21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

# Distributed Multi-Agent Optimization with Local Constraints via a Subgradient Method with Delayed Information of Feasibility*

Izumi Masubuchi[a†], Jun Tsukamoto[b], Takayuki Wada[c], Ryosuke Morita[d], Toru Asai[e], Yuzo Ohta[a] and Yasumasa Fujisaki[c]

*Abstract*— **This paper proposes protocols for a distributed optimization problem to minimize the average of objective functions of the agents in the network with satisfying local constraints of each agent. To decide whether the agents update their decision variables toward the direction of the gradient of objective or constraint functions, the agents receive information on fulfillment of constraints from neighbor agents as well as sharing the decision variables. Then the agents can know whether all the constraints of the agents were satisfied some steps ago. Convergence of subgradient-method-based protocols is proved even with depending on such past information of constraints. Numerical examples are presented to illustrate the protocols.**

## I. INTRODUCTION

In large-scale systems such as energy networks, subsystems are interconnected through physical and/or information networks and have their own purposes, as well as those subsystems are needed to act so that the whole network system attains a certain objective. Distributed optimization methods based on asymptotic consensus making of multi-agent systems have been receiving a great deal of attention as a realistic method by which each subsystem or agent can behave for the purpose of the network system. Seminal works have been presented for consensus and distributed multi-agent optimization problems in e.g. [1]-[8].

In particular, subgradient methods [9], [10] are adopted for distributed optimization, with averaging of decision variables of agents between neighbor agents and projection onto a set that constrains the decision variables within it [3]-[6]. Zhu and Martínez [7] show algorithms based on projected primal-dual subgradient methods that can handle constraints given by equality and inequality constraints that are common among agents. If only inequality constraints are involved, their algorithm can include projections onto uncommon sets. Those previous methods are, however, not applicable if the agents need to satisfy inequality constraints that are not common and not shared among agents. One possible approach is to pass information on fulfillment of constraints at each time as well as decision variable for averaging. Then, in finite-sized networks, after certain steps all the agents in the network can know whether all the agents satisfied their own local constraints some steps ago.

In this paper, we utilize this information for distributed multi-agent optimization with local constraint functions whose feasible set is not common and not a priori known. A linear averaging protocol is presented that depends on delayed information on fulfillment of constraints, where the delay is due to passing information agent-by-agent through the network. We show consensus and convergence of sequences generated by the proposed protocols to an optimum of the average (or the sum) of each objective function of the agents subject to local constraints of agents, even though our protocols are based on delayed information. Numerical examples are presented to illustrate the protocols.

The rest of the paper is organized as follows. Section II formulates a distributed optimization problem with local constraints. In Section III, we show a protocol to share information of satisfaction of local constraints and a protocol to solve the distributed optimization problem based on delayed information on constraints. Section IV considers a slightly different protocol that does not depend on past gradients. In Section V, numerical examples are provided for the protocols of Sections III and IV. Lastly, we conclude this paper in Section VI.

### *Notation*

The set of integers and real numbers are denoted by $\mathbf{Z}$ and $\mathbf{R}$, respectively. $\mathbf{R}^n$ and $\mathbf{R}^{m\times n}$ are the sets of real (column) $n$-vectors and $m \times n$-matrices, respectively. $\|\cdot\|$ is the Euclidean norm for vectors and $\langle\cdot,\cdot\rangle$ is the standard inner product of $\mathbf{R}^n$. Inequalities on real vectors are elementwise in this paper.

## II. PROBLEM FORMULATION

Let $f^i, g^i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \dots, N$ be functions that satisfy the following assumptions:

*Assumption 1:* The functions $f^i, g^i, i = 1, \dots, N$ are subdifferentiable and convex on $\mathbf{R}^n$.

*Assumption 2:* The functions $f^i, g^i$, $i = 1, \dots, N$ are uniformly Lipschitz continuous on $\mathbf{R}^n$ whose Lipschitz

*This work is supported by the Japan Science and Technology Agency, CREST.

