Fifty-first Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 2 - 3, 2013

Continuous-time Distributed Optimization of Homogenous Dynamics

Guodong Shi, Alexandre Proutiere and Karl Henrik Johansson

Abstract— **This paper makes an attempt to explore the fundamental properties of distributed methods for minimizing a sum of objective functions with each component only known to a particular node, given a certain level of node knowledge and computation capacity. The information each node receives from its neighbors can be any nonlinear function of its neighbors' states as long as the function takes value zero within the local consensus manifold. Each node also observes the gradient of its own objective function at its current state. The update dynamics of each node is a first-order integrator. The admissible control input of each node is homogeneous, given by a binary function with each variable corresponding to the neighboring term and the gradient term, respectively. The function determining the control law is assumed to be injective when the first variable is fixed to zero. It is proven that there exists a control rule which guarantees global optimal consensus if and only if the solution sets of the local objectives admit a nonempty intersection set for fixed strongly connected graphs. Then we show that for any tolerated error, we can find a simple control rule that guarantees global optimal consensus within this error for fixed, bidirectional, and connected graphs under mild conditions. For time-varying graphs, we show that optimal consensus can always be achieved by a simple control rule as long as the graph is uniformly jointly strongly connected and the nonempty intersection condition holds. The results illustrate that nonempty intersection for the local optimal solution sets is a critical condition for distributed optimization using consensus processing to connect the information over the nodes.**

Index Terms— **Distributed optimization, Dynamical Systems, Multi-agent systems, Optimal consensus**

I. Introduction

In recent years networked dynamics for various problems such as consensus, formation, coverage, etc., have received much research interest [16], [11], [10], [8], [14], [18], [20], [21], [24], [12], [15]. A central idea is that a collective task can be reached for a group of nodes as long as, the right information is exchanged by each node, the proper design of individual node dynamics, and the communication graph is well structured.

Distributed control, communication and estimation in multi-agent systems and wireless networks naturally result in distributed optimization problems [29], [31], [30], [32]. Minimizing a sum of convex objective functions, $\sum_{i=1}^{N} f_i(z)$, where each component f_i is known only to node i, has served as a basic model. Tremendous research efforts have been devoted to finding solutions to this distributed optimization problem, and a number of methods have been derived which were proven to be effective [35], [32], [33], [25], [26], [27], [36], [28], [41], [39], [40].

However, the literature has not to sufficient extent studied the real meaning of "distributed" optimization, or the level of "distribution" possible for convergence. Some algorithms perform better than others in the sense of faster convergence, while they depend on more information exchange and more complex iteration rule. In order for a precise definition of distributed optimization method, the knowledge set of the sharing information among the nodes, and the computation capacity each node is equipped, should be clarified.

An interesting question arises: fixing the knowledge set and the computation capacity, what is the best one can do for solving the distributed optimization problems? In this paper, we make an attempt to propose an answer to such a question when the knowledge set for each node i is restricted to a neighboring term as any nonlinear function of node i's neighbors' states which takes value zero within the local consensus manifold, and a gradient term as the gradient of f_i at its current estimate, and the node dynamics is restricted to continuous-time feedback laws with respect to the neighboring and gradient terms.

The main results we obtain are stated as follows:

- We prove that there exists a control rule which guarantees global optimal consensus if and only if the solution sets of $f_i, i = 1, \dots, N$ admit a nonempty intersection set for fixed strongly connected graphs.
- We show that given any $\epsilon > 0$, a simple control rule can be found which guarantees global optimal consensus with error no larger than ϵ for fixed, bidirectional, and connected graphs under some mild conditions.
- We show that optimal consensus can always be achieved by a simple control rule with time-varying graphs as long as the graph is uniformly jointly strongly connected and the nonempty intersection condition holds.

The rest of the paper is organized as follows. In Section II, some preliminary mathematical concepts and lemmas are introduced. In Section III, we formulate considered optimization model, node dynamics, and define the problem of interest. Section IV focuses on fixed graphs, where a necessary and sufficient condition will be presented for exact solutions of optimal consensus, and then approximate solutions are investigated by ϵ-optimal consensus. Section V turns to time-varying graphs, and we will show optimal consensus under uniformly jointly strongly connected graphs. Finally, in Section VI some concluding remarks are given.

This work has been supported in part by the NSFC of China under Grant 61120106011, the Knut and Alice Wallenberg Foundation, the Swedish Research Council, and KTH SRA TNG.

The authors are with ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm 10044, Sweden.Email: guodongs@kth.se, alepro@kth.se, kallej@kth.se

II. Preliminaries

In this section, we introduce some notations and provide preliminary results that will be used in the rest of the paper.

A. Directed Graphs

A directed graph (digraph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a finite set $\mathcal{V}$ of nodes and an arc set $\mathcal{E}$, where an arc is an ordered pair of distinct nodes of $\mathcal{V}$. An element $(i, j) \in \mathcal{E}$ describes an arc which leaves i and enters j. A *walk* in digraph $\mathcal{G}$ is an alternating sequence $\mathcal{W} : i_1 e_1 i_2 e_2 \dots e_{m-1} i_m$ of nodes i_κ and arcs $e_\kappa = (i_\kappa, i_{\kappa+1}) \in \mathcal{E}$ for $\kappa = 1, 2, \dots, m-1$. A walk is called a *path* if the nodes of this walk are distinct, and a path from i to j is denoted as $i \to j$. $\mathcal{G}$ is said to be *strongly connected* if it contains path $i \to j$ and $j \to i$ for every pair of nodes i and j. A digraph $\mathcal{G}$ is called *bidirectional* when for any two nodes i and j, $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$. Ignoring the direction of the arcs, the connectivity of a bidirectional digraph will be transformed to that of the corresponding undirected graph. A time-varying graph is defined as $\mathcal{G}_{\sigma(t)} = (\mathcal{V}, \mathcal{E}_{\sigma(t)})$ with $\sigma : [0, +\infty) \to \mathcal{Q}$ as a piecewise constant function, where $\mathcal{Q}$ is a finite set containing all possible graphs with node set $\mathcal{V}$. Moreover, the joint graph of $\mathcal{G}_{\sigma(t)}$ in time interval $[t_1, t_2)$ with $t_1 < t_2 \le +\infty$ is defined as $\mathcal{G}([t_1, t_2)) = \cup_{t \in [t_1, t_2)} \mathcal{G}(t) = (\mathcal{V}, \cup_{t \in [t_1, t_2)} \mathcal{E}_{\sigma(t)})$.

