Distributed Subgradient Methods for Multi-agent Optimization

Angelia Nedić* and Asuman Ozdaglar†

August 16, 2007

Abstract

We study a distributed computation model for optimizing a sum of convex objective functions corresponding to multiple agents. For solving this (not necessarily smooth) optimization problem, we consider a subgradient method that is distributed among the agents. In this model, each agent minimizes his/her own objective while exchanging information directly or indirectly with other agents in the network. We allow such communication to be asynchronous, local, and with time varying connectivity. We provide convergence results and convergence rate estimates for the subgradient method. Our convergence rate results explicitly characterize the tradeoff between a desired accuracy of the generated approximate optimal solutions and the number of iterations needed to achieve the accuracy.

1 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 resource allocation methods that can operate in a decentralized manner with limited local information and rapidly converge to an approximately optimal operating point. Most existing approaches use cooperative and noncooperative distributed optimization frameworks under the assumption that each agent has an objective function that depends only on the resource allocated to that agent. In many practical situations however, individual cost functions or performance measures depend on the entire resource allocation vector.

In this paper, we study a distributed computation model that can be used for general resource allocation problems. We provide a simple algorithm and study its convergence and rate of convergence properties. In particular, we focus on the distributed control of

*Department of Industrial and Enterprise Systems Engineering, University of Illinois, angelia@uiuc.edu

†Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, asuman@mit.edu

a network consisting of m agents. The global objective is to cooperatively minimize the cost function $\sum_{i=1}^{m} f_i(x)$, where the function $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ represents the cost function of agent i, known by this agent only, and $x \in \mathbb{R}^n$ is a decision vector. The decision vector can be viewed as either a resource vector where sub-components correspond to resources allocated to each agent, or a global decision vector upon which the agents are attempting to reach a *consensus*.

Our algorithm is as follows: each agent generates and maintains estimates of the optimal decision vector based on information concerning his own cost function (in particular, the subgradient information of f_i) and exchanges these estimates directly or indirectly with the other agents in the network. We allow such communication to be asynchronous, local, and with time varying connectivity. Under some weak assumptions, we prove that this type of local communication and computation achieves consensus in the estimates and converges to an (approximate) global optimal solution. Moreover, we provide rate of convergence analysis for the estimates maintained by each agent. 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.

Our approach builds on the seminal work of Tsitsiklis [23] (see also Tsitsiklis *et al.* [24], Bertsekas and Tsitsiklis [2]), who developed a framework for the analysis of distributed computation models. This framework focuses on the minimization of a (smooth) function $f(x)$ by distributing the processing of the components of vector $x \in \mathbb{R}^n$ among n agents. In contrast, our focus is on problems in which each agent has a locally known, different, convex and potentially nonsmooth cost function. To the best of our knowledge, this paper presents the first analysis of this type of distributed resource allocation problem.

In addition our work is also related to the literature on reaching consensus on a particular scalar value or computing exact averages of the initial values of the agents, which has attracted much recent attention as natural models of cooperative behavior in networked-systems (see Vicsek *et al.* [25], Jadbabaie *et al.* [8], Boyd *et al.* [4], Olfati-Saber and Murray [17], Cao *et al.* [5], and Olshevsky and Tsitsiklis [18, 19]). Our work is also related to the *utility maximization* framework for resource allocation in networks (see Kelly *et al.* [10], Low and Lapsley [11], Srikant [22], and Chiang *et al.* [7]). In contrast to this literature, we allow the local performance measures to depend on the entire resource allocation vector and provide explicit rate analysis.

The remainder of this paper is organized as follows: In Section 2, we introduce a model for the information exchange among the agents and a distributed optimization model. In Section 3, we establish the basic convergence properties of the transition matrices governing the evolution of the system in time. In Section 4, we establish convergence and convergence rate for the proposed distributed optimization algorithm. Finally, we give our concluding remarks in Section 5.

Basic Notation and Notions:

A vector is viewed as a column, unless clearly stated otherwise. We denote by x_i or $[x]_i$ the i-th component of a vector x. When $x_i \geq 0$ for all components i of a vector x, we write $x \geq 0$. 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.

We denote the nonnegative orthant by $\mathbb{R}_+^n$, i.e., $\mathbb{R}_+^n = \{x \in \mathbb{R}^n \mid x \geq 0\}$. We write x' to denote the transpose of a vector x. The scalar product of two vectors $x, y \in \mathbb{R}^m$ is denoted by $x'y$. We use $\|x\|$ to denote the standard Euclidean norm, $\|x\| = \sqrt{x'x}$. We write $\|x\|_\infty$ to denote the max norm, $\|x\|_\infty = \max_{1 \leq i \leq m} |x_i|$.

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 their sum is equal to 1, i.e., $\sum_{i=1}^m a_i = 1$. 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 a *doubly stochastic* when both A and A' are stochastic matrices.

For a function $F : \mathbb{R}^n \mapsto (-\infty, \infty]$, we denote the domain of F by $\mathrm{dom}(F)$, where

$$\mathrm{dom}(F) = \{x \in \mathbb{R}^n \mid F(x) < \infty\}.$$

We use the notion of a subgradient of a *convex* function $F(x)$ at a given vector $\bar{x} \in \mathrm{dom}(F)$. We say that $s_F(\bar{x}) \in \mathbb{R}^n$ *is a subgradient of the function F at $\bar{x} \in \mathrm{dom}(F)$* when the following relation holds:

$$F(\bar{x}) + s_F(\bar{x})'(x - \bar{x}) \leq F(x) \qquad \text{for all } x \in \mathrm{dom}(F). \tag{1}$$

The set of all subgradients of F at $\bar{x}$ is denoted by $\partial F(\bar{x})$ (see [1]).

2 Multi-agent Model

We are interested in a distributed computation model for a multi-agent system, where each agent processes his/her local information and shares the information with his/her neighbors. To describe such a multi-agent system, we need to specify two models: an *information exchange model* describing the evolution of the agents' information in time and an *optimization model* specifying overall system objective that agents are cooperatively minimizing by individually minimizing their own local objectives. Informally speaking, the first model specifies the rules of the agents' interactions such as how often the agents communicate, how they value their own information and the information received from the neighbors. The second model describes the overall goal that the agents want to achieve through their information exchange. The models are discussed in the following sections.

2.1 Information Exchange Model

We consider a network with node (or agent) set $V = \{1, \ldots, m\}$. At this point, we are not describing what the agents' goal is, but rather what the agents' information exchange process is.

We use an information evolution model based on the model proposed by Tsitsiklis [23] (see also Blondel *et al.* [3]). We impose some rules that govern the information evolution of the agent system in time. These rules include:

- A rule on the weights that an agent uses when combining his information with the information received from his/her neighbors.

- A connectivity rule ensuring that the information of each agent influences the information of any other agent infinitely often in time.

- A rule on the frequency at which an agent sends his information to the neighbors.

We assume that the agents update and (possibly) send their information to their neighbors at discrete times $t_0, t_1, t_2, \ldots$. The neighbors of an agent i are the agents j communicating directly with agent i through a directed link (j, i). We index agents' information states and any other information at time t_k by k. We use $x^i(k) \in \mathbb{R}^n$ to denote agent i information state at time t_k.

We now describe a rule that agents use when updating their information states $x^i(k)$. The information update for agents includes combining their own information state with those received from their neighbors. In particular, we assume that each agent i has a vector of weights $a^i(k) \in \mathbb{R}^m$ at any time t_k. For each j, the scalar $a^i_j(k)$ is the weight that agent i assigns to the information x^j obtained from a neighboring agent j, when the information is received during the time interval (t_k, t_{k+1}). We use the following assumption on these weights.

Assumption 1 *(Weights Rule)* We have:

(a) There exists a scalar η with $0 < \eta < 1$ such that for all $i \in \{1, \ldots, m\}$,

  (i) $a^i_i(k) \geq \eta$ for all $k \geq 0$.

  (ii) $a^i_j(k) \geq \eta$ for all $k \geq 0$, and all agents j communicating directly with agent i in the interval (t_k, t_{k+1}).

  (iii) $a^i_j(k) = 0$ for all $k \geq 0$ 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 her own state $x^i(k)$ and the state information $x^j(k)$ available from her neighboring agents j at the update time t_k. Naturally, an agent i assigns zero weights to the states x^j for those agents j whose state information is not available at the update time.[1] Note that, under Assumption 1, for a matrix $A(k)$ whose columns are $a^1(k), \ldots, a^m(k)$, the transpose $A'(k)$ is a stochastic matrix for all $k \geq 0$.

The following are some examples of weight choices satisfying Assumption 1 (cf. Blondel *et al.* [3]):

[1]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 positive weights $a^i_j(k)$, each agent may use her own η_i, with $0 < \eta_i < 1$. In this case, Assumption 1(a) holds for $\eta = \min_{1 \leq i \leq m} \eta_i$. Moreover, we do not assume that the common bound value η is available to any agent at any time.