[a]Izumi Masubuchi and Yuzo Ohta are with Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada, Kobe 657-8501, Japan.
[b]Jun Tsukamoto was with Faculty of Engineering, Kobe University, 1-1 Rokkodai, Nada, Kobe 657-8501, Japan while this work was done. Currently he is with Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Kyoto 606-8501, Japan.
[c]Takayuki Wada and Yasumasa Fujisaki are with Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan.
[d]Ryosuke Morita was with Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan while this work was done. Currently he is with College of Science and Engineering, Aoyama Gakuin University, 5-10-1 Fuchinobe, Chuo, Sagamihara 252-5258, Japan.
[e]Toru Asai is with Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan.
†Corresponding author. E-mail: msb@harbor.kobe-u.ac.jp

ISBN: 978-90-367-6321-9

constants are no more than $c(>0)$, i.e.,

$$|f^i(x) - f^i(y)| \le c\|x-y\|,$$
$$|g^i(x) - g^i(y)| \le c\|x-y\|, \quad i = 1,\dots,N$$

hold for all $x, y \in \mathbf{R}^n$.

Let $x^i \in \mathbf{R}^n$, $i = 1,\dots,N$ and consider the following minimization problem of the average of $f^i(x^i)$ with constraints $g^i(x^i) \le 0$ and requirement of consensus:

$$\text{PA:} \min_{x^1,\dots,x^N \in \mathbf{R}^n} \frac{1}{N}\sum_{i=1}^N f^i(x^i) \quad \text{subject to}$$
$$g^i(x^i) \le 0, \quad i = 1,\dots,N, \quad x^1 = \dots = x^N. \tag{1}$$

*Assumption 3:* There exists an optimal solution to PA. Let $x^1 = \dots = x^N = x_*$ be any optimum of $PA$ and let $f_*$ be the optimum: $f_* = \frac{1}{N}\sum_{i=1}^N f^i(x_*)$ with $g^i(x_*) \le 0$, $i = 1,\dots,N$. We assume strict feasibility of the constraints:

*Assumption 4:* There exists a vector $x_s \in \mathbf{R}^n$ and a positive scalar $\mu_s > 0$ for which $g^i(x_s) \le -\mu_s$, $i = 1,\dots,N$.

To solve the problem PA in a distributed setting, consider a system of $N$ agents that belong to $\mathcal{A} = \{1,\dots,N\}$. Each agent $i \in \mathcal{A}$ is assigned with the decision variable vector $x^i \in \mathbf{R}^n$ and an additional data $\nu^i$, which we define later. We assume that each agent $i$ only knows the functions $f^i$ and $g^i$. The values of $x^i$ and $\nu^i$ at time $k \in \mathbf{Z}$ are denoted as $x^i_k$ and $\nu^i_k$, respectively. At time $k \in \mathbf{Z}$, an agent $i \in \mathcal{A}$ receives variables $(x^j_k, \nu^j_k)$ of other agents $j$ that belong to $\mathcal{J}^i \subset A$. Setting $\mathcal{E} = \{(j,i) : j \in \mathcal{J}^i, i \in \mathcal{A}\}$ defines a directed graph $\mathcal{G} = (\mathcal{A}, \mathcal{E})$.

*Assumption 5:* The graph $\mathcal{G}$ is strongly connected and balanced.

## III. A DISTRIBUTED OPTIMIZATION PROTOCOL

First, we consider averaging of the decision variables $x^i_k$ between agents, which is standard in distributed optimization with using linear protocols. Let $W = (w^{ij}) \in \mathbf{R}^{N\times N}$ be a doubly stochastic matrix that satisfies

$$W\mathbf{1} = \mathbf{1}, \quad \mathbf{1}^\top W = \mathbf{1}^\top, \tag{2}$$

where $\mathbf{1} = [\ 1 \ \ \cdots \ \ 1\ ]^\top \in \mathbf{R}^n$, and

$$\begin{cases} w^{ij} > 0 & \text{if } j \in \mathcal{J}^i \cup \{i\}, \\ w^{ij} = 0 & \text{if } j \notin \mathcal{J}^i. \end{cases} \tag{3}$$

The assumption that the graph $\mathcal{G}$ is strongly connected and balanced ensures that $W_k$ has an eigenvalue 1 with multiplicity 1 and the other eigenvalues are in $[0,1)$.