B. Dini Derivatives

The upper *Dini derivative* of a continuous function $h : (a, b) \to \mathbb{R}$ $(-\infty \le a < b \le \infty)$ at t is defined as

$$D^+ h(t) = \limsup_{s \to 0^+} \frac{h(t+s) - h(t)}{s}.$$

When h is continuous on (a, b), h is non-increasing on (a, b) if and only if $D^+ h(t) \le 0$ for any $t \in (a, b)$. The next result is given for the calculation of Dini derivative (see [6], [23]).

Lemma 2.1: Let $V_i(t, x) : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ $(i = 1, \dots, n)$ be C^1 and $V(t, x) = \max_{i=1,\dots,n} V_i(t, x)$. If $\mathcal{I}(t) = \{i \in \{1, 2, \dots, n\} : V(t, x(t)) = V_i(t, x(t))\}$ is the set of indices where the maximum is reached at t, then $D^+ V(t, x(t)) = \max_{i \in \mathcal{I}(t)} \dot{V}_i(t, x(t))$.

C. Limit Sets

Consider the following autonomous system

$$\dot{x} = f(x), \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous function. Let $x(t)$ be a solution of (1) with initial condition $x(t_0) = x^0$. Then $\Omega_0 \subset \mathbb{R}^d$ is called a *positively invariant set* of (1) if, for any $t_0 \in R$ and any $x^0 \in \Omega_0$, we have $x(t) \in \Omega_0$ when $t \ge t_0$ along every solution $x(t)$ of (1).

We call y a ω-limit point of $x(t)$ if there exists a sequence $\{t_k\}$ with $\lim_{k \to \infty} t_k = \infty$ such that

$$\lim_{k \to \infty} x(t_k) = y.$$

The set of all ω-limit points of $x(t)$ is called the ω-limit set of $x(t)$, denoted as $\Lambda^+\big(x(t)\big)$. The following lemma is well-known (see [5]).

Lemma 2.2: Let $x(t)$ be a solution of (1). Then $\Lambda^+\big(x(t)\big)$ is positively invariant. Moreover, if $x(t)$ is contained in a compact set, then $\Lambda^+\big(x(t)\big) \neq \emptyset$.

D. Convex Analysis

A set $K \subset \mathbb{R}^d$ is said to be *convex* if $(1-\lambda)x + \lambda y \in K$ whenever $x \in K, y \in K$ and $0 \le \lambda \le 1$. For any set $S \subset \mathbb{R}^d$, the intersection of all convex sets containing S is called the *convex hull* of S, denoted by $co(S)$.

Let K be a closed convex subset in $\mathbb{R}^d$ and denote $|x|_K \doteq \inf_{y \in K} |x - y|$ as the distance between $x \in \mathbb{R}^d$ and K, where $|\cdot|$ denotes the Euclidean norm. There is a unique element $P_K(x) \in K$ satisfying $|x - P_K(x)| = |x|_K$ associated to any $x \in \mathbb{R}^d$ [2]. The map P_K is called the *projector* onto K. The following properties hold (see [2]).

Lemma 2.3: (i). We have

$$\langle P_K(x) - x, P_K(x) - y \rangle \le 0, \quad \forall y \in K. \tag{2}$$

(ii). P_K has the following non-expansiveness property:

$$|P_K(x) - P_K(y)| \le |x - y|, x, y \in \mathbb{R}^d. \tag{3}$$

(iii) $|x|_K^2$ is continuously differentiable at point x with

$$\nabla |x|_K^2 = 2\big(x - P_K(x)\big). \tag{4}$$

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a real-valued function. We call f a convex function if for any $x, y \in \mathbb{R}^d$ and $0 \le \lambda \le 1$, it holds that $f\big((1-\lambda)x + \lambda y\big) \le (1-\lambda) f(x) + \lambda f(y)$. The following lemma states some well-known properties for convex functions.

Lemma 2.4: Let $f : \mathbb{R}^d \to \mathbb{R} \in C^1$ be a convex function.

(i). $f(x) \ge f(y) + \big\langle x - y, \nabla f(y) \big\rangle$.

(ii). Any local minimum is a global minimum, i.e., $\arg\min f = \big\{z : \nabla f(z) = 0\big\}$.

III. Problem Definition

A. Optimization Model

Consider a network with node set $\mathcal{V} = \{1, 2, \dots, N\}$. The state of node i is denoted as $x_i \in \mathbb{R}^m$, and node i is associated with a cost function $f_i : \mathbb{R}^m \to \mathbb{R}$ which is observed by node i only. The objective for this group of autonomous agents is to cooperatively solve the following optimization problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N f_i(z) \\ \text{subject to} & z \in \mathbb{R}^m. \end{array} \tag{5}$$

We impose the following assumption on the functions $f_i, i = 1, \dots, N$.

A1. For all $i = 1, \dots, N$, we have (i) $f_i \in C^1$; (ii) f_i is a convex function; (iii) $\arg\min f_i \neq \emptyset$.

Problem (5) is equivalent with the following problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N f_i(x_i) \\ \text{subject to} & x_i \in \mathbb{R}^m \\ & x_1 = \dots = x_N. \end{array} \tag{6}$$

Then from (6) we see that consensus is a natural mean for solving (5).

B. Dynamics

The dynamics of each node is assumed to be a first-order integrator:

$$\dot{x}_i = u_i, \quad i = 1, \dots, N, \tag{7}$$

where u_i is the control input.

The structure of the information flow through the network is modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. A node j is said to be a *neighbor* of i at time t when there is an arc $(j, i) \in \mathcal{E}$, and we denote $\mathcal{N}_i$ the set of neighbors for node i.

The local optimization information, $g_i(t)$, that node i receives from its objective f_i, is simply assume to be the gradient of f_i at its current state, i.e.,

$$g_i(t) \doteq \nabla f_i\big(x_i(t)\big). \tag{8}$$

We define the neighboring information, $n_i(t)$ which node i receives from its neighbors at time t as

$$n_i(t) \doteq h_i\big(x_i(t), x_j(t) : j \in \mathcal{N}_i\big). \tag{9}$$

where $h_i : \mathbb{R}^m \times \mathbb{R}^{m|\mathcal{N}_i|} \to \mathbb{R}^n$ is a continuous function with $|\mathcal{N}_i|$ denoting the number of elements in $\mathcal{N}_i$ and n a given integer indicating the dimension of the neighboring information. For each h_i, we use the following assumption.

A2. For all $i \in \mathcal{V}$, $h_i \equiv 0$ within the local consensus manifold $\big\{x_i = x_j : j \in \mathcal{N}_i\big\}$.

Remark 3.1: Assumption A2 is to say that the neighboring information a node receives from the neighbors becomes trivial when it is in the same state with all its neighbors. This is a quite natural assumption in the literature on distributed averaging and optimization algorithms [16], [24], [13], [26], [27].