(1) (*Fixed Set of Weights*) Each agent i selects a weight vector from a given finite set of stochastic vectors $\mathcal{B}_i = \{b^1, \ldots, b^K\}$, i.e., $a^i(k) \in \{b^1, \ldots, b^K\}$ for all i and k, where every $b \in \mathcal{B}_i$ is a stochastic vector with entries $b_l > 0$.

(3) (*Equal Neighbor Weights*) Each agent assigns equal weight to his/her state and the state information received from the other agents, i.e., $a^i_j(k) = \frac{1}{n_i(k)+1}$ for each i, k, and those neighbors j whose state information is available to agent i at time t_k; otherwise $a^i_j(k) = 0$. Here, the number $n_i(k)$ is the number of agents communicating with agent i at the given time t_k.

We now discuss the rules we impose on the information exchange. 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 E_k of directed edges given by

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

Note that, by Assumption 1(a), we have $(i, i) \in E_k$ for each agent i and all k. Also, we have $(j, i) \in E_k$ if and only if agent i receives the information x^j from agent j in the time interval (t_k, t_{k+1}).

Motivated by the model of Tsitsiklis [23] and the "consensus" setting of Blondel *et al.* [3]), we impose a minimal assumption on the connectivity of a multi-agent system as follows: 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). In other words, the information state of any agent i has to influence the information state of any other agent infinitely often in time. In formulating this, we use the set E_∞ consisting of edges (j, i) such that j is a neighbor of i who communicates with i infinitely often. The connectivity requirement is formally stated in the following assumption.

Assumption 2 *(Connectivity)* The graph (V, E_∞) is connected, where E_∞ is the set of edges (j, i) representing agent pairs communicating directly infinitely many times, i.e.,

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

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 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.

When analyzing the system state behavior, we use an additional assumption that the intercommunication intervals are bounded for those agents that communicate directly. In particular, we use the following.

Assumption 3 *(Bounded Intercommunication Interval)* There exists an integer $B \geq 1$ such that for every $(j, i) \in E_\infty$, agent j sends his/her information to the neighboring agent i at least once every B consecutive time slots, i.e., at time t_k or at time t_{k+1} or $\ldots$ or (at latest) at time t_{k+B-1} for any $k \geq 0$.

This assumption is equivalent to the requirement that there is $B \geq 1$ such that

$$(j, i) \in E_k \cup E_{k+1} \cup \cdots \cup E_{k+B-1} \qquad \text{for all } (j, i) \in E_\infty \text{ and } k \geq 0.$$

2.2 Optimization Model

We consider a scenario where agents cooperatively minimize a common additive cost. Each agent has information only about one cost component, and minimizes that component while exchanging information with other agents. In particular, the agents want to cooperatively solve the following unconstrained optimization problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m f_i(x) \\ \text{subject to} & x \in \mathbb{R}^n, \end{array} \tag{2}$$

where each $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function. We denote the optimal value of this problem by f^*, which we assume to be finite. We also denote the optimal solution set by X^*, i.e., $X^* = \{x \in \mathbb{R}^n \mid \sum_{i=1}^m f_i(x) = f^*\}$.

In this setting, the information state of an agent i is an estimate of an optimal solution of the problem (2). We denote by $x^i(k) \in \mathbb{R}^n$ the estimate maintained by agent i at the time t_k. The agents update their estimates as follows: When generating a new estimate, agent i combines his/her current estimate x^i with the estimates x^j received from some of the other agents j. Here we assume that there is no communication delay in delivering a message from agent j to agent i [2].

In particular, agent i updates his/her estimates according to the following relation:

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

where the vector $a^i(k) = (a_1^i(k), \ldots, a_m^i(k))'$ 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 agent i objective function $f_i(x)$ at $x = x^i(k)$. We note that the optimization model of Eqs. (2)–(3) reduces to a "consensus" or "agreement" problem when all the objective functions f_i are identically equal to zero; see Jadbabaie *et al.* [8] and Blondel *et al.* [3].

We are interested in conditions guaranteeing convergence of $x^i(k)$ to a common limit vector in $\mathbb{R}^n$. We are also interested in characterizing these common limit points in terms of the properties of the functions f_i. In order to have a more compact representation of the evolution of the estimates $x^i(k)$ of Eq. (3) in time, we rewrite this model in a form similar to that of Tsitsiklis [23]. This form is also more appropriate for our convergence analysis. 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 $k \geq s$,

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

[2] A more general model that accounts for the possibility of such delays is the subject of our current work, see [16].

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

Let us introduce the matrices

$$\Phi(k,s) = A(s)A(s+1)\cdots A(k-1)A(k) \qquad \text{for all } s \text{ and } k \text{ with } k \geq s,$$

where $\Phi(k,k) = A(k)$ for all k. Note that the i-th column of $\Phi(k,s)$ is given by

$$[\Phi(k,s)]^i = A(s)A(s+1)\cdots A(k-1)a^i(k) \qquad \text{for all } i,\ s \text{ and } k \text{ with } k \geq s,$$

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

$$[\Phi(k,s)]^i_j = [A(s)A(s+1)\cdots A(k-1)a^i(k)]_j \qquad \text{for all } i,\ j,\ s \text{ and } k \text{ with } k \geq s.$$

We can now rewrite relation (4) compactly in terms of the matrices $\Phi(k,s)$, as follows: for any $i \in \{1,\ldots,m\}$, and s and k with $k \geq s \geq 0$,

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

We start our analysis by considering the transition matrices $\Phi(k,s)$.

3 Convergence of the Transition Matrices $\Phi(k,s)$

In this section, we study the convergence behavior of the matrices $\Phi(k,s)$ as k goes to infinity. We establish convergence rate estimates for these matrices. Clearly, the convergence rate of these matrices dictates the convergence rate of the agents' estimates to an optimal solution of the overall optimization problem (2). Recall that these matrices are given by

$$\Phi(k,s) = A(s)A(s+1)\cdots A(k-1)A(k) \qquad \text{for all } s \text{ and } k \text{ with } k \geq s, \tag{6}$$

where

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

3.1 Basic Properties

Here, we establish some properties of the transition matrices $\Phi(k,s)$ under the assumptions discussed in Section 2.

Lemma 1 Let Weights Rule (a) hold [cf. Assumption 1(a)]. We then have:

(a) $[\Phi(k,s)]_j^j \geq \eta^{k-s+1}$ for all j, k, and s with $k \geq s$.

(b) $[\Phi(k,s)]_j^i \geq \eta^{k-s+1}$ for all k and s with $k \geq s$, and for all $(j,i) \in E_s \cup \cdots \cup E_k$, where E_t is the set of edges defined by

$$E_t = \{(j,i) \mid a_j^i(t) > 0\} \qquad \text{for all } t.$$

(c) Let $(j,v) \in E_s \cup \cdots \cup E_r$ for some $r \geq s$ and $(v,i) \in E_{r+1} \cup \cdots \cup E_k$ for $k > r$. Then,

$$[\Phi(k,s)]_j^i \geq \eta^{k-s+1}.$$

(d) Let Weights Rule (b) also hold [cf. Assumption 1(b)]. Then, the matrices $\Phi(k,s)$ are stochastic for all k and s with $k \geq s$.

Proof. We let s be arbitrary, and prove the relations by induction on k.
(a) Note that, in view of Assumption 1(a), the matrices $\Phi(k,s)$ have nonnegative entries for all k and s with $k \geq s$. Furthermore, by Assumption 1(a)(i), we have $[\Phi(s,s)]_j^j \geq \eta$. Thus, the relation $[\Phi(k,s)]_j^j \geq \eta^{k-s+1}$ holds for $k = s$.

Now, assume that for some k with $k > s$ we have $[\Phi(k,s)]_j^j \geq \eta^{k-s+1}$, and consider $[\Phi(k+1,s)]_j^j$. By the definition of the matrix $\Phi(k,s)$ [cf. Eq. (6)], we have

$$[\Phi(k+1,s)]_j^j = \sum_{h=1}^{m} [\Phi(k,s)]_j^h a_h^j(k+1) \geq [\Phi(k,s)]_j^j a_j^j(k+1),$$

where the inequality in the preceding relation follows from the nonnegativity of the entries of $\Phi(k,s)$. By using the inductive hypothesis and the relation $a_j^j(k+1) \geq \eta$ [cf. Assumption 1(a)(i)], we obtain

$$[\Phi(k+1,s)]_j^j \geq \eta^{k-s+2}.$$

Hence, the relation $[\Phi(k,s)]_j^j \geq \eta^{k-s+1}$ holds for all $k \geq s$.
(b) Let $(j,i) \in E_s$. Then, by the definition of E_s and Assumption 1(a), we have that $a_j^i(s) \geq \eta$. Since $\Phi(s,s) = A(s)$ [cf. Eq. (7)], it follows that the relation $[\Phi(k,s)]_j^i \geq \eta^{k-s+1}$ holds for $k = s$ and any $(j,i) \in E_s$. Assume now that for some $k > s$ and all $(j,i) \in E_s \cup \cdots \cup E_k$, we have $[\Phi(k,s)]_j^i \geq \eta^{k-s+1}$. Consider $k+1$, and let $(j,i) \in E_s \cup \cdots \cup E_k \cup E_{k+1}$. There are two possibilities $(j,i) \in E_s \cup \cdots \cup E_k$ or $(j,i) \in E_{k+1}$.

