \end{array} \tag{1}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ is a convex function for all i. Let $f(x) = \sum_{i=1}^m f_i(x)$. We denote the optimal value of this problem by f^*. We also denote the optimal solution set by X^*.

We assume that each agent i has information only about his/her cost function f_i. Every agent generates and maintains estimates of the optimal solution to problem (1), and communicates them directly or indirectly to the other agents. Each agent updates his/her estimate by combining it with the estimates received from the other agents (if any) and by using the subgradient information of f_i.

Our model is in the spirit of the distributed computation model proposed by Tsitsiklis [14] (see also Tsitsiklis et al. [15], Bertsekas and Tsitsiklis [3]). There, the main focus is on minimizing a (smooth) function $f(x) = \sum_{i=1}^m f_i(x)$ by distributing the vector components $x_j,\ j = 1, \ldots, n$ among n processors. The possibility of distributing the component functions f_i among m agents has been suggested in [14], but has not been explored fully. Here, we pursue this idea in depth with focus on the case where the component functions f_i are convex but not necessarily smooth. To our knowledge this is the first distributed model of such kind that is rigorously analyzed. For this model, we present convergence results and estimates for the rate of convergence. In particular, we show that there is a tradeoff between the quality of an approximate optimal solution and the computation load required to generate such a solution. Our convergence rate estimate captures this dependence explicitly in terms of the system and algorithm parameters.

Related computational models for reaching consensus on a particular scalar value have attracted a lot of recent attention as natural models of cooperative behavior in networked-systems (see Vicsek et al. [16] and Jadbabaie et al. [6]). There is also another line of related work that focuses on computing exact averages of the initial values of the agents (see Boyd et al. [5] and Kashyap et al. [1]).

The remainder of this paper is organized as follows: in Section 2, we describe a distributed computation model and the assumptions on the interaction rules among the agents. In Section 3, we present our assumptions and preliminary results. In Section 4, we provide our main convergence and rate of convergence results. Finally, in Section 5, we present our concluding remarks.

Notation and Basic Notions For a matrix A, we write A_i^j or $[A]_i^j$ to denote the matrix entry in the i-th row and j-th column. We write $[A]_i$ to denote the i-th row of the matrix A, and $[A]^j$ to denote the j-th column of A. A vector $a \in \mathbb{R}^m$ is said to be a *stochastic vector* when its components a_i, $i = 1, \ldots, m$, are nonnegative and $\sum_{i=1}^m a_i = 1$. A square $m \times m$ matrix

A is said to be a *stochastic matrix* when each row of A is a stochastic vector. A square $m \times m$ matrix A is said to be *doubly stochastic* when both A and A' are stochastic matrices.

For a function $F : \mathbb{R}^n \to (-\infty, \infty]$, we denote the domain of F by $\mathrm{dom}(F) = \{x \in \mathbb{R}^n \mid F(x) < \infty\}$. We use the notion of a subgradient of a *convex* function: a vector $s_F(\bar{x}) \in \mathbb{R}^n$ *is a subgradient of a convex function F at $\bar{x} \in \mathrm{dom}(F)$* when the following relation holds:

$$F(\bar{x}) + s_F(\bar{x})'(x - \bar{x}) \leq F(x) \qquad \text{for all } x \in \mathrm{dom}(F). \quad (2)$$

The set of all subgradients of F at $\bar{x}$ is denoted by $\partial F(\bar{x})$ (see [2]).

II. Computation Model

We consider a model where all agents update their estimates x^i at some discrete times $t_k, k = 0, 1, \ldots$. We denote by $x^i(k)$ the vector estimate computed by agent i at the time t_k. The agents exchange their estimates as follows: When generating a new estimate, an agent i combines his/her current estimate x^i with the estimates x^j received from some of the other agents j. For simplicity, in this paper we focus on a model in the absence of delays.[1] Specifically, agent i updates his/her estimates according to the following relation:

$$x^i(k+1) = \sum_{j=1}^{m} a^i_j(k) x^j(k) - \alpha^i(k) d_i(k), \qquad (3)$$

where the vector $a^i = (a^i_1, \ldots, a^i_m)'$ is a vector of weights and the scalar $\alpha^i(k) > 0$ is a stepsize used by agent i. The vector $d_i(k)$ is a subgradient of the agent i objective function $f_i(x)$ at $x = x^i(k)$.