For the control strategy of the nodes, we impose the following assumption.

A3. There exists a function $\mathcal{J} \in \mathscr{F}_* \doteq \big\{\mathcal{F}(\cdot,\cdot) \in C^0 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m,\ \mathcal{F}(0,\cdot) \text{ is injective}\big\}$ such that

$$u_i = \mathcal{J}\big(n_i, g_i\big),\ i = 1, \dots, N. \tag{10}$$

Remark 3.2: Assumption A3 indicates that control protocols are applied to each node without difference with respect to individual local optimization information and neighboring information. Note that our network model is an equitable world because one cannot tell the difference from one node to another. The admissible control rule requires $\mathcal{J}(0,\cdot)$ to be injective, which indicates that each node should take different response to different gradient information on the local consensus manifold. Assumption A3 indeed has been widely applied in the literature, see [16], [24], [13], [26], [27].

C. Problem

Let $x(t)$ be the trajectory of system (7) with control law (10) for initial condition $x^0 = x(t_0) = (x_1^T(t_0), \dots, x_N^T(t_0))^T \in \mathbb{R}^{mN}$. Denote $F(z) = \sum_{i=1}^N f_i(z)$. We introduce the following definition.

Definition 3.1: Global *optimal consensus* is achieved if for all $x^0 \in \mathbb{R}^{mN}$, we have

$$\limsup_{t\to+\infty} F\big(x_i(t)\big) = \min_{z\in\mathbb{R}^m} F(z) \tag{11}$$

and

$$\lim_{t\to+\infty} \big|x_i(t) - x_j(t)\big| = 0, \quad i, j = 1, \dots, N. \tag{12}$$

We see from the definition of optimal consensus that it means a consensus meanwhile solving the optimization problem (5).

IV. Fixed Graphs: Fundamental Limit and Approximate Solution

In this section, we consider the possibility of solving optimal consensus using feedback law (10) under fixed communication graphs. We first discuss whether exact optimal consensus can be reached for the network with directed node interactions. Then we show the existence of an approximate solution for optimal consensus with bidirectional graphs.

A. Exact Solution

This section presents the result for optimal consensus. We make another assumption on the solution set of $F = \sum_{i=1}^N f_i$.

A4. $\arg\min F(z) \neq \emptyset$ is a bounded set.

The main result we obtain on the existence of a control law solving optimal consensus is stated as follows.

Theorem 4.1: Assume that A1–A4 hold. Let the communication graph $\mathcal{G}$ be fixed and strongly connected. There exists a distributed control which guarantees global optimal consensus if and only if

$$\bigcap_{i=1}^N \arg\min f_i(z) \neq \emptyset. \tag{13}$$

According to Theorem 4.1, the optimal solution sets of $f_i,\ i = 1, \dots, N$ having nonempty intersection is a critical condition for the existence of a simple control rule (10) that solves the optimal consensus problem. The $\arg\min f_i's$ admit an nonempty intersection is a strong constraint which in general does not hold. Therefore, basically Theorem 4.1 suggests that exact solution for optimal consensus is impossible of the considered model. The sufficiency claim of Theorem 4.1 follows from the upcoming Theorem 5.2. The proof of the necessity part is as follows.

Proof of Theorem 4.1: (Necessity) Suppose $\bigcap_{i=1}^N \arg\min f_i(z) = \emptyset$ and there exists a distributed control in the form of (10), say $\mathcal{J}_0\big(n_i, g_i\big)$, under which global optimal consensus is reached. Let $x(t)$ be a trajectory of system (7) with control $\mathcal{J}_0\big(n_i, g_i\big)$ and $\Lambda^+(x(t))$ be its ω-limit set. The definition of optimal consensus leads to that $x(t)$ converges to the bounded set $\Big(\arg\min F(z)\Big)^N \bigcap \mathcal{M}$, where $\Big(\arg\min F(z)\Big)^N$ denotes the N'th power set of

$\arg\min F(z)$ and $\mathcal{M}$ denotes the consensus manifold, defined by

$$\mathcal{M} \doteq \big\{x = (x_1^T \dots x_N^T)^T \in \mathbb{R}^{mN}: \ x_1 = \dots = x_N\big\}. \tag{14}$$

Therefore, each trajectory $x(t)$ is contained in a compact set.

Based on Lemma 2.2, we conclude that $\Lambda^+(x(t)) \neq \emptyset$ and

$$\Lambda^+(x(t)) \subseteq \Big(\arg\min F(z)\Big)^N \bigcap \mathcal{M}, \tag{15}$$

Moreover, $\Lambda^+(x(t))$ is positively invariant since system (7) is autonomous under control $\mathscr{J}_0\big(n_i, g_i\big)$ when the communication graph is fixed. This is to say, any trajectory of system (7) under control $\mathscr{J}_0\big(n_i, g_i\big)$ must stay within $\Lambda^+(x(t))$ for any initial value in $\Lambda^+(x(t))$.

Now we take $y \in \Lambda^+(x(t))$. Then we have $y \in \Big(\arg\min F(z)\Big)^N \bigcap \mathcal{M}$ according to (15), and thus $y = (z_*^T \dots z_*^T)^T$ for some $z_* \in \arg\min F(z)$. With Assumption A1, the convexity of f_i's implies that

$$\arg\min F(z) = \big\{z \in \mathbb{R}^m: \ \sum_{i=1}^N \nabla f_i(z) = 0\big\}. \tag{16}$$

On the other hand, we have

$$\bigcap_{i=1}^N \arg\min f_i(z) = \bigcap_{i=1}^N \big\{z \in \mathbb{R}^m: \ \nabla f_i(z) = 0\big\} = \emptyset.$$

Therefore, there exist two indices $i_1, i_2 \in \{1, \dots, N\}$ with $i_1 \neq i_2$ such that

$$\nabla f_{i_1}(z_*) \neq \nabla f_{i_2}(z_*). \tag{17}$$

Consider the solution of (7) under control $\mathscr{J}_0\big(n_i, g_i\big)$ for initial time t_0 and initial value y. The fact that y belongs to the consensus manifold guarantees

$$n_{i_1}(t_0) = n_{i_2}(t_0) = 0. \tag{18}$$

With Assumption A4, we have

$$\begin{aligned}\mathscr{J}_0\big(n_{i_1}(t_0), g_{i_1}(t_0)\big) &= \mathscr{J}_0\big(0, \nabla f_{i_1}(z_*)\big)\\ &\neq \mathscr{J}_0\big(0, \nabla f_{i_2}(z_*)\big)\\ &= \mathscr{J}_0\big(n_{i_2}(t_0), g_{i_2}(t_0)\big).\end{aligned} \tag{19}$$

This implies $\dot{x}_{i_1}(t_0) \neq \dot{x}_{i_2}(t_0)$. As a result, there exists a constant $\varepsilon > 0$ such that $x_{i_1}(t) \neq x_{i_2}(t)$ for $t \in (t_0, t_0 + \varepsilon)$. In other word, the trajectory will leave the set

$$\Big(\arg\min F(z)\Big)^N \bigcap \mathcal{M}$$

for $(t_0, t_0+\varepsilon)$, and therefore will also leave the set $\Lambda^+(x(t))$. This contradicts the fact that $\Lambda^+(x(t))$ is positively invariant. The necessity part of Theorem 4.1 has been proved.