Since the problem PA involves constraints $g^i(x^i) \le 0$ that are *not* common among agents, protocols for distributed optimization with satisfying the constraints need some information on whether all of the constraints are satisfied or not. As well as the decision variable vectors $x^i_k$ are exchanged between agents, we consider sending data vectors $\nu^j$ from agents $j \in \mathcal{J}^i$ to agent $i$ at each $k$. Let $\nu^j$ have $m$ bit memories $\nu^j[\tau] \in \{0,1\}$, $\tau = 1,\dots,m$, where $m$ is the diameter of the graph $\mathcal{G}$ Each $\nu^i_k[\tau]$ is set 1 if agent $i$ was feasible at time $(k-m+\tau)$ and if the agent $i$ has not received information that any other agent's constraint was infeasible at time $(k-m+\tau)$ and otherwise set 0. More exactly, each agent $i$ initializes and updates $\nu^i_k[j]$, $j = 1,\dots,m$ below. Let $\xi^i_k$ be the vector generated by motion toward consensus:

$$\xi^i_k = \sum_{j=1}^N w^{ij} x^j_k, \quad k = 0,1,2,\dots. \tag{4}$$

Then $\nu^i_k$ is determined as follows:

- Initialization:

$$\nu^i_0[\tau] = 0, \quad \tau \in [1, m-1],$$
$$\nu^i_0[m] = \begin{cases} 1 & \text{if } g^i(\xi^i_0) \le 0, \\ 0 & \text{if } g^i(\xi^i_0) > 0. \end{cases}$$

- Update: for $k = 0,1,2,\dots$,

$$\nu^i_{k+1}[\tau] = \prod_{j \in \mathcal{J}^i \cup \{i\}} \nu^j_k[\tau+1], \quad \tau \in [1, m-1],$$
$$\nu^i_{k+1}[m] = \begin{cases} 1 & \text{if } g^i(\xi^i_k) \le 0, \\ 0 & \text{if } g^i(\xi^i_k) > 0. \end{cases}$$

To run this protocol, the agent $i$ needs to receive $\nu^j_k$'s that belong to $\mathcal{J}_i$ as well as $x^j_k$. Since the graph $\mathcal{G}$ is strongly connected, it obviously holds that, if $k \ge m$,

$$\begin{aligned} \eta^i_k &:= \prod_{j \in \mathcal{J}^i \cup \{i\}} \nu^j_k[1] \\ &= \begin{cases} 1 & \text{if } \forall i \in \{1,\dots,N\} \ \ g^i(\xi^i_{k-m}) \le 0, \\ 0 & \text{if } \exists i \in \{1,\dots,N\} \ \ g^i(\xi^i_{k-m}) > 0. \end{cases} \end{aligned} \tag{5}$$

Note that $\eta^i_k$'s are identical for all $i \in \mathcal{A}$ if $k \ge m$.

Next, we show a distributed optimization protocol that uses $\nu^i_k$. Consider the following update law of $x^i_k$:

$$x^i_{k+1} = \xi^i_k - b_k h^i_k, \quad k = 0,1,2,\dots \tag{6}$$

where $\xi^i_k$ is given in (4). We set $h^i_k = 0$ for $k < m$ and

$$h^i_k \in \begin{cases} \partial f^i(\xi^i_{k-m}) & \text{if } \eta^i_k = 1, \\ \partial g^i(\xi^i_{k-m}) & \text{if } \eta^i_k = 0 \text{ and } g^i(\xi^i_{k-m}) > 0, \\ \{0\} & \text{otherwise} \end{cases} \tag{7}$$

for $k \ge m$. From Assumption 2, it holds that $\|h^i_k\| \le c$. We make the following assumption on the step size $b_k$.

*Assumption 6:* $\{b_k : k = 0,1,2,\dots\}$ is a sequence of positive numbers that is monotonically decreasing, square summable and not summable.