In our convergence analysis of the model in Eq. (3), we use a linear description of the estimate evolution in time, which is motivated by work of Tsitsiklis [14]. In particular, we introduce matrices $A(s)$ whose i-th column is the vector $a^i(s)$. Using these matrices we can relate estimate $x^i(k+1)$ to the estimates $x^1(s), \ldots, x^m(s)$ for any $s \leq k$. In particular, it is straightforward to verify that for the iterates generated by Eq. (3), we have for any i, and any s and k with $k \geq s$,

$$\begin{aligned} x^i(k+1) = &\sum_{j=1}^{m} [A(s)A(s+1)\cdots A(k-1)a^i(k)]_j x^j(s) \\ &- \sum_{j=1}^{m} [A(s+1)\cdots A(k-1)a^i(k)]_j \alpha^j(s) d_j(s) \\ &- \cdots - \sum_{j=1}^{m} [A(k-1)a^i(k)]_j \alpha^j(k-2) d_j(k-2) \\ &- \sum_{j=1}^{m} [a^i(k)]_j \alpha^j(k-1) d_j(k-1) - \alpha^i(k) d_i(k). \quad (4) \end{aligned}$$

For all s and k with $k \geq s$, we introduce the matrices

$$\Phi(k,s) = A(s)A(s+1)\cdots A(k-1)A(k),$$

where

$$\Phi(k,k) = A(k) \qquad \text{for all } k.$$

Note that the i-th column of $\Phi(k,s)$ is given by

$$[\Phi(k,s)]^i = A(s)A(s+1)\cdots A(k-1)a^i(k),$$

while the entry in i-th column and j-th row of $\Phi(k,s)$ is given by

$$[\Phi(k,s)]^i_j = [A(s)A(s+1)\cdots A(k-1)a^i(k)]_j.$$

Using the preceding relations, we can now rewrite relation (4) in terms of the matrices $\Phi(k,s)$, as follows:

$$\begin{aligned} x^i(k+1) = &\sum_{j=1}^{m} [\Phi(k,s)]^i_j x^j(s) - \sum_{j=1}^{m} [\Phi(k,s+1)]^i_j \alpha^j(s) d_j(s) \\ &- \cdots - \sum_{j=1}^{m} [\Phi(k,k-1))]^i_j \alpha^j(k-2) d_j(k-2) \\ &- \sum_{j=1}^{m} [\Phi(k,k)]^i_j \alpha^j(k-1) d_j(k-1) - \alpha^i(k) d_i(k). \end{aligned}$$

More compactly, we have for any $i \in \{1, \ldots, m\}$, and s and k with $k \geq s$,

$$\begin{aligned} x^i(k+1) \;=\; &\sum_{j=1}^{m} [\Phi(k,s)]^i_j x^j(s) \qquad (5) \\ &- \sum_{r=s+1}^{k} \left(\sum_{j=1}^{m} [\Phi(k,r)]^i_j \alpha^j(r-1) d_j(r-1) \right) \\ &- \alpha^i(k) d_i(k). \end{aligned}$$

We next consider some rules on agent interactions and study the properties of the matrices $\Phi(k,s)$ resulting from these rules.

III. Assumptions and Preliminary Results

We impose some rules on the information exchange among the agents that can guarantee the convergence of the estimates generated by the model given in Eq. (3). These rules are in the spirit of those proposed by Tsitsiklis [14] (see also the recent paper of Blondel et al. [4]).

We start with the rules on agent interactions. In particular, at first, we specify a rule that agents use when combining their estimates x^i. This rule is described in terms of the weights $a^i_j(k)$ in relation Eq. (3).

Assumption 1: (Weights Rule) We have:

(a) There exists a scalar η with $0 < \eta < 1$ such that
  - (i) $a^i_i(k) \geq \eta$ for all i and k.
  - (ii) $a^i_j(k) \geq \eta$ for all i, k, and j when agent j communicates with agent i in the time interval (t_k, t_{k+1}).
  - (iii) $a^i_j(k) = 0$ for all i, k, and j otherwise.

(b) The vectors $a^i(k)$ are stochastic, i.e., $\sum_{j=1}^{m} a^i_j(k) = 1$ for all i and k.