B. Approximate Solution

Theorem 4.1 indicates that optimal consensus is impossible no matter how the control rule is chosen from $\mathscr{F}_*$ as long as the nonempty intersection condition (13) is not fulfilled. In this subsection, we discuss the approximate solution of optimal consensus in the absence of (13). We introduce the following definition.

Definition 4.1: Global *ϵ-optimal consensus* is achieved if for all $x^0 \in \mathbb{R}^{mN}$, we have

$$\limsup_{t\to+\infty} F\big(x_i(t)\big) \leq \min_{z\in\mathbb{R}^m} F(z) + \epsilon \tag{20}$$

and

$$\lim_{t\to+\infty} \big|x_i(t) - x_j(t)\big| \leq \epsilon, \quad i, j = 1, \dots, N. \tag{21}$$

Let the communication graph $\mathcal{G}$ be bidirectional. In this case, we will use an unordered pair $\{i, j\}$ to denote the edge between node i and j in $\mathcal{E}$. We also allow a weight $a_{ij} > 0$ marking the strength of the information flow of edge $\{i, j\}$. Hence we have $a_{ij} = a_{ji}$ in this case.

Denoting $F_{\mathcal{G}}(x; K) = \sum_{i=1}^N f_i(x_i) + \frac{K}{2} \sum_{\{j,i\}\in\mathcal{E}} a_{ij}\big|x_j - x_i\big|^2$, we impose the following assumption.

A5. (i) $\arg\min F(z) \neq \emptyset$, where $F(z) = \sum_{i=1}^N f_i(z)$; (ii) $\arg\min F_{\mathcal{G}}(x; K) \neq \emptyset$ for all $K \geq 0$; (iii) $\bigcup_{K\geq 0} \arg\min F_{\mathcal{G}}(x; K)$ is bounded.

For ϵ-optimal consensus, we present the following result.

Theorem 4.2: Assume that A1, A2, A3, and A5 hold. Let the communication graph $\mathcal{G}$ be fixed, bidirectional, and connected. Then for any $\epsilon > 0$, there exists a constant $K(\epsilon) > 0$ such that simple control rule

$$u_i = \mathscr{J}_K(n_i, g_i) \doteq K(\epsilon) \sum_{j\in\mathcal{N}_i} a_{ij}\big(x_j - x_i\big) - \nabla f_i\big(x_i\big) \tag{22}$$

guarantees global ϵ-optimal consensus.

Proof. It is straightforward to see that

$$\begin{aligned}\mathscr{J}_K(n_i, g_i) &= K \sum_{j\in\mathcal{N}_i} a_{ij}\big(x_j - x_i\big) - \nabla f_i\big(x_i\big)\\ &= -\nabla_{x_i}\Big(\frac{K}{2} \sum_{j\in\mathcal{N}_i} a_{ij}\big|x_j - x_i\big|^2 + f_i(x_i)\Big).\end{aligned} \tag{23}$$

Thus, System (7) with control law $u_i = \mathscr{J}_K(n_i, g_i)$ can be written into the following compact form

$$\dot{x} = -\nabla F_{\mathcal{G}}(x; K), \quad x = (x_1^T \dots x_N^T)^T \in \mathbb{R}^{mN}. \tag{24}$$

Then the convexity of $F_{\mathcal{G}}(x; K)$ ensures that every trajectory of (22) asymptotically solves the following convex optimization problem

$$\begin{array}{ll}\text{minimize} & F_{\mathcal{G}}(x; K)\\ \text{subject to} & x_i \in \mathbb{R}^m, \ i = 1, \dots, N.\end{array} \tag{25}$$

Convexity gives

$$\begin{aligned}\arg\min F_{\mathcal{G}}(x; K) = \Big\{x: \ &-K(L \otimes I_m)x\\ &= \Big(\big(\nabla f_1(x_1)\big)^T \dots \big(\nabla f_N(x_N)\big)^T\Big)^T\Big\},\end{aligned} \tag{26}$$

where $\otimes$ represents the Kronecker product, I_m is the identity matrix in $\mathbb{R}^m$, and $L = D - A$ is the Laplacian of graph $\mathcal{G}$ with $A = [a_{ij}]$ and $D = \mathrm{diag}(d_1, \ldots, d_N)$, where $d_i = \sum_{j=1}^{n} a_{ij}$. Under Assumptions A1 and A5, we have

$$L_0 \doteq \sup\Big\{\big|\nabla \tilde{F}(x)\big| : \ x \in \bigcup_{K\geq 0} \arg\min F_{\mathcal{G}}(x;K)\Big\} \quad (27)$$

is a finite number, where $\tilde{F}(x) = \sum_{i=1}^{N} f_i(x_i)$. We also define

$$\begin{aligned} D_0 \doteq \sup\Big\{\big|z_* - x_i\big| : \ & i = 1,\ldots,N, \\ & x \in \bigcup_{K\geq 0} \arg\min F_{\mathcal{G}}(x;K)\Big\}, \end{aligned} \quad (28)$$

where $z_* \in \arg\min F$ is an arbitrarily chosen point.

Let $p = (p_1^T \ldots p_N^T)^T \in \arg\min F_{\mathcal{G}}(x;K)$. Since the graph is bidirectional and connected, we can sort the eigenvalues of the Laplacian $L \otimes I_m$ as

$$0 = \lambda_1 = \cdots = \lambda_m < \lambda_{m+1} \leq \cdots \leq \lambda_{mN}.$$

Let $l_1 \ldots, l_{mN}$ be the orthonormal basis of $\mathbb{R}^{mN}$ formed by the right eigenvectors of $L \otimes I_m$, where $l_1, \ldots, l_m$ are eigenvectors corresponding to the zero eigenvalue. Suppose $p = \sum_{k=1}^{mN} c_k l_k$ with $c_k \in \mathbb{R}, k = 1, \ldots, mN$.