Suppose that $(j,i) \in E_s \cup \cdots \cup E_k$. Then, by the induction hypothesis, we have

$$[\Phi(k,s)]_j^i \geq \eta^{k-s+1}.$$

Therefore

$$[\Phi(k+1,s)]_j^i = \sum_{h=1}^{m} [\Phi(k,s)]_j^h a_h^i(k+1) \geq [\Phi(k,s)]_j^i a_i^i(k+1),$$

where the inequality in the preceding relation follows from the nonnegativity of the entries of $\Phi(k,s)$. By combining the preceding two relations, and using the fact $a(r)^i_i \geq \eta$ for all i and r [cf. Assumption 1(a)(i)], we obtain

$$[\Phi(k+1,s)]^i_j \geq \eta^{k-s+2}.$$

Suppose now that $(j,i) \in E_{k+1}$. Then, by the definition of E_{k+1}, we have $a^i_j(k+1) \geq \eta$. Furthermore, by part (a), we have

$$[\Phi(k,s)]^j_j \geq \eta^{k-s+1}.$$

Therefore

$$[\Phi(k+1,s)]^i_j = \sum_{h=1}^m [\Phi(k,s)]^h_j a^i_h(k+1) \geq [\Phi(k,s)]^j_j a^i_j(k+1) \geq \eta^{k-s+2}.$$

Hence, $[\Phi(k,s)]^i_j \geq \eta^{k-s+1}$ holds for all $k \geq s$ and all $(j,i) \in E_s \cup \cdots \cup E_k$.
(c) Let $(j,v) \in E_s \cup \cdots \cup E_r$ for some $r \geq s$ and $(v,i) \in E_{r+1} \cup \cdots \cup E_k$ for $k > r$. Then, by the nonnegativity of the entries of $\Phi(r,s)$ and $\Phi(k,r+1)$, we have

$$[\Phi(k,s)]^i_j = \sum_{h=1}^m [\Phi(r,s)]^h_j [\Phi(k,r+1)]^i_h \geq [\Phi(r,s)]^v_j [\Phi(k,r+1)]^i_v.$$

By part (b), we further have

$$[\Phi(r,s)]^v_j \geq \eta^{r-s+1} \qquad [\Phi(k,r+1)]^i_v \geq \eta^{k-r},$$

implying that

$$[\Phi(k,s)]^i_j \geq \eta^{r-s+1}\eta^{k-r} = \eta^{k-s+1}.$$

(d) Recall that, for each k, the columns of the matrix $A(k)$ are the weight vectors $a^1(k),\ldots,a^m(k)$. Hence, by Assumption 1, the matrix $A'(k)$ is stochastic for all k. From the definition of $\Phi(k,s)$ in Eqs. (6)–(7), we have $\Phi'(k,s) = A'(k)\cdots A'(s+1)A'(s)$, thus implying that $\Phi'(k,s)$ is stochastic for all k and s with $k \geq s$. ∎

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

$$[\Phi(s+(m-1)B-1,s)]^i_j \geq \eta^{(m-1)B} \qquad \text{for all } s,\ i, \text{ and } j,$$

where η is the lower bound of Assumption 1(a) and B is the intercommunication interval bound of Assumption 3.

Proof. Let s, i, and j be arbitrary. If $j = i$, then by Lemma 1(a), we have

$$[\Phi(s+(m-1)B-1,s)]^i_i \geq \eta^{(m-1)B}.$$

Assume now that $j \neq i$. By Connectivity [cf. Assumption 2], there is a path $j = i_0 \to i_1 \to \ldots \to i_{r-1} \to i_r = i$ from j to i, passing through distinct nodes i_κ, $\kappa = 0, \ldots, r$ and with edges $(i_{\kappa-1}, i_\kappa)$ in the set

$$E_\infty = \{(h, \bar{h}) \mid (h, \bar{h}) \in E_k \text{ for infinitely many indices } k\}.$$

Because each edge $(i_{\kappa-1}, i_\kappa)$ of the path belongs to E_∞, by using Assumption 3 [with $k = s + (\kappa - 1)B$ for edge $(i_{\kappa-1}, i_\kappa)$], we obtain

$$(i_{\kappa-1}, i_\kappa) \in E_{s+(\kappa-1)B} \cup \cdots \cup E_{s+\kappa B-1} \qquad \text{for } \kappa = 1, ..., r.$$

By using Lemma 1 (b), we have

$$[\Phi(s + \kappa B - 1, s + (\kappa - 1)B)]_{i_{\kappa-1}}^{i_\kappa} \geq \eta^B \qquad \text{for } \kappa = 1, ..., r.$$

By Lemma 1(c), it follows that

$$[\Phi(s + rB - 1, s)]_j^i \geq \eta^{rB}.$$

Since there are m agents, and the nodes in the path $j = i_0 \to i_1 \to \ldots \to i_{r-1} \to i_r = i$ are distinct, it follows that $r \leq m - 1$. Hence, we have

$$\begin{aligned}
[\Phi(s + (m-1)B - 1, s)]_j^i &= \sum_{h=1}^{m} [\Phi(s + rB - 1, s)]_j^h [\Phi(s + (m-1)B - 1, s + rB)]_h^i \\
&\geq [\Phi(s + rB - 1, s)]_j^i [\Phi(s + (m-1)B - 1, s + rB)]_i^i \\
&\geq \eta^{rB} \eta^{(m-1)B - rB} \\
&= \eta^{(m-1)B},
\end{aligned}$$

where the last inequality follows from $[\Phi(k, s]_i^i \geq \eta^{k-s+1}$ for all i, k, and s with $k \geq s$ [cf. Lemma 1(a)]. ■

Our ultimate goal is to study the limit behavior of $\Phi(k, s)$ as $k \to \infty$ for a fixed $s \geq 0$. For this analysis, we introduce the matrices $D_k(s)$ as follows: for a fixed $s \geq 0$,

$$D_k(s) = \Phi'(s + kB_0 - 1, s + (k-1)B_0) \qquad \text{for } k = 1, 2, \ldots, \tag{8}$$

where $B_0 = (m - 1)B$. We show that, for each $s \geq 0$, the product of these matrices converges as k increases to infinity. In particular, we have the following result.

Lemma 3 Let Weights Rule, Connectivity, and Bounded Intercommunication Interval assumptions hold [cf. Assumptions 1, 2, and 3]. Let the matrices $D_k(s)$ for $k \geq 1$ and a fixed $s \geq 0$ be given by Eq. (8). We then have:

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

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

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

where $e \in \mathbb{R}^m$ is a vector of ones and $\phi(s) \in \mathbb{R}^m$ is a stochastic vector.

(c) The convergence of $D_k(s)\cdots D_1(s)$ to $\bar{D}(s)$ is geometric: for every $x \in \mathbb{R}^m$,

$$\left\|(D_k(s)\cdots D_1(s))\,x - \bar{D}(s)x\right\|_\infty \leq 2\left(1+\eta^{-B_0}\right)\left(1-\eta^{B_0}\right)^k \|x\|_\infty \quad \text{for all } k \geq 1.$$

In particular, for every j, the entries $[D_k(s)\cdots D_1(s)]_i^j$, $i = 1,\ldots,m$, converge to the same limit $\phi_j(s)$ as $k \to \infty$ with a geometric rate: for every j,

$$\left|[D_k(s)\cdots D_1(s)]_i^j - \phi_j(s)\right| \leq 2\left(1+\eta^{-B_0}\right)\left(1-\eta^{B_0}\right)^k \qquad \text{for all } k \geq 1 \text{ and } i,$$

where η is the lower bound of Assumption 1(a), $B_0 = (m-1)B$, m is the number of agents, and B is the intercommunication interval bound of Assumption 3.

Proof. In this proof, we suppress the explicit dependence of the matrices D_i on s to simplify our notation.