Assumption 1(a) states that each agent gives significant weights to his/her own estimate $x^i(k)$ and all estimates $x^j(k)$

[1] In ongoing work, we consider a more general setting where there is a delay in delivering a message from one agent to another.

available from other agents j at the update time. Naturally, an agent i assigns zero weights to the estimates x^j for those agents j whose estimate information is not available at the update time.[2] Consider the matrix $A(k)$ whose columns are $a^1(k), \ldots, a^m(k)$. We note that under Assumption 1, the transpose $A'(k)$ is a stochastic matrix for all k.

At each update time t_k, the information exchange among the agents may be represented by a directed graph (V, E_k) with the set of directed edges E_k given by

$$E_k = \{(j,i) \mid a^i_j(k) > 0\}.$$

From Assumption 1(a), it follows that $(i,i) \in E_k$ for all i and k, and that $(j,i) \in E_k$ if and only if agent i receives an estimate $x^j(k)$ from agent j in the time interval (t_k, t_{k+1}).

The next assumption states that following any time t_k, the information of an agent j reaches each and every agent i directly or indirectly (through a sequence of communications between the other agents).

Assumption 2: (Connectivity) The graph (V, E_∞) with the set of edges E_∞ consisting of agent pairs (j,i) communicating infinitely many times, i.e.,

$$E_\infty = \{(j,i) \mid (j,i) \in E_k \text{ for infinitely many indices } k\},$$

is connected.

In other words, this assumption states that for any k and any agent pair (j,i), there is a directed path from agent j to agent i with the path edges in the set $\cup_{l \geq k} E_l$. Thus, Assumption 2 is equivalent to the assumption that the composite directed graph $(V, \cup_{l \geq k} E_l)$ is connected for all k.

In our convergence analysis, we also adopt the following assumption, which states that the intercommunication intervals are bounded for those agents that communicate directly.

Assumption 3: (Bounded Intercommunication Interval) There exists an integer $B \geq 1$ such that

$$(j,i) \in E_k \cup E_{k+1} \cup \cdots \cup E_{k+B-1}$$

for all k and any agent j that communicates directly to an agent i infinitely many times.

Worded differently, this assumption requires that, for any agent pair (j,i) such that agent j communicates directly to agent i infinitely many times [i.e., $(j,i) \in E_\infty$], agent j provides information to agent i at least once every B consecutive updates.

A. Properties of the Matrices $\Phi(k,s)$

In this section, we establish some properties of the matrices $\Phi(k,s)$ for $k \geq s$. The convergence properties of these matrices as $k \to \infty$ have been extensively studied and well-established (see [14], [6], [17]). Our contribution in this section is the provision of the convergence rate estimates explicitly as a function of the system parameters. We skip the proofs of the lemmas due to space constraints and refer the interested reader to Tsitsiklis [14] and the technical report [12].

Recall that for all k and s with $k \geq s$,

$$\Phi(k,s) = A(s)A(s+1)\cdots A(k-1)A(k), \tag{6}$$

where

$$\Phi(k,k) = A(k) \qquad \text{for all } k. \tag{7}$$

Let s be arbitrary and consider the matrices

$$\Phi(s + k\bar{B} - 1, s + (k-1)\bar{B}) \qquad \text{for } k = 1, 2, \ldots,$$

with the scalar $\bar{B}$ given by

$$\bar{B} = (m-1)B, \tag{8}$$

where m is the number of agents and B is the intercommunication interval bound of Assumption 3. The following lemma states that, for any fixed s, the product of the matrices $\Phi(k,s)$ converges as $k \to \infty$.

Lemma 1: Let Weights Rule, Connectivity, and Bounded Intercommunication Interval assumptions hold [cf. Assumptions 1, 2, and 3]. Let s be arbitrary and define

$$D_k = \Phi'\left(s + k\bar{B} - 1, s + (k-1)\bar{B}\right) \qquad \text{for } k \geq 1, \tag{9}$$

where the scalar $\bar{B}$ is given by Eq. (8). We then have:

(a) The limit $\bar{D} = \lim_{k\to\infty} D_k \cdots D_1$ exists.

(b) The limit $\bar{D}$ is a stochastic matrix with identical rows i.e.,

$$\bar{D} = e\phi'$$

for a stochastic vector $\phi \in \mathbb{R}^m$.