According to (26), we have

$$\Big|K(L \otimes I_m)p\Big|^2 = K^2 \sum_{k=m+1}^{mN} c_k^2 \lambda_k^2 \leq L_0^2, \quad (29)$$

which yields

$$\sum_{k=m+1}^{mN} c_k^2 \leq \Big(\frac{L_0}{K\lambda_2^*}\Big)^2, \quad (30)$$

where $\lambda_2^* = \lambda_{m+1} > 0$ denotes the second smallest eigenvalue of L.

Now recall that

$$\mathcal{M} \doteq \big\{x = (x_1^T \ldots x_N^T)^T : \ x_1 = \cdots = x_N\big\}. \quad (31)$$

is the consensus manifold. Noticing that $\mathcal{M} = \mathrm{span}\{l_1, \ldots, l_m\}$, we conclude from (30) that

$$\begin{aligned} \sum_{k=m+1}^{mN} c_k^2 &= \Big|\sum_{k=m+1}^{mN} c_k l_k\Big|^2 \\ &= |p|_{\mathcal{M}}^2 \\ &= \sum_{i=1}^{N} \Big|p_i - \frac{\sum_{i=1}^{N} p_i}{N}\Big|^2 \\ &\leq \Big(\frac{L_0}{K\lambda_2^*}\Big)^2. \end{aligned} \quad (32)$$

Thus, for any $\varsigma > 0$, there is $K_1(\varsigma) > 0$ such that when $K \geq K_1(\varsigma)$,

$$\Big|p_i - p_{\mathrm{ave}}\Big| \leq \varsigma, \ i = 1, \ldots, N$$

and

$$\Big|F(p_i) - F(p_{\mathrm{ave}})\Big| \leq \varsigma, \ i = 1, \ldots, N,$$

where $p_{\mathrm{ave}} = \frac{\sum_{i=1}^{N} p_i}{N}$.

On the other hand, with (26), we have

$$\sum_{i=1}^{N} \nabla f_i(p_i) = \sum_{i=1}^{N} \nabla f_i(p_{\mathrm{ave}} + \hat{p}_i) = 0, \quad (33)$$

where and $\hat{p}_i = p_i - p_{\mathrm{ave}}$. Now since each $f_i \in C^1$, for any $\varsigma > 0$, there is $K_2(\varsigma) > 0$ such that when $K \geq K_2(\varsigma)$,

$$\Big|\sum_{i=1}^{N} \nabla f_i(p_{\mathrm{ave}})\Big| \leq \frac{\varsigma}{D_0}. \quad (34)$$

This implies

$$\begin{aligned} F(p_{\mathrm{ave}}) &\leq F(z_*) + |z_* - p_{\mathrm{ave}}| \times \Big|\sum_{i=1}^{N} \nabla f_i(p_{\mathrm{ave}})\Big| \\ &\leq F(z_*) + \varsigma. \end{aligned} \quad (35)$$

Therefore, for any $\epsilon > 0$, we can take $K_0 = \max\{K_1(\epsilon/2), K_2(\epsilon/2)\}$. Then when $K \geq K_0$, we have

$$|p_i - p_j| \leq \epsilon; \quad F(p_i) \leq \min_z F(z) + \epsilon \quad (36)$$

for all i and j. Now that $F_{\mathcal{G}}(x;K)$ is a convex function and observing (24), every limit point of System (7) with control rule $\mathcal{J}_K(n_i, g_i)$ is contained in the set $\arg\min F_{\mathcal{G}}(x;K)$. Noting that p is arbitrarily chosen from $\arg\min F_{\mathcal{G}}(x;K)$, ϵ-optimal consensus is achieved as long as we choose $K \geq K_0$. This completes the proof. □

With Theorem 4.2, we see that even though we know that without nonempty intersection condition, it is impossible to reach optimal consensus via any feedback control in the form of (10), it is still possible to find a control law (10) which guarantees approximate optimal consensus with arbitrary accuracy.

It is worth pointing out that to determine the proper K in (22) for a given ϵ relies on the knowledge of structure of the network, and the information of all $f_i, i = 1, \ldots, N$. This to say, finding a proper control (22) for ϵ-optimal consensus requires the global knowledge of the whole network. Apparently the nonempty intersection condition in Theorem 4.1 is also a global knowledge. Then we see that from Theorem 4.1 to Theorem 4.2, some global information (or constraint) is always needed to guarantee a collective convergence.

C. Assumption Feasibility

This subsection discusses the feasibility of Assumptions A4 and A5 and shows that some mild conditions are enough to ensure A4 and A5.

Proposition 4.1: Assume that A1 holds. If $\tilde{F}(x) = \sum_{i=1}^{N} f_i(x_i)$ is coercive, i.e., $\tilde{F}(x) \to \infty$ as long as $|x| \to \infty$, then A4 and A5 hold.

Proof. a). Since $\tilde{F}(x) = \sum_{i=1}^{N} f_i(x_i)$ is coercive, it follows straightforwardly that $F(z) = \sum_{i=1}^{N} f_i(z)$ is also coercive. As a result, $\arg\min F(z) \neq \emptyset$ is a bounded set. Thus, A4 and A5.(i) hold.

b). Observing that $\frac{K}{2} \sum_{\{j,i\}\in\mathcal{E}} a_{ij}\big|x_j - x_i\big|^2 \geq 0$ for all $x = (x_1^T \ldots x_N^T)^T \in \mathbb{R}^{mN}$ and that $\tilde{F}(x) = \sum_{i=1}^{N} f_i(x_i)$

is coercive, we obtain that $\arg\min F_{\mathcal{G}}(x;K) \neq \emptyset$ for all $K \geq 0$. Thus, A5.(ii) holds.

c). Based on a), we can denote $F_* = \min_z F(z) = F(z_*)$. Since $\sum_{i=1}^N f_i(x_i)$ is coercive, there exists a constant $M(F_*) > 0$ such that $\sum_{i=1}^N f_i(x_i) > F_*$ for all $|x| > M$. This implies

$$F_{\mathcal{G}}(x;K) > F_{\mathcal{G}}(\mathbf{1}_N \otimes z_*;K) = F_* \tag{37}$$

for all $|x| > M$. That is to say, the global minimum of $F_{\mathcal{G}}(x;K)$ is reached within the set $\{|x| \leq M\}$ for all $K > 0$. Therefore, we have

$$\bigcup_{K\geq 0} \arg\min F_{\mathcal{G}}(x;K) \subseteq \big\{|x| \leq M\big\}. \tag{38}$$

This proves A5.(iii). □

V. Time-varying Graphs

Now we consider time-varying graphs. The communication in the multi-agent network is modeled as $\mathcal{G}_{\sigma(t)} = (\mathcal{V}, \mathcal{E}_{\sigma(t)})$ with $\sigma : [0,+\infty) \to \mathcal{Q}$ as a piecewise constant function, where $\mathcal{Q}$ is a finite set indicating all possible graphs. In this case the neighbor set for each node is time-varying, and we let $\mathcal{N}_i(\sigma(t))$ represent the set of agent i's neighbors at time t. As usual in the literature [16], [23], [20], an assumption is given to the variation of $\mathcal{G}_{\sigma(t)}$.