(a) We prove that the limit $\lim_{k\to\infty}(D_k\cdots D_1)$ exists by showing that the sequence $\{(D_k\cdots D_1)x\}$ converges for every $x \in \mathbb{R}^m$. To show this, let $x_0 \in \mathbb{R}^m$ be arbitrary, and consider the vector sequence $\{x_k\}$ defined by

$$x_k = D_k\cdots D_1 x_0 \qquad \text{for } k \geq 1.$$

Each vector x_k we write as follows:

$$x_k = z_k + c_k e, \tag{9}$$

where c_k is the minimum component of x_k, i.e,

$$c_k = \min_{1\leq i\leq m}[x_k]_i \qquad \text{for all } k. \tag{10}$$

Therefore,

$$z_k = x_k - c_k e \geq 0 \qquad \text{for all } k, \tag{11}$$

$$[z_k]_i = 0 \quad \text{ for at least one } i \in \{1,\ldots,m\}. \tag{12}$$

We now consider x_{k+1}. Note that, for each k, the matrix D_k is stochastic since it is a product of stochastic matrices [cf. Eq. (8) and Lemma 1(d)]. By the decomposition rule of Eq. (9) and the stochasticity of D_{k+1} [i.e., $D_{k+1}e = e$], we have

$$x_{k+1} = D_{k+1}x_k = D_{k+1}z_k + c_k e.$$

From the definition of z_k in Eq. (11), it follows that for each component $[z_{k+1}]_j$ we have

$$[z_{k+1}]_j = [D_{k+1}]_j z_k - [D_{k+1}]_{j^*}z_k, \qquad [z_{k+1}]_j \geq 0,$$

$$c_{k+1} = [D_{k+1}]_{j^*}z_k + c_k, \tag{13}$$

where $[D_{k+1}]_j$ is the j-th row vector of the matrix D_{k+1} and j^* is the index of the row vector $[D_{k+1}]_j$ achieving the minimum in $\min_{1\leq j\leq m}[D_{k+1}]_j z_k$. Since the vectors

$[D_{k+1}]_j$ and z_k have nonnegative entries, it follows that $[D_{k+1}]_{j^*} z_k \geq 0$. Therefore, for all $j = 1, \ldots, m$,

$$[z_{k+1}]_j \leq [D_{k+1}]_j z_k = \sum_{i=1}^{m} [D_{k+1}]_j^i [z_k]_i \leq \sum_{i:[z_k]_i > 0} [D_{k+1}]_j^i \max_{i:[z_k]_i > 0} [z_k]_i,$$

where $[D_{k+1}]_j^i$ denotes the (j,i)-th entry of the matrix D_{k+1}. Because z_k has nonnegative entries $[z_k]_i$ and at least one of them is zero [cf. Eq. (12)], say $[z_k]_{i^*} = 0$, it follows that

$$[z_{k+1}]_j \leq \left(\sum_{i \neq i^*} [D_{k+1}]_j^i \right) \|z_k\|_\infty = \left(1 - [D_{k+1}]_j^{i^*}\right) \|z_k\|_\infty, \qquad \text{for all } j, \tag{14}$$

where the last equality in the preceding relation follows from the stochasticity of the matrix D_{k+1} (i.e., $\sum_{i=1}^m [D_{k+1}]_j^i = 1$ for each row j). By the definition of D_k [cf. Eq. (8)], we have that $[D_{k+1}]_j^{i^*}$ is the entry in the i^*-th row and j-th column of the matrix $\Phi\left(s + kB_0, s + (k-1)B_0\right)$, i.e.,

$$[D_{k+1}]_j^{i^*} = [\Phi\left(s + kB_0, s + (k-1)B_0\right)]_{i^*}^j .$$

Furthermore, by Lemma 2 it follows that

$$[\Phi\left(s + kB_0, s + (k-1)B_0\right)]_{i^*}^j \geq \eta^{B_0} \qquad \text{for any } j,$$

with $B_0 = (m-1)B$. Therefore, from the preceding and relation (14) we obtain

$$\|z_{k+1}\|_\infty \leq \left(1 - \eta^{B_0}\right) \|z_k\|_\infty \qquad \text{for all } k,$$

implying that

$$\|z_k\|_\infty \leq \left(1 - \eta^{B_0}\right)^k \|z_0\|_\infty \qquad \text{for all } k. \tag{15}$$

Hence $z_k \to 0$ with a geometric rate.

Consider now the sequence c_k of Eq. (13). Since the vectors $[D_{k+1}]_j$ and z_k have nonnegative entries, it follows that

$$c_k \leq c_{k+1} = c_k + [D_{k+1}]_{j^*} z_k.$$

Furthermore, by using the stochasticity of the matrix D_{k+1}, we obtain for all k,

$$c_{k+1} \leq c_k + \sum_{i=1}^{m} [D_{k+1}]_{j^*}^i \|z_k\|_\infty = c_k + \|z_k\|_\infty.$$

From the preceding two relations and Eq. (15), it follows that

$$0 \leq c_{k+1} - c_k \leq \|z_k\|_\infty \leq \left(1 - \eta^{B_0}\right)^k \|z_0\|_\infty \qquad \text{for all } k.$$

Therefore, we have for any $k \geq 1$ and $r \geq 1$,

$$0 \leq c_{k+r} - c_k \leq c_{k+r} - c_{k+r-1} + \cdots + c_{k+1} - c_k \leq (q^{k+r-1} + \cdots + q^k)\|z_0\|_\infty = \frac{1-q^r}{1-q}\, q^k \|z_0\|_\infty,$$

where $q = 1 - \eta^{B_0}$. Hence, $\{c_k\}$ is a Cauchy sequence and therefore it converges to some $\bar{c} \in \mathbb{R}$. By letting $r \to \infty$ in the preceding relation, we obtain

$$0 \leq \bar{c} - c_k \leq \frac{q^k}{1-q} \|z_0\|_\infty. \tag{16}$$

From the decomposition of x_k [cf. Eq. (9)], and the relations $z_k \to 0$ and $c_k \to \bar{c}$, it follows that $(D_k \cdots D_1)x_0 \to \bar{c}e$ for any $x_0 \in \mathbb{R}^m$, with $\bar{c}$ being a function of x_0. Therefore, the limit of $D_k \cdots D_1$ as $k \to \infty$ exists. We denote this limit by $\bar{D}$, for which we have

$$\bar{D}x_0 = \bar{c}(x_0)e \qquad \text{for all } x_0 \in \mathbb{R}^m.$$

(b) Since each D_k is stochastic, the limit matrix $\bar{D}$ is also stochastic. Furthermore, because $(D_k \cdots D_1)x \to \bar{c}(x)e$ for any $x \in \mathbb{R}^m$, the limit matrix $\bar{D}$ has rank one. Thus, the rows of $\bar{D}$ are collinear. Because the sum of all entries of $\bar{D}$ in each of its rows is equal to 1, it follows that the rows of $\bar{D}$ are identical. Therefore, for some stochastic vector $\phi(s) \in \mathbb{R}^m$ [a function of the fixed s], we have

$$\bar{D} = e\phi(s)'.$$

(c) Let $x_k = (D_k \cdots D_1)x_0$ for an arbitrary $x_0 \in \mathbb{R}^m$. By omitting the explicit dependence on x_0 in $\bar{c}(x_0)$, and by using the decomposition of x_k [cf. Eq. (9)], we have

$$(D_k \cdots D_1)x_0 - \bar{D}x_0 = z_k + (c_k - \bar{c})e \qquad \text{for all } k.$$

Using the estimates in Eqs. (15) and (16), we obtain for all $k \geq 1$,

$$\left\|(D_k \cdots D_1)x_0 - \bar{D}x_0\right\|_\infty \leq \|z_k\|_\infty + |c_k - \bar{c}| \leq \left(1 + \frac{1}{1-q}\right) q^k \|z_0\|_\infty.$$

In view of the decomposition relations (10) and (11), we have $z_0 = x_0 - \min_{1 \leq i \leq m}[x_0]_i$, implying that $\|z_0\|_\infty \leq 2\|x_0\|_\infty$. Therefore,

$$\left\|(D_k \cdots D_1)x_0 - \bar{D}x_0\right\|_\infty \leq 2\left(1 + \frac{1}{1-q}\right) q^k \|x_0\|_\infty \qquad \text{for all } k,$$

with $q = 1 - \eta^{B_0}$, or equivalently

$$\left\|(D_k \cdots D_1)x_0 - \bar{D}x_0\right\|_\infty \leq 2\left(1 + \eta^{-B_0}\right)\left(1 - \eta^{B_0}\right)^k \|x_0\|_\infty \qquad \text{for all } k. \tag{17}$$

Thus, the first relation of part (c) of the lemma is established.