(c) The convergence of $D_k \cdots D_1$ to $\bar{D}$ is geometric, i.e., for all $k \geq 1$ and $u \in \mathbb{R}^m$,

$$\left\|D_k \cdots D_1 u - \bar{D}u\right\|_\infty \leq 2\left(1 + \eta^{-\bar{B}}\right)\left(1 - \eta^{\bar{B}}\right)^k \|u\|_\infty,$$

where η is the lower bound of Assumption 1(a). In particular, for every j, the entries $[D_k \cdots D_1]^j_i$, $i = 1, \ldots, m$, converge to the same limit ϕ_j as $k \to \infty$ with a geometric rate, i.e., for every j,

$$\left|[D_k \cdots D_1]^j_i - \phi_j\right| \leq 2\left(1 + \eta^{-\bar{B}}\right)\left(1 - \eta^{\bar{B}}\right)^k.$$

The explicit form of the bound in part (c) of Lemma 1 is new; the other parts have been established by Tsitsiklis [14].

In the following lemma, we present convergence results for the matrices $\Phi(k,s)$ as k goes to infinity. Lemma 1 plays a crucial role in establishing these results. In particular, we show that the matrices $\Phi(k,s)$ have the same limit as the matrices $[D_1 \cdots D_k]'$, when k increases to infinity (see [12] for the proof).

Lemma 2: Let Weights Rule, Connectivity, and Bounded Intercommunication Interval assumptions hold [cf. Assumptions 1, 2, and 3]. We then have:

(a) The limit $\bar{\Phi}(s) = \lim_{k\to\infty} \Phi(k,s)$ exists for each s.

(b) The limit matrix $\bar{\Phi}(s)$ has identical columns and the columns are stochastic i.e.,

$$\bar{\Phi}(s) = \phi(s)e',$$

[2] For Assumption 1(a) to hold, the agents need not be given the lower bound η for their weights $a^i_j(k)$. In particular, as a lower bound for the weights $a^i_j(k)$, $j = 1, \ldots, m$, each agent may use his/her own bound η_i, with $0 < \eta_i < 1$. In this case, Assumption 1(a) holds for $\eta = \min_{1\leq i\leq m} \eta_i$, and this common bound is used only in the analysis of the model.

where $\phi(s)$ is a stochastic vector for each s.

(c) For every i, the entries $[\Phi(k,s)]_j^i$, $j = 1, ..., m$, converge to the same limit $[\phi(s)]_i$ as $k \to \infty$ with a geometric rate, i.e., for every $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, m\}$, and $k \geq s$,

$$\left|[\Phi(k,s)]_i^j - [\phi(s)]_i\right| \leq 2\,\frac{1+\eta^{-\bar{B}}}{1-\eta^{\bar{B}}}\,\left(1-\eta^{\bar{B}}\right)^{\frac{k-s}{\bar{B}}},$$

where η is the lower bound of Assumption 1(a) and $\bar{B}$ is given by Eq. (8).

B. Limit Vectors $\phi(s)$

In our convergence analysis, it is crucial that the limit vectors $\phi(s)$ converge to a uniform distribution, i.e., $\lim_{s\to\infty} \phi(s) = \frac{1}{m}\,e$ for all s. A special case when this holds is the case when $\phi(s) = \frac{1}{m}\,e$ for all s. In the following, we show that each $\phi(s)$ corresponds to a uniform distribution when the matrices $A(k) = [a^1(k), \ldots, a^m(k)]$ formed from the weights $a^i(k)$, $i = 1, \ldots, m$, are doubly stochastic. We formally impose this assumption in the following.

Assumption 4: (Doubly Stochastic Weights) The weight matrix $A(k) = [a^1(k), \ldots, a^m(k)]$ is doubly stochastic for all k.

Under this and some additional assumptions, all the limit vectors $\phi(s)$ are the same and correspond to the uniform distribution $\frac{1}{m}\,e$. This is an immediate consequence of Lemma 2, as seen in the following.