A6 *(Dwell Time)* There is a lower bound constant $\tau_D > 0$ between two consecutive switching time instants of $\sigma(t)$.

We also allow the arc weights a_{ij} to be time-varying and another assumption is made on each $a_{ij}(t), i,j = 1,2,...,N$.

A7 *(Weights Rule)* (i) Each $a_{ij}(t)$ is piece-wise continuous and $a_{ij}(t) \geq 0$ for all i and j.
(ii). There are $a^* > 0$ and $a_* > 0$ such that $a_* \leq a_{ij}(t) \leq a^*, \quad t \in \mathbb{R}^+$.

We have the following definition.

Definition 5.1: (i) $\mathcal{G}_{\sigma(t)}$ is said to be *uniformly jointly strongly connected* (UJSC) if there exists a constant $T > 0$ such that $\mathcal{G}([t,t+T))$ is strongly connected for any $t \geq 0$.

(ii) $\mathcal{G}_{\sigma(t)}$ is said to be *uniformly jointly quasi-strongly connected* (UQSC) if there exists a constant $T > 0$ such that $\mathcal{G}([t,t+T))$ has a spanning tree for any $t \geq 0$.

For optimal consensus with time-varying graphs, we present the following result.

Theorem 5.1: Suppose A1–A3, A6–A7 hold and $\mathcal{G}_{\sigma(t)}$ is UJSC. Suppose $\bigcap_{i=1}^N \arg\min f_i \neq \emptyset$, and it contains at least one interior point. Then simple control rule

$$u_i = \mathcal{J}_\star(n_i, g_i) \doteq \sum_{j\in\mathcal{N}_i(\sigma(t))} a_{ij}(t)\big(x_j - x_i\big) - \nabla f_i\big(x_i\big)$$

guarantees global optimal consensus. Furthermore, there exists a point $x_* \in \bigcap_{i=1}^N \arg\min f_i$ such that

$$\lim_{t\to\infty} x_i(t) = x_*. \tag{39}$$

Note that (39) indeed a stronger conclusion than our definition of optimal consensus since it guarantees all the node states converge to a common point in the global solution set of $F(z)$. We will see from the proof of Theorem 5.1 that this state convergence highly relies on the existence of a interior point $\bigcap_{i=1}^N \arg\min f_i$. In the absence of this interior point condition, it turns out that optimal consensus still stands. We present another theorem.

Theorem 5.2: Suppose A1–A3, A6–A7 hold and $\mathcal{G}_{\sigma(t)}$ is UJSC. Suppose also $\bigcap_{i=1}^N \arg\min f_i \neq \emptyset$. Then control rule

$$u_i = \mathcal{J}_\star(n_i, g_i) \doteq \sum_{j\in\mathcal{N}_i(\sigma(t))} a_{ij}(t)\big(x_j - x_i\big) - \nabla f_i\big(x_i\big)$$

guarantees global optimal consensus.

A. Preliminary Lemmas

We establish several useful lemmas in this subsection.

Suppose $\bigcap_{i=1}^N \arg\min f_i \neq \emptyset$ and take $z_* \in \bigcap_{i=1}^N \arg\min f_i$. We define

$$V_i(t) = \big|x_i(t) - z_*\big|^2,\ i = 1,\dots,N, \tag{40}$$

and then

$$V(t) = \max_{i=1,\dots,N} V_i(t). \tag{41}$$

The following lemma holds.

Lemma 5.1: Let A1, A3 and A7 hold. Suppose $\bigcap_{i=1}^N \arg\min f_i \neq \emptyset$. Then along any trajectory of system (7) with control rule $\mathcal{J}_\star(n_i, g_i)$, we have $D^+V(t) \leq 0$ for all t.

Proof. Based on Lemma 2.1, we have

$$\begin{aligned} D^+V(t) &= \max_{i\in\mathcal{I}(t)} \frac{d}{dt} V_i(t) \\ &= \max_{i\in\mathcal{I}(t)} 2\Big\langle x_i(t) - z_*, \\ &\qquad \sum_{j\in\mathcal{N}_i(\sigma(t))} a_{ij}(t)\big(x_j - x_i\big) - \nabla f_i\big(x_i\big)\Big\rangle, \end{aligned} \tag{42}$$

where $\mathcal{I}(t)$ denotes the index set which contains all the nodes reaching the maximum for $V(t)$.

Let $m \in \mathcal{I}(t)$. Denote

$$Z_t = \big\{z:\ |z - z_*| \leq \sqrt{V(t)}\big\}$$

as the disk centered at z_* with radius $\sqrt{V(t)}$. Take $y = x_m(t) + (x_m(t) - z_*)$. Then it is obvious to see that $P_{Z_t}(y) = x_m(t)$, where P_{Z_t} is the *projector* onto Z_t. Thus, for all $j \in \mathcal{N}_m(\sigma(t))$, we obtain

$$\begin{aligned} \big\langle x_m(t) - z_*, x_j(t) - x_m(t)\big\rangle &= \big\langle y - x_m(t), x_j(t) - x_m(t)\big\rangle \\ &= \big\langle y - P_{Z_t}(y), x_j(t) - P_{Z_t}(y)\big\rangle \\ &\leq 0 \end{aligned} \tag{43}$$

according to inequality (2) in Lemma 2.3 since $x_j(t) \in Z_t$. On the other hand, based on inequality (i) in Lemma 2.4, we also have

$$\big\langle x_m(t) - z_*, -\nabla f_m\big(x_m(t)\big)\big\rangle \leq f_m(z_*) - f_m\big(x_m(t)\big) \leq 0 \tag{44}$$

in light of the definition of z_*.

With (42), (43) and (44), we conclude that

$$D^{+}V(t) \leq 0, \tag{45}$$

which completes the proof. $\square$

A direct consequence of Lemma 5.1 is that when $\bigcap_{i=1}^{N} \arg\min f_i \neq \emptyset$, we have

$$\lim_{t\to\infty} V(t) = V_*^2 \tag{46}$$

for some $V_* \in \mathbb{R}$ along any trajectory of system (7) with control rule $\mathcal{J}_\star(n_i, g_i)$. However, it is still unclear whether $V_i(t)$ converges, or $V_i(t)$ converges to what if it did converge.