Define

$$\bar{x}_k := \frac{1}{N}\sum_{i=1}^N x^i_k, \quad \bar{h}_k := \frac{1}{N}\sum_{i=1}^N h^i_k,$$
$$x^\delta_k := \begin{bmatrix} x^1_k - x^2_k \\ \vdots \\ x^{N-1}_k - x^N_k \end{bmatrix}.$$

Then it holds that $\bar{x}_{k+1} = \bar{x}_k - b_k \bar{h}_k$ and for some positive constant $d > 0$ we have $\|x^i_k - x^j_k\| \le d\|x^\delta_k\|$, $i,j = 1,\dots,N$, $k = 0,1,2\dots$. From the assumptions on $W$, $b_k$ and $h^i_k$, $x^\delta_k$ is square summable [3].

To prove convergence of $x_k^i$ to an optimum $x_*$, following the proof of the standard constrained subgradient method [10], we assume

$$\exists \varepsilon > 0 \quad \exists k_0 \geq m \quad \forall k \geq k_0$$
$$\begin{cases} \exists i \in \{1, \ldots, N\} \ \ g^i(\xi_k^i) > 0 \ \text{ or} \\ \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} f^i(\xi_k^i) > f_* + \varepsilon \end{cases} \tag{8}$$

and will lead to a contradiction. First, difference of the variables $\xi^i$ at time $k$ and $(k-m)$ is bounded with error of consensus of $x^i$ and the step size $b_k$ as

$$\|\xi_k^i - \xi_{k-m}^i\| \leq d(\|x_k^\delta\| + \|x_{k-m}^\delta\|) + mcb_{k-1}. \tag{9}$$

This is immediately seen from

$$\begin{aligned} \left\|\xi_k^i - \bar{x}_k\right\| &= \left\|\frac{1}{N}\sum_{l=1}^{N}\sum_{j=1}^{N} w^{ij}(x_k^j - x_k^l)\right\| \\ &\leq \frac{1}{N}\sum_{l=1}^{N}\sum_{j=1}^{N} w^{ij}\|x_k^j - x_k^l\| \\ &\leq d\left\|x_k^\delta\right\| \end{aligned}$$

and

$$\|\bar{x}_k - \bar{x}_{k-m}\| = \left\|\sum_{l=k-m}^{k-1} b_l \bar{h}_l\right\| \leq mcb_{k-1}.$$

From Assumption 4, for any $\varepsilon > 0$, there exist $\mu > 0$ and $\tilde{x} \in \mathbf{R}^n$ that satisfy

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} f^i(\tilde{x}) &\leq f_* + \frac{\varepsilon}{2}, \\ g^i(\tilde{x}) &\leq -\mu. \quad i = 1, \ldots, N, \end{aligned} \tag{10}$$

Let $k \geq m$ and consider

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\|x_{k+1}^i - \tilde{x}\|^2 &= \frac{1}{N}\sum_{i=1}^{N}\left\|\xi_k^i - b_k h_k^i - \tilde{x}\right\|^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\|\xi_k^i - \tilde{x}\|^2 + \frac{1}{N}\sum_{i=1}^{N}\left\{-2b_k\left\langle h_k^i,\, \xi_k^i - \tilde{x}\right\rangle\right\} \\ &\quad + \frac{1}{N}\sum_{i=1}^{N} b_k^2\|h_k^i\|^2. \end{aligned} \tag{11}$$

The first term of the RHS is bounded as $\frac{1}{N}\sum_{i=1}^{N}\|\xi_k^i - \tilde{x}\|^2 \leq \frac{1}{N}\sum_{j=1}^{N}\|x^j - \tilde{x}\|^2$, which is deduced from convexity, and the third term as $\|h_k^i\| \leq c$. Denote

$$\iota_k^i = -2b_k\left\langle h_k^i,\, \xi_k^i - \tilde{x}\right\rangle \tag{12}$$

and observe that

$$\begin{aligned} \iota_k^i &= -2b_k\left\langle h_k^i, \xi_k^i - \tilde{x}\right\rangle \\ &= -2b_k\left\langle h_k^i, \xi_{k-m}^i - \tilde{x}\right\rangle + 2b_k\left\langle h_k^i, \xi_{k-m}^i - \xi_k^i\right\rangle \\ &\leq -2b_k\left\langle h_k^i, \xi_{k-m}^i - \tilde{x}\right\rangle + 2b_k c\|\xi_{k-m}^i - \xi_k^i\| \\ &\leq -2b_k\left\langle h_k^i, \xi_{k-m}^i - \tilde{x}\right\rangle + 2b_k e_k \end{aligned} \tag{13}$$