Proposition 1: Let Weights Rule, Connectivity, Bounded Intercommunication Interval, and Doubly Stochastic Weights assumptions hold [cf. Assumptions 1, 2, 3, and 4]. We then have:

(a) The limit matrices $\Phi(s) = \lim_{k\to\infty} \Phi(k,s)$ are doubly stochastic and correspond to a uniform steady state distribution for all s, i.e.,

$$\Phi(s) = \frac{1}{m}\,ee' \qquad \text{for all } s.$$

(b) The entries $[\Phi(k,s)]_j^i$ converge to $\frac{1}{m}$ as $k \to \infty$ with a geometric rate uniformly with respect to i and j, i.e., for all $i, j \in \{1, \ldots, m\}$, and all k, s with $k \geq s$,

$$\left|[\Phi(k,s)]_i^j - \frac{1}{m}\right| \leq 2\,\frac{1+\eta^{-\bar{B}}}{1-\eta^{\bar{B}}}\,\left(1-\eta^{\bar{B}}\right)^{\frac{k-s}{\bar{B}}}.$$

The requirement that $A(k)$ are doubly stochastic for all k inherently dictates that the agents share the information about their weights and coordinate the choices of the weights when updating their estimates. In this scenario, we view the weights of the agents as being of two types: *planned weights* and *actual weights* they use in their updates. Specifically, let the weight $p_j^i(k) > 0$ be the weight that agent i plans to use at the update time t_{k+1} provided that an estimate $x^j(k)$ is received from agent j in the interval (t_k, t_{k+1}). If agent j communicates with agent i during the time interval (t_k, t_{k+1}), then these agents communicate to each other their estimates $x^j(k)$ and $x^i(k)$ as well as their planned weights $p_i^j(k)$ and $p_j^i(k)$. In the next update time t_{k+1}, the actual weight $a_j^i(k)$ that agent i assigns to the estimate $x^j(k)$ is a combination of the agent j planned weight $p_i^j(k)$ and the agent i planned weight $p_j^i(k)$. We summarize this in the following assumptions.

Assumption 5: (Simultaneous Information Exchange) The agents exchange information simultaneously: if agent j communicates to agent i at some time, then agent i also communicates to agent j at that time, i.e.,

$$\text{if } (j,i) \in E_k \text{ for some } k, \text{ then } (i,j) \in E_k.$$

Furthermore, when agents i and j communicate, they exchange their estimates $x^i(k)$ and $x^j(k)$, and their planned weights $p_j^i(k)$ and $p_i^j(k)$.

Assumption 6: (Symmetric Weights) Let the agent planned weights $p_j^i(k)$, $i, j = 1\ldots, m$, be such that for some scalar η, with $0 < \eta < 1$, we have $p_j^i(k) \geq \eta$ for all i, j and k, and $\sum_{j=1}^m p_j^i(k) = 1$ for all i and k. Furthermore, let the actual weights $a_j^i(k)$, $i, j = 1, \ldots, m$ that agents use in their updates be given by

(i) $a_j^i(k) = \min\{p_j^i(k), p_i^j(k)\}$ when agents i and j communicate during the time interval (t_k, t_{k+1}), and $a_j^i(k) = 0$ otherwise.

(ii) $a_i^i(k) = 1 - \sum_{j\neq i} a_j^i(k)$.

The next proposition shows that when agents take the minimum of their planned weights and these weights are stochastic, then the actual weights form a doubly stochastic matrix.

Proposition 2: Let Connectivity, Bounded Intercommunication Interval, Simultaneous Information Exchange, and Symmetric Weights assumptions hold [cf. Assumptions 2, 3, 5, and 6]. We then have:

(a) The limit matrices $\Phi(s) = \lim_{k\to\infty} \Phi(k,s)$ are doubly stochastic and correspond to a uniform steady state distribution for all s, i.e.,

$$\Phi(s) = \frac{1}{m}\,ee' \qquad \text{for all } s.$$

(b) The entries $[\Phi(k,s)]_j^i$ converge to $\frac{1}{m}$ as $k \to \infty$ with a geometric rate uniformly with respect to i and j, i.e., for all $i, j \in \{1, \ldots, m\}$, and all k, s with $k \geq s$,

$$\left|[\Phi(k,s)]_i^j - \frac{1}{m}\right| \leq 2\,\frac{1+\eta^{-\bar{B}}}{1-\eta^{\bar{B}}}\,\left(1-\eta^{\bar{B}}\right)^{\frac{k-s}{\bar{B}}}.$$