We establish another lemma indicating that with proper connectivity condition for the communication graph, all $V_i(t)$'s have the same limit V_*^2.

Lemma 5.2: Let A1, A6, and A7 hold. Suppose $\bigcap_{i=1}^{N} \arg\min f_i \neq \emptyset$ and $\mathcal{G}_{\sigma(t)}$ is UJSC. Then along any trajectory of system (7) with control rule $\mathcal{J}_\star(n_i, g_i)$, we have $\lim_{t\to\infty} V_i(t) = V_*^2$ for all i.

The next lemma shows that each node will reach its own optimum along the trajectories of system (7) under control rule $\mathcal{J}_\star(n_i, g_i)$.

Lemma 5.3: Let A1, A6, and A7 hold. Suppose $\bigcap_{i=1}^{N} \arg\min f_i \neq \emptyset$ and $\mathcal{G}_{\sigma(t)}$ is UJSC. Then along any trajectory of system (7) with control rule $\mathcal{J}_\star(n_i, g_i)$, we have $\limsup_{t\to\infty} \big|x_i(t)\big|_{\arg\min f_i} = 0$ for all i.

The proofs of Lemmas 5.2 and 5.3 are based on contradiction arguments, and are omitted due to space limitations.

B. Proof of Theorem 5.1

The proof of Theorem 5.1 relies on the following lemma.

Lemma 5.4: Let $z_1, \dots, z_{m+1} \in \mathbb{R}^m$ and $d_1, \dots, d_{m+1} \in \mathbb{R}^+$. Suppose there exist solutions to equations (with variable y)

$$\begin{cases} |y - z_1| = d_1; \\ \dots \\ |y - z_{m+1}| = d_{m+1}. \end{cases} \tag{47}$$

Then the the solution of (47) is unique if $\operatorname{rank}\big(z_2 - z_1, \dots, z_{m+1} - z_1\big) = m$.

Proof. Take $j > 1$ and let y be a solution to the equations. Noticing that

$$\langle y - z_1, y - z_1 \rangle = d_1; \quad \langle y - z_j, y - z_j \rangle = d_j$$

we obtain

$$\langle y, z_j - z_1 \rangle = \frac{1}{2}\Big(d_1 - d_j + |z_j|^2 - |z_1|^2\Big),\ j = 2, \dots, m+1. \tag{48}$$

The desired conclusion following immediately. $\square$

We now prove Theorem 5.1. Let $r_\star = (r_1^T \dots r_N^T)^T$ be a limit point of a trajectory of system (7) with control rule $\mathcal{J}_\star(n_i, g_i)$.

We first show consensus. Based on Lemma 5.2, we have $\lim_{t\to\infty} V_i(t) = V_*$ for all $z_* \in \bigcap_{i=1}^{N} \arg\min f_i$. This is to say, $|r_i - z_*| = V_*$ for all i and $z_* \in \bigcap_{i=1}^{N} \arg\min f_i$. Since $\bigcap_{i=1}^{N} \arg\min f_i \neq \emptyset$ contains at least one interior point, it is obvious to see that we can find $z_1, \dots, z_{m+1} \in \bigcap_{i=1}^{N} \arg\min f_i$ with $\operatorname{rank}\big(z_2 - z_1, \dots, z_{m+1} - z_1\big) = m$ and $d_1, \dots, d_{m+1} \in \mathbb{R}^+$, such that each $r_i, i = 1, \dots, N$ is a solution of equations (47). Then based on Lemma 5.4, we conclude that $r_1 = \dots = r_N$. Next, with Lemma 5.3, we have $|r_i|_{\arg\min f_i} = 0$. This implies that $r_1 = \dots = r_N \in \bigcap_{i=1}^{N} \arg\min f_i$, i.e., optimal consensus is achieved.

We turn to state convergence. We only need to show that $r_\star$ is unique along any trajectory of system (7) with control rule $\mathcal{J}_\star(n_i, g_i)$. Now suppose $r_\star^1 = \mathbf{1}_N \otimes r^1$ and $r_\star^2 = \mathbf{1}_N \otimes r^2$ are two different limit points with $r^1 \neq r^2 \in \bigcap_{i=1}^{N} \arg\min f_i$. According to the definition of limit point, we have that for any $\varepsilon > 0$. there exists a time instant t_ε such that $|x_i(t_\varepsilon) - r^1| \leq \varepsilon$ for all i. Note that Lemma 5.1 indicates that the disc $B(r^1, \varepsilon) = \{y : |y - r^1| \leq \varepsilon\}$ is an invariant set for initial time t_ε. While taking $\varepsilon = |r^1 - r^2|/4$, we see that $r^2 \notin B(r^1, |r^1 - r^2|/4)$. Thus, r^2 cannot be a limit point.

Now since the limit point is unique, we denote it as $\mathbf{1}_N \otimes x_*$ with $x_* \in \bigcap_{i=1}^{N} \arg\min f_i$. Then we have $\lim_{t\to\infty} x_i(t) = x_*$ for all $i = 1, \dots, N$. This completes the proof.

C. Proof of Theorem 5.2

In this subsection, we prove Theorem 5.2. We need the following lemma on robust consensus, which can be found in [22].

Lemma 5.5: Consider a network with node set $\mathcal{V} = \{1, \dots, N\}$ with time-varying communication graph $\mathcal{G}_{\sigma(t)}$. Let the dynamics of node i be

$$\dot{x}_i = \sum_{j \in \mathcal{N}_i(\sigma(t))} a_{ij}(t)\big(x_j - x_i\big) + w_i(t), \tag{49}$$

where $w_i(t)$ is a piecewise continuous function. Suppose A6 and A7 hold and $\mathcal{G}_{\sigma(t)}$ is UQSC. Then we have

$$\lim_{t\to+\infty} \big|x_i(t) - x_j(t)\big| = 0, \quad i, j = 1, \dots, N \tag{50}$$

if $\lim_{t\to\infty} w_i(t) = 0$ for all i.

Lemma 5.3 indicates that $\limsup_{t\to\infty} \big|x_i(t)\big|_{\arg\min f_i} = 0$ for all i, which yields

$$\lim_{t\to\infty} \nabla f_i\big(x_i(t)\big) = 0 \tag{51}$$

for all i according to Assumption A1. Then the consensus part in the definition of optimal consensus follows immediately from Lemma 5.5. Again by Lemma 5.3, we further conclude that $\limsup_{t\to\infty} \operatorname{dist}\big(x_i(t), \bigcap_{i=1}^{N} \arg\min f_i\big) = 0$. The desired conclusion thus follows.