To show the second relation of part (c) of the lemma, let $j \in \{1, \ldots, m\}$ be arbitrary. Let $e_j \in \mathbb{R}^m$ be a vector with j-th entry equal to 1 and the other entries equal to 0. By setting $x_0 = e_j$ in Eq. (17), and by using $\bar{D} = e\phi'(s)$ and $\|e_j\|_\infty = 1$, we obtain

$$\left\|[D_k \cdots D_1]^j - \phi_j(s)e\right\|_\infty \leq 2\left(1 + \eta^{-B_0}\right)\left(1 - \eta^{B_0}\right)^k \qquad \text{for all } k.$$

Thus, it follows that

$$\left|[D_k \cdots D_1]_i^j - \phi_j(s)\right| \leq 2\left(1 + \eta^{-B_0}\right)\left(1 - \eta^{B_0}\right)^k \qquad \text{for all } k \geq 1 \text{ and } i.$$

■

In the following lemma, we present convergence results for the matrices $\Phi(k,s)$ as k goes to infinity. Lemma 3 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 $[D_1(s)\cdots D_k(s)]'$, when k increases to infinity.

Lemma 4 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.,

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

where $\phi(s) \in \mathbb{R}^m$ is a stochastic vector for each s.

(c) For every i, the entries $[\Phi(k,s)]_i^j$, $j = 1,...,m$, converge to the same limit $\phi_i(s)$ as $k \to \infty$ with a geometric rate, i.e., for every $i \in \{1,\ldots,m\}$ and all $s \geq 0$,

$$\left|[\Phi(k,s)]_i^j - \phi_i(s)\right| \leq 2\,\frac{1+\eta^{-B_0}}{1-\eta^{B_0}}\ \left(1-\eta^{B_0}\right)^{\frac{k-s}{B_0}} \quad \text{for all } k \geq s \text{ and } j \in \{1,\ldots,m\},$$

where η is the lower bound of Assumption 1(a), $B_0 = (m-1)B$, m is the number of agents, and B is the intercommunication interval bound of Assumption 3.

Proof. For a given s and $k \geq s + B_0$, there exists $\kappa \geq 1$ such that $s + \kappa B_0 \leq k < s + (\kappa+1)B_0$. Then, by the definition of $\Phi(k,s)$ [cf. Eqs. (6)-(7)], we have

$$\begin{aligned}
\Phi(k,s) &= \Phi\left(s+\kappa B_0 - 1, s\right)\Phi\left(k, s+\kappa B_0\right)\\
&= \Phi(s+B_0-1,s)\cdots\Phi\left(s+\kappa B_0 - 1, s+(\kappa-1)B_0\right)\Phi\left(k, s+\kappa B_0\right)\\
&= \left(\Phi'\left(s+\kappa B_0-1, s+(\kappa-1)B_0\right)\cdots\Phi'(s+B_0-1,s)\right)'\Phi\left(k,s+\kappa B_0\right).
\end{aligned}$$

By using the matrices

$$D_k = \Phi'\left(s+kB_0-1, s+(k-1)B_0\right) \qquad \text{for } k \geq 1 \tag{18}$$

[the dependence of D_k on s is suppressed], we can write

$$\Phi(k,s) = (D_\kappa\cdots D_1)'\Phi\left(k, s+\kappa B_0\right).$$

Therefore, for any i, j and $k \geq s + B_0$, we have

$$[\Phi(k,s)]_i^j = \sum_{h=1}^m [D_\kappa\cdots D_1]_h^i\left[\Phi\left(k,s+\kappa B_0\right)\right]_h^j \leq \max_{1\leq h\leq m}[D_\kappa\cdots D_1]_h^i \sum_{h=1}^m\left[\Phi\left(k,s+\kappa B_0\right)\right]_h^j.$$

Since the columns of the matrix $\Phi\left(k, s+\kappa B_0\right)$ are stochastic vectors, it follows that for any i, j and $k \geq s+B_0$,

$$[\Phi(k, s)]_i^j \leq \max_{1 \leq h \leq m} [D_\kappa \cdots D_1]_h^i. \tag{19}$$

Similarly, it can be seen that for any i, j and $k \geq s+B_0$,

$$[\Phi(k, s)]_i^j \geq \min_{1 \leq h \leq m} [D_\kappa \cdots D_1]_h^i. \tag{20}$$

In view of Lemma 3, for a given s, there exists a stochastic vector $\phi(s)$ such that

$$\lim_{k \to \infty} D_k \cdots D_1 = e\phi'(s).$$

Furthermore, by Lemma 3(c) we have for every $\tilde{h} \in \{1, \ldots, m\}$,

$$\left|[D_\kappa \cdots D_1]_h^{\tilde{h}} - [\phi(s)]_{\tilde{h}}\right| \leq 2\left(1+\eta^{-B_0}\right)\left(1-\eta^{B_0}\right)^\kappa,$$

for $\kappa \geq 1$ and $h \in \{1, \ldots, m\}$. From the preceding relation, and inequalities (19) and (20), it follows that for $k \geq s+B_0$ and any $i, j \in \{1, \ldots, m\}$,

$$\begin{aligned}
&\left|[\Phi(k, s)]_i^j - [\phi(s)]_i\right| \\
&\quad \leq \max\left\{\left|\max_{1 \leq h \leq m} [D_\kappa \cdots D_1]_h^i - [\phi(s)]_i\right|, \left|\min_{1 \leq h \leq m} [D_\kappa \cdots D_1]_h^i - [\phi(s)]_i\right|\right\} \\
&\quad \leq 2\left(1+\eta^{-B_0}\right)\left(1-\eta^{B_0}\right)^\kappa.
\end{aligned}$$

Since $\kappa \geq 1$ and $s+\kappa B_0 \leq k < s+(\kappa+1)B_0$, we have

$$\begin{aligned}
\left(1-\eta^{B_0}\right)^\kappa &= \left(1-\eta^{B_0}\right)^{\kappa+1} \frac{1}{1-\eta^{B_0}} \\
&= \left(1-\eta^{B_0}\right)^{\frac{s+(\kappa+1)B_0-s}{B_0}} \frac{1}{1-\eta^{B_0}} \\
&\leq \left(1-\eta^{B_0}\right)^{\frac{k-s}{B_0}} \frac{1}{1-\eta^{B_0}},
\end{aligned}$$

where the last inequality follows from the relations $0 < 1-\eta^{B_0} < 1$ and $k < s+(\kappa+1)B_0$. By combining the preceding two relations, we obtain for $k \geq s+B_0$ and any $i, j \in \{1, \ldots, m\}$,

$$\left|[\Phi(k, s)]_i^j - [\phi(s)]_i\right| \leq 2\,\frac{1+\eta^{-B_0}}{1-\eta^{B_0}}\left(1-\eta^{B_0}\right)^{\frac{k-s}{B_0}}. \tag{21}$$

Therefore, we have

$$\lim_{k \to \infty} \Phi(k, s) = \phi(s)e' = \bar{\Phi}(s), \qquad \text{for every } s,$$

thus showing part (a) of the lemma. Furthermore, we have that all the columns of $\bar{\Phi}(s)$ coincide with the vector $\phi(s)$, which is a stochastic vector by Lemma 3(b). This shows part (b) of the lemma.

Note that relation (21) holds for $k \geq s + B_0$ and any $i, j \in \{1, \dots, m\}$. To prove part (c) of the lemma, we need to show that the estimate of Eq. (21) holds for arbitrary s and for k with $s + B_0 > k \geq s$, and any $i, j \in \{1, \dots, m\}$. Thus, let s be arbitrary and let $s + B_0 > k \geq s$. Because $\Phi'(k, s)$ is a stochastic matrix, we have for all i and j,

$$0 \leq [\Phi(k, s)]_i^j \leq 1.$$

Therefore, for k with $s + B_0 > k \geq s$, and any $i, j \in \{1, \dots, m\}$,

$$\begin{aligned}
\left| [\Phi(k,s)]_i^j - [\phi(s)]_i \right| &\leq 2 \\
&< 2\left(1 + \eta^{-B_0}\right) \\
&= 2\,\frac{1 + \eta^{-B_0}}{1 - \eta^{B_0}}\,\left(1 - \eta^{B_0}\right)^{\frac{s + B_0 - s}{B_0}} \\
&< 2\,\frac{1 + \eta^{-B_0}}{1 - \eta^{B_0}}\,\left(1 - \eta^{B_0}\right)^{\frac{k - s}{B_0}},
\end{aligned}$$

where the last inequality follows from the relations $0 < 1 - \eta^{B_0} < 1$ and $k < s + B_0$. From the preceding relation and Eq. (21) it follows that for every s and $i \in \{1, \dots, m\}$,

$$\left| [\Phi(k,s)]_i^j - [\phi(s)]_i \right| \leq 2\,\frac{1 + \eta^{-B_0}}{1 - \eta^{B_0}}\,\left(1 - \eta^{B_0}\right)^{\frac{k - s}{B_0}} \qquad \text{for all } k \geq s \text{ and } j \in \{1, \dots, m\},$$

thus showing part (c) of the lemma. ■

The preceding results are shown by following the line of analysis of Tsitsiklis [23] (see Lemma 5.2.1 in [23]; see also Bertsekas and Tsitsiklis [2]). The rate estimate given in part (c) is new and provides the explicit dependence of the convergence of the transition matrices on system parameters and problem data. This estimate will be essential in providing convergence rate results for the subgradient method of Section 3 [cf. Eq. (5)].