IV. Convergence Analysis

In this section, we study the convergence behavior of the subgradient method introduced in Section II. In particular, we consider the subgradient method with a constant stepsize that is common to all agents, i.e., $\alpha^j(r) = \alpha$ for all r and all agents j. As seen in Section II, the iterates generated by the subgradient method satisfy Eq. (5), which for a constant step reduces to the following: for any $i \in \{1, \ldots, m\}$, and s and k with $k \geq s$,

$$x^i(k+1) \quad = \quad \sum_{j=1}^m [\Phi(k,s)]_j^i x^j(s)$$

$$- \quad \alpha \sum_{r=s+1}^{k} \left(\sum_{j=1}^{m} [\Phi(k,r)]_j^i d_j(r-1) \right) - \alpha d_i(k).$$

To analyze this model, we consider a related "stopped" model whereby the agents stop computing the subgradients $d_j(k)$ at some time, but they keep exchanging their information and updating their estimates using the weights only for the rest of the time. To describe the "stopped" model, we use $s = 0$ in the preceding relation and obtain

$$x^i(k+1) = \sum_{j=1}^{m} [\Phi(k,0)]_j^i x^j(0) - \alpha \sum_{s=1}^{k} \left(\sum_{j=1}^{m} [\Phi(k,s)]_j^i d_j(s-1) \right) - \alpha d_i(k). \tag{10}$$

Suppose that agents cease computing $d_j(k)$ after some time $t_{\bar{k}}$, so that

$$d^j(k) = 0 \qquad \text{for all } j \text{ and all } k \text{ with } k \geq \bar{k}.$$

Let $\{\bar{x}^i(k)\}$, $i = 1, \ldots, m$ be the sequences of the estimates generated by the agents in this case. Then, from relation (10) we have for all i,

$$\bar{x}^i(k) = x^i(k) \qquad \text{for all } k \leq \bar{k},$$

and for all $k > \bar{k}$,

$$\bar{x}^i(k) = \sum_{j=1}^{m} [\Phi(k-1,0)]_j^i x^j(0) - \alpha \sum_{s=1}^{\bar{k}} \left(\sum_{j=1}^{m} [\Phi(k-1,s)]_j^i d_j(s-1) \right).$$

By letting $k \to \infty$ and by using Proposition 2(b), we see that the limit vector $\lim_{k\to\infty} \bar{x}^i(k)$ exists. Furthermore, the limit vector does not depend on i, but does depend on $\bar{k}$. We denote this limit by $y(\bar{k})$, i.e.,

$$\lim_{k\to\infty} \bar{x}^i(k) = y(\bar{k}),$$

for which, by Proposition 2(a), we have

$$y(\bar{k}) = \frac{1}{m} \sum_{j=1}^{m} x^j(0) - \alpha \sum_{s=1}^{\bar{k}} \left(\sum_{j=1}^{m} \frac{1}{m} d_j(s-1) \right).$$

We re-index these relations by using k, and thus obtain

$$y(k+1) = y(k) - \frac{\alpha}{m} \sum_{j=1}^{m} d_j(k) \qquad \text{for all } k. \tag{11}$$

The vector $d_j(k)$ is a subgradient of the agent j objective function $f_j(x)$ at $x = x^j(k)$, so that the preceding iteration can be viewed as an iteration of an approximate subgradient method. Specifically, for each j, the method uses a subgradient of f_j at the estimate $x^j(k)$ approximating the vector $y(k)$ [instead of a subgradient of $f_j(x)$ at $x = y(k)$].

We start with a lemma providing some basic relations used in the analysis of subgradient methods (see Nedić and Bertsekas [9], [10], and Nedić et al. [11]). Recall our notation $f(x) = \sum_{i=1}^{m} f_i(x)$.