VI. Conclusions

Various algorithms have been established in the literature for the distributed optimization problem of minimizing $\sum_{i=1}^{N} f_i$ with f_i only known to node i. This paper made an attempt to explore the fundamental limit that distributed methods can do for this problem given a certain level of node knowledge and computation capacity. We assume that the neighboring information each node receives from its neighbors can be any nonlinear function of its neighbors' states which takes value zero within the local consensus manifold, and each node i can get the gradient of f_i at its current state. The dynamics of each node is a first-order integrator with the control input given by a binary function, each variable corresponds to the neighboring information and the gradient information, respectively. This function determining the control law is assumed to be injective when the first variable is fixed to zero. We proved that there exists a control rule which guarantees global optimal consensus if and only if $\arg\min f_i, i = 1, \dots, N$ admit a nonempty intersection set for fixed strongly connected graphs. Then we showed that for any tolerated error, we can find a simple control rule which guarantees global optimal consensus within this error for fixed, bidirectional, and connected graphs under some mild conditions such as f_i is coercive for some i. For time-varying graphs, it was proven that optimal consensus can always be achieved by simple control rule as long as the graph is uniformly jointly strongly connected and the nonempty intersection condition holds. It was then concluded that nonempty intersection for the local optimal solution sets is a critical condition for distributed optimization using consensus processing to connect the information over the nodes.

REFERENCES

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[2] J. Aubin and A. Cellina, *Differential Inclusions*. Berlin: Speringer-Verlag, 1984.

[3] R. T. Rockafellar, *Convex Analysis.* New Jersey: Princeton University Press, 1972.

[4] C. Godsil and G. Royle, *Algebraic Graph Theory.* New York: Springer-Verlag, 2001.

[5] N. Rouche, P. Habets, and M. Laloy, *Stability Theory by Liapunov's Direct Method*. New York: Springer-Verlag, 1977.

[6] J. Danskin, "The theory of max-min, with applications," *SIAM J. Appl. Math.*, vol. 14, 641-664, 1966.

[7] F. Clarke, Yu.S. Ledyaev, R. Stern, and P. Wolenski, *Nonsmooth Analysis and Control Theory*. Berlin: Speringer-Verlag, 1998.

[8] S. Martinez, J. Cortes, and F. Bullo, " Motion coordination with distributed information," *IEEE Control Syst. Mag.*, vol. 27, no. 4, pp. 75-88, 2007.

[9] J. Cortés, "Discontinuous dynamical systems–A tutorial on solutions, nonsmooth analysis, and stability," *IEEE Control Systems Magazine*, vol. 28, no. 3, 36-73, 2008.

[10] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: algorithms and theory," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 401-420, 2006.

[11] M. Cao, A. S. Morse and B. D. O. Anderson, "Reaching a consensus in a dynamically changing environment: a graphical approach," *SIAM J. Control Optim.*, vol. 47, no. 2, 575-600, 2008.

[12] J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Distributed anonymous discrete function computation," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2276-2289, 2011.

[13] R. Olfati-Saber and R. Murray, "Consensus problems in the networks of agents with switching topology and time delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520-1533, 2004.

[14] J. Fax and R. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Automatic Control*, vol. 49, no. 9, 1465-1476, 2004.

[15] H. G. Tanner, A. Jadbabaie, G. J. Pappas, "Flocking in fixed and switching networks," *IEEE Trans. Autom. Control*, vol. 52, no.5, pp. 863-868, 2007.

[16] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988-1001, 2003.

[17] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803-812, 1986.

[18] W. Ren and R. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control*, Springer-Verlag, London, 2008.

[19] W. Ren and R. Beard, "Consensus seeking in multi-agent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655-661, 2005.

[20] G. Shi and Y. Hong, "Global target aggregation and state agreement of nonlinear multi-agent systems with switching topologies," *Automatica*, vol. 45, no. 5, pp. 1165-1175, 2009.

[21] G. Shi, Y. Hong and K. H. Johansson, "Connectivity and set tracking of multi-agent systems guided by multiple moving leaders," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 663-676, 2012.

[22] G. Shi and K. H. Johansson,"Multi-agent robust consensus-Part I: convergence analysis," IEEE CDC, pp. 5744-5749, 2011.

[23] Z. Lin, B. Francis, and M. Maggiore, "State agreement for continuous-time coupled nonlinear systems," *SIAM J. Control Optim.*, vol. 46, no. 1, pp. 288-307, 2007.

[24] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, no. 2, pp. 169-182, 2005.

[25] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Proc. IEEE Conf. on Decision and Control, Cancun, Mexico*, pp. 4177-4184, 2008.

[26] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48-61, 2009.

[27] A. Nedić, A. Ozdaglar and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922-938, 2010.

[28] G. Shi and K. H. Johansson, "Randomized optimal consensus of multi-agent systems," *Automatica*, to appear.

[29] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *IPSN'04*, pp. 20-27, 2004.

[30] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proc. IEEE Conf. on Decision and Control*, Cancun, Mexico, pp. 4185-4190, 2008.

[31] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Stochastic incremental gradient descent for estimation in sensor networks," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, pp. 582-586, 2007.

[32] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, "Gossip algorithms for convex consensus optimization over networks," *IEEE Trans. Autom. Control*, vol. 56, pp. 2917-2923, 2011.

[33] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case," *IEEE Trans. Autom. Control*, vol. 59, no. 7, pp. 2348-2354, 2012.

[34] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Incremental stochastic subgradient algorithms for convex optimization," *SIAM Journal on Optimization*, vol. 20, no. 2, 691-717, 2009.

[35] B. Johansson, M. Rabi and M. Johansson. "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1157-1170, 2009.

[36] D. Jakovetić, J. Xavier and J. M. F. Moura, "Cooperative convex optimization in networked systems: augmented lagrangian algorithms with directed gossip communication," http://arxiv.org/abs/1007.3706, 2011.

[37] K. Arrow, L. Hurwicz and H. Uzawa, *Studies in linear and non-linear programming*. Stanford University Press, 1958.

[38] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems," *Linear Algebra and its Applications*, vol. 146, pp. 79-91, 1991.

[39] J. Wang and N. Elia, "Control approach to distributed optimization," *The 48th Annual Allerton Conference*, pp. 557-561, 2010.

[40] H.-B. Dürr and C. Ebenbauer, "A smooth vector field for saddle point problems," *The 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 4654-4660, 2011.

[41] G. Shi, K. H. Johansson and Y. Hong, "Reaching an optimal consensus: dynamical systems that compute intersections of convex sets," *IEEE Transactions on Automatic Control*, to appear.