holds from (9), where $e_k$ is defined as

$$e_k := c\{d(\|x_k^\delta\| + \|x_{k-m}^\delta\|) + mcb_{k-1}\}. \tag{14}$$

Note that $e_k$ is square summable. Consider further $\iota_k^i$, which depends on the selection of $h_k^i$ and it is determined whether $\eta_k^i$, which we have seen coincides with each other for all $i$, is 0 or 1.

1) Suppose that $\eta_k^1 = \cdots = \eta_k^N = 1$. Then $h_k^i \in \partial f^i(\xi_{k-m}^i)$ from (7) and $g^i(\xi_{k-m}^i) \leq 0$, $i = 1, \ldots, N$ and hence it holds that

$$\frac{1}{N}\sum_{i=1}^{N} f^i(\xi_{k-m}^i) > f_* + \varepsilon$$

from (8). Combining this with (10), we obtain

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} f^i(\tilde{x}) - \frac{1}{N}\sum_{i=1}^{N} f^i(\xi_{k-m}^i) \\ &\qquad < f_* + \frac{\varepsilon}{2} - (f_* + \varepsilon) \leq -\frac{\varepsilon}{2}. \end{aligned}$$

Applying this inequality and (7) to (13), we derive

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N}\iota_k^i \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\left\{-2b_k(f^i(\xi_{k-m}^i) - f^i(\tilde{x})) + 2b_k e_k\right\} \\ &\leq -b_k\varepsilon + 2b_k e_k. \end{aligned}$$

2) Suppose that $\eta_k^1 = \cdots = \eta_k^N = 0$. Then, from (7),

$$h_k^i \in \begin{cases} \partial g^i(\xi_{k-m}^i) & \text{if } g^i(\xi_{k-m}^i) > 0, \\ \{0\} & \text{otherwise.} \end{cases}$$

If $g^i(\xi_{k-m}^i) > 0$, applying (10) and (7) to (13), we see

$$\begin{aligned} \iota_k^i &\leq -2b_k(g^i(\xi_{k-m}^i) - g^i(\tilde{x})) + 2b_k e_k \\ &\leq -2b_k\mu + 2b_k e_k. \end{aligned}$$

If $g^i(\xi_{k-m}^i) \leq 0$, we have $\iota_k^i = 0$ from (12). Since at least one agent $i$ violates the constraint, we have

$$\frac{1}{N}\sum_{i=1}^{N}\iota_k^i \leq -b_k\frac{2\mu}{N} + 2b_k e_k.$$

Thus, setting $\varepsilon' = \min\{\varepsilon, 2\mu/N\} > 0$, we see

$$\begin{aligned} &(\text{RHS of (11)}) \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\|x_k^i - \tilde{x}\|^2 - b_k\varepsilon' + 2b_k e_k + c^2 b_k^2 \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\|x_k^i - \tilde{x}\|^2 - b_k\varepsilon' + e_k^2 + (c^2+1)b_k^2 \end{aligned}$$

and from this we have

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\|x_{k_1+1}^i - \tilde{x}\|^2 &\leq \frac{1}{N}\sum_{i=1}^{N}\|x_{k_0}^i - \tilde{x}\|^2 - \varepsilon'\sum_{k=k_0}^{k_1} b_k \\ &\quad + \sum_{k=k_0}^{k_1}\{e_k^2 + (c^2+1)b_k^2\}. \end{aligned} \tag{15}$$

Since the LHS is not negative, it holds that

$$\varepsilon' \leq \frac{\frac{1}{N}\sum_{i=1}^{N}\|x_{k_0}^i - \tilde{x}\|^2 + \sum_{k=k_0}^{k_1}\{e_k^2 + (c^2+1)b_k^2\}}{\sum_{k=k_0}^{k_1} b_k},$$

which implies $\varepsilon' = 0$ because the numerator of the RHS consists of constant and square summable terms and the denominator is not summable. Thus this contradicts $\varepsilon' > 0$ and hence (8) is false, which means that

$$\forall \varepsilon > 0 \quad \forall k_0 \geq m \quad \exists k \geq k_0$$
$$g^i(\xi_k^i) \leq 0, \;\; i = 1, \ldots, N, \;\; \frac{1}{N}\sum_{i=1}^{N} f^i(x_k^i) \leq f_* + \varepsilon.$$

Now we conclude the following.