3.2 Limit Vectors $\phi(s)$

The agents' objective is to cooperatively minimize the additive cost function $\sum_{i=1}^m f_i(x)$, while each agent individually performs his own state updates according to the subgradient method of Eq. (5). In order to reach a "consensus" on the optimal solution of the problem, it is essential that the agents process their individual objective functions with the same frequency in the long run. This amounts to having 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. One way of ensuring this is to have $\phi(s) = \frac{1}{m}\,e$ for all s, which holds when the weight matrices $A(k)$ are doubly stochastic. We formally impose this condition in the following.

Assumption 4 *(Doubly Stochastic Weights)* Let the weight vectors $a^1(k), \dots, a^m(k)$, $k = 0, 1, \dots$, satisfy Weights Rule [cf. Assumption 1]. Assume further that the matrices $A(k) = [a^1(k), \dots, a^m(k)]$ are doubly stochastic for all k.

Under this and some additional assumptions, we show that 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 4, as seen in the following.

Proposition 1 *(Uniform Limit Distribution)* Let Connectivity, Bounded Intercommunication Interval, and Doubly Stochastic Weights assumptions hold [cf. Assumptions 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 $[\Phi(k, s)]_i^j$ converge to $\frac{1}{m}$ as $k \to \infty$ with a geometric rate uniformly with respect to i and j, i.e., for all $i, j \in \{1, \ldots, m\}$,

$$\left| [\Phi(k, s)]_i^j - \frac{1}{m} \right| \leq 2\, \frac{1 + \eta^{-B_0}}{1 - \eta^{B_0}} \, \left(1 - \eta^{B_0}\right)^{\frac{k-s}{B_0}} \quad \text{for all } s \text{ and } k \text{ with } k \geq s,$$

where η is the lower bound of Assumption 1(a), $B_0 = (m - 1)B$, m is the number of agents, and B is the intercommunication interval bound of Assumption 3.

Proof. (a) Since the matrix $A(k)$ is doubly stochastic for all k, the matrix $\Phi(k, s)$ [cf. Eqs. (6)-(7)] is also doubly stochastic for all s and k with $k \geq s$. In view of Lemma 4, the limit matrix $\bar{\Phi}(s) = \phi(s)e'$ is doubly stochastic for every s. Therefore, we have $\phi(s)e'e = e$ for all s, implying that

$$\phi(s) = \frac{1}{m}\, e \qquad \text{for all } s.$$

(b) The geometric rate estimate follows directly from Lemma 4(c). ∎

The requirement that $A(k)$ is 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 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 assumption.

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 $p_i^j(k)$.

We now show that when agents choose the smallest of their planned weights and the planned weights are stochastic, then the actual weights form a doubly stochastic matrix.

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

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

(ii) $a^i_i(k) = 1 - \sum_{j \neq i} a^i_j(k)$.

The preceding discussion, combined with Lemma 1, yields the following result.

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

$$\left| [\Phi(k, s)]^j_i - \frac{1}{m} \right| \leq 2\, \frac{1 + \eta^{-B_0}}{1 - \eta^{B_0}} \; \left(1 - \eta^{B_0}\right)^{\frac{k-s}{B_0}} \qquad \text{for all } s \text{ and } k \text{ with } k \geq s,$$

where η is the lower bound of Symmetric Weights Assumption 6, $B_0 = (m - 1)B$, m is the number of agents, and B is the intercommunication interval bound of Assumption 3.

Proof. In view of Uniform Limit Distribution [cf. Proposition 1], it suffices to show that Simultaneous Information Exchange and Symmetric Weights assumptions [cf. Assumptions 5 and 6], imply that Assumption 4 holds. In particular, we need to show that the actual weights $a^i_j(k)$, $i, j = 1, \ldots, m$ satisfy Weights Rule [cf. Assumption 1], and that the vectors $a^i(k)$, $i = 1, \ldots, m$ form a doubly stochastic matrix.

First, note that by Symmetric Weights assumption, the weights $a^i_j(k)$, $i, j = 1, \ldots, m$ satisfy Weights Rule. Thus, the agent weight vectors $a^i(k)$, $i = 1, \ldots, m$ are stochastic, and hence, the weight matrix $A'(k)$ with rows $a^i(k)$, $i = 1, \ldots, m$ is stochastic for all k. Second, note that by Simultaneous Information Exchange and Symmetric Weights assumptions, it follows that the weight matrix $A(k)$ is symmetric for all k. Since $A'(k)$ is stochastic for any k, we have that $A(k)$ is doubly stochastic for all k. ■

4 Convergence of the Subgradient Method

Here, we study the convergence behavior of the subgradient method introduced in Section 2. In particular, we have shown that the iterations of the method satisfy the following relation: for any $i \in \{1, \ldots, m\}$, and s and k with $k \geq s$,

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

[cf. Eq. (5)]. We analyze this model under the symmetric weights assumption (cf. Assumption 6). Also, we consider the case of a constant stepsize that is common to all agents, i.e., $\alpha^j(r) = \alpha$ for all r and all agents j, so that the model reduces to the following: for any $i \in \{1, \ldots, m\}$, and s and k with $k \geq s$,

$$x^i(k+1) = \sum_{j=1}^{m} [\Phi(k,s)]_j^i x^j(s) - \alpha \sum_{r=s+1}^{k} \left(\sum_{j=1}^{m} [\Phi(k,r)]_j^i d_j(r-1) \right) - \alpha d_i(k). \tag{22}$$

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 only the weights for the rest of the time. To describe the "stopped" model, we use relation (22) with $s = 0$, from which we obtain

$$x^i(k+1) = \sum_{j=1}^{m} [\Phi(k,0)]_j^i x^j(0) - \alpha \sum_{s=1}^{k} \left(\sum_{j=1}^{m} [\Phi(k,s)]_j^i d_j(s-1) \right) - \alpha d_i(k). \tag{23}$$

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 (23) we have for all i,

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

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

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

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

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

Note that this relation holds for any $\bar{k}$, so may re-index these relations by using k, and thus obtain

$$y(k+1) = y(k) - \frac{\alpha}{m}\sum_{j=1}^{m} d_j(k) \qquad \text{for all } k. \tag{24}$$

Recall that the vector $d_j(k)$ is a subgradient of the agent j objective function $f_j(x)$ at $x = x^j(k)$. Thus, 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. Similar relations have been used in various ways to analyze subgradient approaches (for example, see Shor [21], Polyak [20], Nedić and Bertsekas [12], [13], and Nedić, Bertsekas, and Borkar [14]). In the following lemma and thereafter, we use notation $f(x) = \sum_{i=1}^{m} f_i(x)$.

Lemma 5 *(Basic Iterate Relation)* Let the sequence $\{y(k)\}$ be generated by the iteration (24), and the sequences $\{x^j(k)\}$ for $j \in \{1,\ldots,m\}$ be generated by the iteration (23). Let $\{g_j(k)\}$ be a sequence of subgradients such that $g_j(k) \in \partial f_j(y(k))$ for all $j \in \{1,\ldots,m\}$ and $k \geq 0$. We then have:

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

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

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

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

Proof. From relation (24) we obtain for any $x \in \mathbb{R}^n$ and all $k \geq 0$,

$$\|y(k+1) - x\|^2 = \left\| y(k) - \frac{\alpha}{m} \sum_{j=1}^{m} d_j(k) - x \right\|^2,$$

implying that

$$\|y(k+1) - x\|^2 \leq \|y(k) - x\|^2 - \frac{2\alpha}{m} \sum_{j=1}^{m} d_j(k)'(y(k) - x) + \frac{\alpha^2}{m^2} \sum_{j=1}^{m} \|d_j(k)\|^2. \tag{25}$$

We now consider the term $d_j(k)'(y(k) - x)$ for any j, for which we have

$$\begin{aligned} d_j(k)'(y(k) - x) &= d_j(k)'(y(k) - x^j(k)) + d_j(k)'(x^j(k) - x) \\ &\geq -\|d_j(k)\| \, \|y(k) - x^j(k)\| + d_j(k)'(x^j(k) - x). \end{aligned}$$

Since $d_j(k)$ is a subgradient of f_j at $x^j(k)$ [cf. Eq. (1)], we further have for any j and any $x \in \mathbb{R}^n$,

$$d_j(k)'(x^j(k) - x) \geq f_j(x^j(k)) - f_j(x).$$

Furthermore, by using a subgradient $g_j(k)$ of f_j at $y(k)$ [cf. Eq. (1)], we also have for any j and $x \in \mathbb{R}^n$,

$$\begin{aligned} f_j(x^j(k)) - f_j(x) &= f_j(x^j(k)) - f_j(y(k)) \; + \; f_j(y(k)) - f_j(x) \\ &\geq g_j(k)'(x^j(k) - y(k)) \; + \; f_j(y(k)) - f_j(x) \\ &\geq -\|g_j(k)\| \, \|x^j(k) - y(k)\| \; + \; f_j(y(k)) - f_j(x). \end{aligned}$$