Lemma 3: Let the sequence $\{y(k)\}$ be generated by the iteration (11), and let the sequences $\{x^j(k)\}$ be generated by the iteration (10). Let $\{g_j(k)\}$ be a sequence of subgradients such that $g_j(k) \in \partial f_j(y(k))$ for all j and k. We then have:

(a) For any $x \in \mathbb{R}^n$ and all $k \geq 0$,

$$\|y(k+1) - x\|^2 \leq \|y(k) - x\|^2 + \frac{2\alpha}{m} \sum_{j=1}^{m} \left(\|d_j(k)\| + \|g_j(k)\| \right) \|y(k) - x^j(k)\| - \frac{2\alpha}{m} [f(y(k)) - f(x)] + \frac{\alpha^2}{m^2} \sum_{j=1}^{m} \|d_j(k)\|^2.$$

(b) When the optimal solution set X^* is nonempty, there holds for all $k \geq 0$,

$$dist^2(y(k+1), X^*) \leq dist^2(y(k), X^*) + \frac{2\alpha}{m} \sum_{j=1}^{m} (\|d_j(k)\| + \|g_j(k)\|) \|y(k) - x^j(k)\| - \frac{2\alpha}{m} [f(y(k)) - f^*] + \frac{\alpha^2}{m^2} \sum_{j=1}^{m} \|d_j(k)\|^2.$$

In our convergence analysis, we assume that the subgradients of the agent functions f_j are uniformly bounded. Specifically, we assume that there is a scalar L such that

$$\max_{j,k} \{\|d_j(k)\|, \|g_j(k)\|\} \leq L, \tag{12}$$

where $g_j(k)$ is a subgradient of $f_j(y(k))$ for all j and k. This assumption is standard in the convergence analysis of subgradient methods; it is satisfied, for example, when each f_i is polyhedral (i.e., f_i is the pointwise maximum of a finite number of affine functions).

The next proposition contains our main convergence result. The proof is omitted due to space constraints and it can be found in [12].

Proposition 3: Let Connectivity, Bounded Intercommunication Interval, Simultaneous Information Exchange, and Symmetric Weights assumptions hold [cf. Assumptions 2, 3, 5, and 6]. Let the subgradients be bounded in sense of Eq. (12), and assume that the optimal set X^* of problem (1) is nonempty. Let $x^j(0)$ denote the initial vector of agent j and assume that

$$\max_{1\leq j\leq m} \|x^j(0)\| \leq \alpha L.$$

Let the sequence $\{y(k)\}$ be generated by the iteration (11), and let the sequences $\{x^i(k)\}$, $i = 1, \ldots, m$, be generated by the iteration (10). We then have:

(a) A uniform upper bound on $\|y(k) - x^i(k)\|$: for every $i \in \{1, \ldots, m\}$ and for all $k \geq 0$,

$$\|y(k) - x^i(k)\| \leq 2\alpha L C_1,$$

$$C_1 = 1 + \frac{m}{1 - (1 - \eta^{\bar{B}})^{\frac{1}{\bar{B}}}} \, \frac{1 + \eta^{-\bar{B}}}{1 - \eta^{\bar{B}}}.$$

(b) An upper bound on the objective values $f(y(k))$: for all $K \geq K^*$,

$$\min_{0\leq k\leq K} f(y(k)) \leq f^* + \alpha L^2 C.$$

When there are subgradients $g_{ij}(k)$ of f_j at $x^i(k)$ that are bounded uniformly by some constant L_1, an upper bound on the objective values $f(x^i(k))$ is given by: for all $K \geq K^*$,

$$\min_{0\leq k\leq K} f(x^i(k)) \leq f^* + \alpha L^2 C + 2\alpha m L_1 L C_1 \quad \text{for all } i.$$

(c) Let $\hat{y}(k)$ be the average of $y(0), \ldots, y(k-1)$ for all $k \geq 1$, and similarly for each $i \in \{1, \ldots, m\}$, let $\hat{x}^i(k)$ be the average of $\hat{x}^i(0), \ldots, \hat{x}^i(k-1)$. An upper bound on the objective cost $f(\hat{y}(k))$ is given by: for all $K \geq K^*$,

$$f(\hat{y}(K)) \leq f^* + \alpha L^2 C.$$

When there are subgradients $\hat{g}_{ij}(k)$ of f_j at $\hat{x}^i(k)$ that are bounded uniformly by some constant $\hat{L}_1$, an upper bound on the objective values $f(\hat{x}^i(k))$ is given by: for all $K \geq K^*$,

$$f(\hat{x}^i(K)) \leq f^* + \alpha L^2 C + 2\alpha m \hat{L}_1 L C_1 \quad \text{for all } i,$$

where L is the subgradient norm bound of Eq. (12) and K^* is the number of iterations given by

$$K^* = \left\lceil \frac{m\, dist^2(y(0), X^*)}{\alpha^2 L^2 C} \right\rceil,$$

$y(0) = \frac{1}{m}\sum_{j=1}^m x^j(0)$ and $C = 1 + 8mC_1$.