*Theorem 1:* Suppose that Assumptions 1-6 hold. Then, for any initial value $(x_0^1, \ldots, x_0^N)$, the protocol (6) generates a sequence $\{(\xi_k^1, \ldots, \xi_k^N),\, k = 0, 1, 2, \ldots\}$ that satisfies

$$\lim_{k\to\infty}(\xi_k^i - \xi_k^j) = 0, \quad i, j = 1, \ldots, N. \tag{16}$$

Moreover, there exists a subsequence of $\xi_k^i$ represented by a strictly increasing integer sequence $k_l, l = 0, 1, 2, \ldots$ such that

$$g^i(\xi_{k_l}^i) \leq 0, \quad i = 1, \ldots, N, \quad l = 0, 1, 2, \ldots, \tag{17}$$
$$\liminf_{l\to\infty} \frac{1}{N}\sum_{i=1}^{N} f^i(\xi_{k_l}^i) = f_*. \tag{18}$$

## IV. A PROTOCOL THAT USES THE NEWEST SUBGRADIENT

The protocol in the previous section takes subgradients at time $(k - m)$ since the agents know at time $k$ whether all of the constraints $g^i(x^i) \leq 0$ hold at time $(k - m)$. In this section, we consider a different protocol that only uses subgradients at time $k$ to compute of $x_{k+1}^i$. This makes each agent free from storing subgradients for $m$ steps, while the protocol then depends on subgradients at a point $x_k^i$ that may not be feasible.

Now consider again the problem PA and the protocol (6) with the gradient definitions (7) modified for $k \geq m$ as follows:

$$h_k^i \in \begin{cases} \partial f^i(\xi_k^i) & \text{if } \eta_k^i = 1, \\ \partial g^i(\xi_k^i) & \text{if } \eta_k^i = 0 \text{ and } g^i(\xi_k^i) > 0, \\ \{0\} & \text{otherwise.} \end{cases} \tag{19}$$

We require the following additional assumption on $b_k$:

*Assumption 7:* In addition to Assumption 6, we assume that, every subsequence $b_{k_l}$ with $k_l$ satisfying $k_{l+1} - k_l \leq m + 1$ is also not summable.

*Remark 1:* It is easy to see that $b_k = 1/(k + 1)$ satisfies Assumptions 6 and 7. In fact, since $k_l \leq k_0 + l(m + 1)$,

$$\sum_{l=1}^{p} b_{k_l} = \sum_{l=1}^{p} \frac{1}{k_l + 1} \geq \sum_{l=1}^{p} \frac{1}{k_0 + 1 + l(m+1)} \to \infty$$

as $p \to \infty$.

To prove convergence, we assume the following condition slightly different from (8) that utilized to obtain Theorem 1:

$$\exists \varepsilon > 0 \quad \exists k_0 \geq m \quad \forall k \geq k_0$$
$$\begin{cases} \exists i \in \{1, \ldots, N\} \;\; g^i(\xi_k^i) > e_k \;\; \text{or} \\ \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} f^i(\xi_k^i) > f_* + \varepsilon, \end{cases} \tag{20}$$

where $e_k$ is as defined in (14). Using

$$|f^i(\xi_k^i) - f^i(\xi_{k-m}^i)| \leq e_k,$$
$$|g^i(\xi_k^i) - g^i(\xi_{k-m}^i)| \leq e_k, \quad i = 1, \ldots, N, \tag{21}$$

we manipulate $\iota_k^i$ defined in (12) as follows:

1) If $\eta_k^1 = \cdots = \eta_k^N = 1$, then $h_k^i \in \partial f^i(\xi_k^i)$ from (19) and $g^i(\xi_{k-m}^i) \leq 0$, $j = 1, \ldots, N$. Let

$$b_k' = b_k$$

for later use. From (21),

$$g^i(\xi_k^i) \leq g^i(\xi_{k-m}^i) + e_k \leq e_k$$

and hence

$$\frac{1}{N}\sum_{i=1}^{N} f^i(\xi_k^i) > f_* + \varepsilon$$

holds from (20). Combining this with (10), we get

$$\frac{1}{N}\sum_{i=1}^{N} f^i(\tilde{x}) - \frac{1}{N}\sum_{i=1}^{N} f^i(\xi_k^i)$$
$$< f_* + \frac{\varepsilon}{2} - (f_* + \varepsilon) = -\frac{\varepsilon}{2}.$$

From this inequality and (19), we can deduce

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} \iota_k^i &\leq \frac{1}{N}\sum_{i=1}^{N} \left\{-2b_k(f^i(\xi_k^i) - f^i(\tilde{x}))\right\} \\ &\leq -b_k\varepsilon = -b_k'\varepsilon. \end{aligned}$$

2) If $\eta_k^1 = \cdots = \eta_k^N = 0$, then, from (19),

$$h_k^i \in \begin{cases} \partial g^i(\xi_k^i) & \text{if } \eta_k^i = 0 \text{ and } g^i(\xi_k^i) > 0, \\ \{0\} & \text{otherwise.} \end{cases}$$

If $g^i(\xi_k^i) > 0$,

$$\begin{aligned} \iota_k^i &\leq -2b_k(g^i(\xi_k^i) - g^i(\tilde{x})) \\ &< 2b_k g^i(\tilde{x}) \;\leq\; -2b_k\mu. \end{aligned}$$

If $g^i(\xi_k^i) \leq 0$, we have $\iota_k^i = 0$ from (12).
Next, if there exists at least one $i$ for which $g^i(\xi_k^i) > 0$, let $b_k' = b_k$. We have

$$\frac{1}{N}\sum_{i=1}^{N} \iota_k^i \leq -2b_k\frac{\mu}{N} = -2b_k'\frac{\mu}{N}.$$

Otherwise, i.e., if $g^i(\xi_k^i) \leq 0$ for all $i \in \mathcal{A}$, let $b_k' = 0$. We can write

$$\frac{1}{N}\sum_{i=1}^{N} \iota_k^i = 0 \leq -2b_k'\frac{\mu}{N}.$$

Now consider $b'_k$ defined above. Suppose that $b'_k = 0$ for some $k$. This means that $g^i(\xi^i_k) \le 0$ for all $i$. Then, the definition of $\eta^i_k$ implies that $\eta^1_{k+m} = \cdots = \eta^N_{k+m} = 1$ and hence $b'_{k+m} = b_{k+m}$. Thus $b'_k$ coincides with $b_k$ at least once every $(m+1)$ iterations. Therefore we can see that, from Assumption 7, $\sum_{k=k_0}^{k_1} b'_k \to \infty$ as $k_1 \to \infty$.

Similarly to the previous section, we can derive

$$\varepsilon' \le \frac{\frac{1}{N}\sum_{i=1}^{N}\|x^i_{k_0} - \tilde{x}\|^2 + \sum_{k=k_0}^{k_1} c^2 b_k^2}{\sum_{k=k_0}^{k_1} b'_k},$$

which allows as to conclude the following result.

*Theorem 2:* Suppose that Assumptions 1-7 hold. Then, for any initial value $(x^1_0, \ldots, x^N_0)$, the protocol (6) with $h^i_k$ defined by (19) instead of (7) generates a sequence $\{(\xi^1_k, \ldots, \xi^N_k),\ k = 0, 1, 2, \ldots\}$ that satisfies

$$\lim_{k\to\infty}(\xi^i_k - \xi^j_k) = 0, \quad i, j = 1, \ldots, N \tag{22}$$

and there exist a subsequence of $\xi^i_k$ represented by a strictly increasing integer sequence $k_l, l = 0, 1, 2, \ldots$ and a sequence of positive numbers $e_k$, $k = 0, 1, 2, \ldots$ which is square summable (and hence converges to zero as $k \to \infty$) such that

$$g^i(\xi^i_{k_l}) \le e_{k_l}, \quad i = 1, \ldots, N, \quad l = 0, 1, 2, \ldots, \tag{23}$$

$$\liminf_{l\to\infty} \frac{1}{N}\sum_{i=1}^{N} f^i(\xi^i_{k_l}) = f_*. \tag{24}$$