By combining the preceding three relations it follows that for any j and $x \in \mathbb{R}^n$,

$$d_j(k)'(y(k) - x) \geq -\left(\|d_j(k)\| + \|g_j(k)\|\right) \|y(k) - x^j(k)\| + f_j(y(k)) - f_j(x).$$

Summing this relation over all j, we obtain

$$\sum_{j=1}^{m} d_j(k)'(y(k) - x) \geq -\sum_{j=1}^{m} \left(\|d_j(k)\| + \|g_j(k)\|\right) \|y(k) - x^j(k)\| + f(y(k)) - f(x).$$

By combining the preceding inequality with Eq. (25) the relation in part (a) follows, i.e., for all $x \in \mathbb{R}^n$ and all $k \geq 0$,

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

The relation in part (b) follows by letting $x \in X^*$ and by taking the infimum over $x \in X^*$ in both sides of the preceding relation. ∎

We adopt the following assumptions for our convergence analysis:

Assumption 7 *(Bounded Subgradients)* The subgradient sequences $\{d_j(k)\}$ and $\{g_j(k)\}$ are bounded for each j, i.e., there exists a scalar $L > 0$ such that

$$\max\{\|d_j(k)\|, \|g_j(k)\|\} \leq L \qquad \text{for all } j = 1, \dots, m, \text{ and all } k \geq 0.$$

This assumption 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).

Assumption 8 *(Nonempty Optimal Solution Set)* The optimal solution set X^* is nonempty.

Our main convergence results are given in the following proposition. In particular, we provide a uniform bound on the norm of the difference between $y(k)$ and $x^i(k)$ that holds for all $i \in \{1, \dots, m\}$ and all $k \geq 0$. We also consider the averaged-vectors $\hat{y}(k)$ and $\hat{x}^i(k)$ defined for all $k \geq 1$ as follows:

$$\hat{y}(k) = \frac{1}{k} \sum_{h=0}^{k-1} y(h), \qquad \hat{x}^i(k) = \frac{1}{k} \sum_{h=0}^{k-1} x^i(h) \qquad \text{for all } i \in \{1, \dots, m\}. \tag{26}$$

We provide upper bounds on the objective function value of the averaged-vectors. Note that averaging allows us to provide our estimates per iteration[3].

Proposition 3 Let Connectivity, Bounded Intercommunication Interval, Simultaneous Information Exchange, and Symmetric Weights assumptions hold [cf. Assumptions 2, 3, 5, and 6]. Let the Bounded Subgradients and Nonempty Optimal Set assumptions hold [cf. Assumptions 7 and 8]. 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 (24), and let the sequences $\{x^i(k)\}$ be generated by the iteration (23). We then have:

(a) For every $i \in \{1, \dots, m\}$, a uniform upper bound on $\|y(k) - x^i(k)\|$ is given by:

$$\|y(k) - x^i(k)\| \leq 2\alpha L C_1 \qquad \text{for all } k \geq 0,$$

$$C_1 = 1 + \frac{m}{1 - (1 - \eta^{B_0})^{\frac{1}{B_0}}} \, \frac{1 + \eta^{-B_0}}{1 - \eta^{B_0}}.$$

[3] See also our recent work [15] which uses averaging to generate approximate primal solutions with convergence rate estimates for dual subgradient methods.

(b) Let $\hat{y}(k)$ and $\hat{x}^i(k)$ be the averaged-vectors of Eq. (26). An upper bound on the objective cost $f(\hat{y}(k))$ is given by:

$$f(\hat{y}(k)) \leq f^* + \frac{m\ dist^2(y(0), X^*)}{2\alpha k} + \frac{\alpha L^2 C}{2} \qquad \text{for all } k \geq 1.$$

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 value $f(\hat{x}^i(k))$ for each i is given by:

$$f(\hat{x}^i(k)) \leq f^* + \frac{m\ dist^2(y(0), X^*)}{2\alpha k} + \frac{\alpha L^2 C}{2} + 2\alpha m \hat{L}_1 L C_1 \qquad \text{for all } k \geq 1,$$

where L is the subgradient norm bound of Assumption 7, $y(0) = \frac{1}{m}\sum_{j=1}^m x^j(0)$, and $C = 1 + 8mC_1$. The constant B_0 is given by $B_0 = (m-1)B$ and B is the intercommunication interval bound of Assumption 3.

Proof. (a) From Eq. (24) it follows that $y(k) = y(0) - \frac{\alpha}{m}\sum_{s=0}^{k-1}\sum_{j=1}^m d_j(s)$ for all $k \geq 1$. Using this relation, the relation $y(0) = \frac{1}{m}\sum_{j=1}^m x^j(0)$ and Eq. (23), we obtain for all $k \geq 0$ and $i \in \{1, \ldots, m\}$,

$$\begin{aligned} \|y(k) - x^i(k)\| \ &\leq \ \Big\| \sum_{j=1}^m x^j(0)\Big(\frac{1}{m} - [\Phi(k-1,0)]_j^i\Big) \\ &- \ \alpha \sum_{s=1}^{k-1}\sum_{j=1}^m \Big(\frac{1}{m} - [\Phi(k-1,s)]_j^i\Big) d_j(s-1) \\ &- \ \alpha\Big(\frac{1}{m}\sum_{j=1}^m d_j(k-1) - d_i(k-1)\Big)\Big\|. \end{aligned}$$

Therefore, for all $k \geq 0$ and $i \in \{1, \ldots, m\}$,

$$\begin{aligned} \|y(k) - x^i(k)\| \ &\leq \ \max_{1\leq j\leq m}\|x^j(0)\| \sum_{j=1}^m \left|\frac{1}{m} - [\Phi(k-1,0)]_j^i\right| \\ &+ \ \alpha \sum_{s=1}^{k-1}\sum_{j=1}^m \|d_j(s-1)\| \left|\frac{1}{m} - [\Phi(k-1,s)]_j^i\right| \\ &+ \ \frac{\alpha}{m}\sum_{j=1}^m \|d_j(k-1) - d_i(k-1)\|. \end{aligned}$$

Using the estimates for $\left|[\Phi(k-1,0)]_j^i - \frac{1}{m}\right|$ of Proposition 2(b), the assumption that $\max_{1\leq j\leq m}\|x^j(0)\| \leq \alpha L$, and the subgradient boundedness [cf. Assumption 7], from the preceding relation we obtain for all $k \geq 0$ and $i \in \{1, \ldots, m\}$,

$$\begin{aligned} \|y(k) - x^i(k)\| \ &\leq \ 2\alpha L m \,\frac{1+\eta^{-B_0}}{1-\eta^{B_0}} \sum_{s=0}^{k-1}(1-\eta^{B_0})^{\frac{k-1-s}{B_0}} + 2\alpha L \\ &\leq \ 2\alpha L\left(1 + \frac{m}{1-(1-\eta^{B_0})^{\frac{1}{B_0}}}\,\frac{1+\eta^{-B_0}}{1-\eta^{B_0}}\right). \end{aligned}$$

(b) By using Lemma 5(b) and the subgradient boundedness [cf. Assumption 7], we have for all $k \geq 0$,

$$dist^2(y(k+1), X^*) \leq dist^2(y(k), X^*) + \frac{4\alpha L}{m} \sum_{j=1}^{m} \|y(k) - x^j(k)\| - \frac{2\alpha}{m} [f(y(k)) - f^*] + \frac{\alpha^2 L^2}{m}.$$

Using the estimate of part (a), we obtain for all $k \geq 0$,

$$\begin{aligned} dist^2(y(k+1), X^*) &\leq dist^2(y(k), X^*) + \frac{4\alpha L}{m} 2m\alpha L C_1 - \frac{2\alpha}{m}[f(y(k)) - f^*] + \frac{\alpha^2 L^2}{m} \\ &= dist^2(y(k), X^*) + \frac{\alpha^2 L^2}{m} C - \frac{2\alpha}{m}[f(y(k)) - f^*], \end{aligned}$$

where $C = 1 + 8mC_1$. Therefore, we have

$$f(y(k)) - f^* \leq \frac{dist^2(y(k), X^*) - dist^2(y(k+1), X^*)}{2\alpha/m} + \frac{\alpha L^2 C}{2} \qquad \text{for all } k \geq 0.$$