Note that the number K^* of iterations of the preceding proposition is inversely proportional to the stepsize α. By part (a) of the proposition, it follows that the error between $y(k)$ and $x^i(k)$ for all i is bounded from above by a constant that is proportional to the stepsize α. This dependence captures the tradeoff between the quality of an approximate solution and the computation load required to generate such a solution.

V. Conclusions

We analyzed the convergence and rate of convergence properties of a distributed computation model for optimizing the sum of objective functions of multiple agents using local information. The model involves distributed computations and local message exchanges. Our analysis explicitly characterizes the tradeoff between the accuracy of the approximate optimal solutions generated and the number of iterations needed. One area that is left for future work is the incorporation of constraints into the computation model, which is the focus of our current research.

References

[1] Kashyap A., Basar T., and Srikant R., "Quantized Consensus," *Proceedings of CDC*, 2006.

[2] Bertsekas D. P., Nedić A., and Ozdaglar A., Convex Analysis and Optimization, Athena Scientific, Belmont, MA, 2003.

[3] Bertsekas D. P. and Tsitsiklis J. N., Parallel and Distributed Computation: Numerical Methods, Athena Scientific, Belmont, MA, 1997.

[4] Blondel V. D., Hendrickx, J. M., Olshevsky A., and Tsitsiklis J. N., "Convergence in Multiagent Coordination, Consensus, and Flocking," *Proceedings of CDC*, 2005.

[5] Boyd S., Ghosh A., Prabhakar B., and Shah D., "Gossip Algorithms: Design, Analysis, and Applications," *Proceedings of Infocom*, 2005.

[6] Jadbabaie A., Lin J. and Morse S., "Coordination of Groups of Mobile Autonomous Agents using Nearest Neighbor Rules," *IEEE Transactions on Automatic Control,* vol. 48, no. 6, pp. 988–1001, 2003.

[7] Kelly, F. P., Maulloo A. K., and Tan D. K.,"Rate control for communication networks: shadow prices, proportional fairness, and stability," *Jour. of the Operational Research Society*, vol. 49, pp. 237-252, 1998.

[8] Low, S. and Lapsley, D. E., "Optimization flow control, I: Basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7(6), pp. 861-874, 1999.

[9] Nedić A. and Bertsekas D. P. "Incremental Subgradient Methods for Nondifferentiable Optimization," *SIAM J. on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[10] Nedić A. and Bertsekas D. P. "Convergence rate of incremental subgradient algorithms", Stochastic Optimization: Algorithms and Applications Eds. Uryasev S. and Pardalos P., Kluwer Academic Publishers, 2000.

[11] Nedić A., Bertsekas D. P., and Borkar V. S. "Distributed Asynchronous Incremental Subgradient Methods", Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, Eds. Butnariu D., Censor Y., and Reich S., Studies Comput. Math., Elsevier, 2001.

[12] Nedić A. and Ozdaglar A., "Distributed Subgradient Methods for Multi-agent Optimization," Laboratory for Information and Decision Systems, *LIDS Technical Report 2755*, Massachusetts Institute of Technology, 2007.

[13] Srikant R., *Mathematics of Internet Congestion Control*, Birkhauser, 2004.

[14] Tsitsiklis J. N., "Problems in Decentralized Decision Making and Computation" Ph.D. Dissertation, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1984.

[15] Tsitsiklis J. N., Bertsekas D. P., and Athans M, "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[16] Vicsek T., Czirok A., Ben-Jacob E., Cohen I., and Schochet O., "Novel Type of Phase Transitions in a System of Self-driven Particles," *Physical Review Letters*, vol. 75, no. 6, pp. 1226–1229, 1995.

[17] Wolfowitz J., "Products of Indecomposable, Aperiodic, Stochastic Matrices," *Proceedings of the American Mathematical Society*, vol. 14, no. 4, pp. 733–737, Oct. 1963.