By summing preceding relations over $0, \ldots, k-1$ and dividing the sum by k, we have for any $k \geq 1$,

$$\begin{aligned} \frac{1}{k} \sum_{k=0}^{k-1} f(y(h)) - f^* &\leq \frac{dist^2(y(0), X^*) - dist^2(y(h), X^*)}{2\alpha/m} + \frac{\alpha L^2 C}{2} \\ &\leq \frac{dist^2(y(0), X^*)}{2\alpha k/m} + \frac{\alpha L^2 C}{2} \end{aligned} \tag{27}$$

By the convexity of the function f, we have

$$\frac{1}{k} \sum_{k=0}^{k-1} f(y(h)) \geq f(\hat{y}(k)) \quad \text{where } \hat{y}(k) = \frac{1}{k} \sum_{k=0}^{k-1} y(h).$$

Therefore, by using the relation in (27), we obtain

$$f(\hat{y}(k)) \leq f^* + \frac{m\, dist^2(y(0), X^*)}{2\alpha k} + \frac{\alpha L^2 C}{2} \qquad \text{for all } k \geq 1. \tag{28}$$

We now show the estimate for $f(\hat{x}^i(k))$. By the subgradient definition, we have

$$f(\hat{x}^i(k)) \leq f(\hat{y}(k)) + \sum_{j=1}^{m} \hat{g}_{ij}(k)'(\hat{x}^i(k) - \hat{y}(k)) \qquad \text{for all } i \in \{1, \ldots, m\} \text{ and } k \geq 1,$$

where $\hat{g}_{ij}(k)$ is a subgradient of f_j at $\hat{x}^i(k)$. Since $\|\hat{g}_{ij}(k)\| \leq \hat{L}_1$ for all $i,\ j \in \{1, \ldots, m\}$, and $k \geq 1$, it follows that

$$f(\hat{x}^i(k)) \leq f(\hat{y}(k)) + m\hat{L}_1 \|\hat{x}^i(k) - \hat{y}(k)\|.$$

Using the estimate in part (a), the relation $\|\hat{x}^i(k) - \hat{y}(k)\| \leq \sum_{l=0}^{k-1} \|x^i(l) - y(l)\|/k$, and Eq. (28), we obtain for all $i \in \{1, \ldots, m\}$ and $k \geq 1$,

$$f(\hat{x}^i(k)) \leq f^* + \frac{m\ dist^2(y(0), X^*)}{2\alpha k} + \frac{\alpha L^2 C}{2} + 2\alpha m \hat{L}_1 L C_1.$$

■

Part (a) of the preceding proposition shows 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, i.e., by picking a smaller stepsize in the subgradient method, one can guarantee a smaller error between the vectors $y(k)$ and $x^i(k)$ for all $i \in \{1, \ldots, m\}$ and all $k \geq 0$. In part (b) of the proposition, we provide upper bounds on the objective function values of the averaged-vectors $\hat{y}(k)$ and $\hat{x}^i(k)$. The upper bounds on $f(\hat{x}^i(k))$ provide estimates for the error from the optimal value f^* at each iteration k. More importantly, they show that the error consists of two additive terms: The first term is inversely proportional to the stepsize α and goes to zero at a rate $1/k$. The second term is a constant that is proportional to the stepsize α, the subgradient bound L, and the constants C and C_1, which are related to the convergence of the transition matrices $\Phi(k, s)$. Hence, our bounds provide explicit per-iteration error expressions for the estimates maintained at each agent i.

The fact that there is a constant error term in the estimates which is proportional to the stepsize value α is not surprising and is due to the fact that the constant stepsize rule is used in the subgradient method of Eq. (22). It is possible to use different stepsize rules (e.g. diminishing stepsize rule or adaptive stepsize rules; see [1], [13], and [12]) to drive the error to zero in the limit. We use constant stepsize rule in view of its simplicity and since our goal is to generate approximate optimal solutions in relatively few number of iterations. Our analysis explicitly characterizes the tradeoff between the quality of an approximate solution and the computation load required to generate such a solution in terms of the stepsize value α.

5 Conclusions

In this paper, we presented an analysis of a distributed computation model for optimizing the sum of objective functions of multiple agents, which are convex but not necessarily smooth. In this model, every agent generates and maintains estimates of the optimal solution of the global optimization problem. These estimates are communicated (directly or indirectly) to other agents asynchronously and over a time-varying connectivity structure. Each agent updates his estimates based on local information concerning the estimates received from his immediate neighbors and his own cost function using a subgradient method.

We provide convergence results for this method focusing on the objective function values of the estimates maintained at each agent. To achieve this, we first analyze the convergence behavior of the transition matrices governing the information exchange among the agents. We provide explicit rate results for the convergence of the transition matrices. We use these rate results in the analysis of the subgradient method. For the constant stepsize rule, we provide bounds on the error between the objective function values of the estimates at each agent and the optimal value of the global optimization problem. Our bounds are per-iteration and explicitly characterize the dependence of the error on the algorithm parameters and the underlying connectivity structure.

The results in this paper add to the growing literature on the cooperative control of multi-agent systems. The framework provided in this paper motivates further analysis of a number of interesting questions:

- Our model assumes that there are no constraints in the global optimization problem. One interesting area of research is to incorporate constraints into the distributed computation model. The presence of constraints may destroy the linearity in the information evolution and will necessitate a different line of analysis.

- The update rule studied in this paper assumes that there is no delay in receiving the estimates of the other agents. This is a restrictive assumption in many applications in view of communication and other types of delays. The convergence and convergence rate analysis of this paper can be extended to this more general case and is the focus of our current research [16].

- The update rule assumes that agents can send and process real-valued estimates, thus excluding the possibility of communication bandwidth constraints on the information exchange. This is a question that is attracting much recent attention in the context of consensus algorithms (see Kashyap *et al.* [9] and Carli *et al.* [6]). Understanding the implications of communication bandwidth constraints on the performance of the asynchronous distributed optimization algorithms both in terms of convergence rate and error is an important area for future study.

References

[1] D.P. Bertsekas, A. Nedić, and A.E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Cambridge, Massachusetts, 2003.

[2] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont, MA, 1997.

[3] V.D. Blondel, J.M. Hendrickx, A. Olshevsky, and J.N. Tsitsiklis, *Convergence in multiagent coordination, consensus, and flocking*, Proceedings of IEEE CDC, 2005.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Gossip algorithms: Design, analysis, and applications*, Proceedings of IEEE INFOCOM, 2005.

[5] M. Cao, D.A. Spielman, and A.S. Morse, *A lower bound on convergence of a distributed network consensus algorithm*, Proceedings of IEEE CDC, 2005.

[6] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in coordinated consensus problems*, Proceedings of IEEE ACC, 2006.

[7] M. Chiang, S.H. Low, A.R. Calderbank, and J.C. Doyle, *Layering as optimization decomposition: A mathematical theory of network architectures*, Proceedings of the IEEE **95** (2007), no. 1, 255–312.

[8] A. Jadbabaie, J. Lin, and S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Transactions on Automatic Control **48** (2003), no. 6, 988–1001.

[9] A. Kashyap, T. Basar, and R. Srikant, *Quantized consensus*, Proceedings of IEEE CDC, 2006.

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

[11] S. Low and D.E. Lapsley, *Optimization flow control, I: Basic algorithm and convergence*, IEEE/ACM Transactions on Networking **7** (1999), no. 6, 861–874.

[12] A. Nedić and D.P. Bertsekas, *Convergence rate of incremental subgradient algorithms*, Stochastic Optimization: Algorithms and Applications (S. Uryasev and P. Pardalos, eds.), Kluwer Academic Publishers, 2000.

[13] ______, *Incremental subgradient methods for nondifferentiable optimization*, SIAM Journal on Optimization **56** (2001), no. 1, 109–138.

[14] A. Nedić, D.P. Bertsekas, and V.S. Borkar, *Distributed asynchronous incremental subgradient methods*, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Studies Comput. Math., Elsevier, 2001.

[15] A. Nedić and A. Ozdaglar, *Approximate primal solutions and rate analysis for dual subgradient methods*, Preprint, 2007.

[16] ______, *Asynchronous consensus with delays*, Preprint, 2007.

[17] 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** (2004), no. 9, 1520–1533.

[18] A. Olshevsky and J.N. Tsitsiklis, *Convergence rates in distributed consensus averaging*, Proceedings of IEEE CDC, 2006.

[19] ______, *Convergence speed in distributed consensus and averaging*, Preprint, 2006.

[20] B.T. Polyak, *Introduction to optimisation*, Optimization Software, Inc., New York, 1987.

[21] N.Z. Shor, *Minimization methods for nondifferentiable functions*, Translated from Russian by K.C. Kiwiel and A. Ruszczynski, Springer, Berlin, 1985.

[22] R. Srikant, *Mathematics of Internet congestion control*, Birkhauser, 2004.

[23] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, 1984.

[24] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control **31** (1986), no. 9, 803–812.

[25] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and Schochet O., *Novel type of phase transitions in a system of self-driven particles*, Physical Review Letters **75** (1995), no. 6, 1226–1229.