## V. Numerical examples

Let $N = 5$, $n = 3$ and consider the following objective functions

$$\begin{aligned} f^1(x) &= 2x_1 + 6x_2 + 2x_3 + 2, \\ f^2(x) &= x_1 - 3x_2 + 5x_3 + 8, \\ f^3(x) &= -3x_1 + x_2 - 4x_3 + 1, \\ f^4(x) &= -2x_1 - 2x_2 + 7x_3 + 1, \\ f^5(x) &= -3x_2 - 5x_3 + 3 \end{aligned}$$

and constraints

$$g^i(x) = G^i \begin{bmatrix} x_1 & x_2 & x_3 & 1 \end{bmatrix}^\top + \tilde{g}^i(x) \le 0,$$

where

$$G^1 = \begin{bmatrix} 3 & 7 & 9 & -9 \\ -4 & 1 & -5 & -8 \\ 9 & -8 & 2 & -13 \end{bmatrix},$$

$$G^2 = \begin{bmatrix} -1 & -4 & 3 & -6 \\ 2 & 2 & 5 & -13 \\ 1 & -8 & 0 & -8 \end{bmatrix},$$

$$G^3 = \begin{bmatrix} 8 & -3 & 2 & -2 \\ -2 & 5 & 6 & -9 \\ -1 & 0 & 3 & -5 \end{bmatrix},$$

$$G^4 = \begin{bmatrix} -1 & 5 & 6 & -5 \\ -2 & -2 & -5 & -8 \\ -1 & -8 & 9 & -3 \end{bmatrix},$$

$$G^5 = \begin{bmatrix} 8 & 0 & 0 & -7 \\ 0 & 1 & -5 & -4 \\ 1 & 1 & 1 & -10 \end{bmatrix}$$

and

$$\begin{aligned} &\left[\ \tilde{g}^1(x) \mid \cdots \mid \tilde{g}^5(x)\ \right] \\ &= \left[\begin{array}{c|c|c|c|c} (x^1)^2 & 0 & 0 & (x^1)^2 & 0 \\ 0 & (x^2)^2 & 0 & 0 & (x^2)^2 \\ 0 & 0 & (x^3)^2 & (x^3)^2 & 0 \end{array}\right]/10. \end{aligned}$$

The optimum is $f_* = 2.0008$ with $x^1 = x^2 = x^3 = x_* = [\ 0.5417 \quad 0.2955 \quad -0.7235\ ]^\top$. These are obtained by solving $\min_{x\in\mathbf{R}^n} \frac{1}{N}\sum_{i=1}^{N} f^i(x)$ subject to $g^i(x) \le 0$, $i = 1, \ldots, N$. We set $\mathcal{J}_1 = \{2\}$, $\mathcal{J}_2 = \{1, 3, 4\}$, $\mathcal{J}_3 = \{2, 4, 5\}$, $\mathcal{J}_4 = \{2, 3, 5\}$ and $\mathcal{J}_5 = \{3, 4\}$ and $W$ as

$$W = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 \end{bmatrix}/4.$$

The maximal length $m$ of the shortest paths from one agent to another is 3. The initial values are set as

$$\left[\ x^1_0 \mid \cdots \mid x^5_0\ \right] = \left[\begin{array}{c|c|c|c|c} -1 & -2 & 0 & -1 & -1 \\ -2 & 0 & 0 & -2 & -1 \\ -1 & -1 & 0 & -1 & 0 \end{array}\right].$$

Lastly, $b_k = 1/(k+1)$ and the number of iterations is fixed to $10^4$. Not that, though in this example $g^i(x)$'s are not uniformly Lipschitz, the results of the paper are valid eventually if the generated trajectories of $\xi^i_k$ are bounded.

The execution of the protocols in Section III with these settings is shown by Figures 1–3. Figure 1 shows the trajectories of $x^i_k$, which converge to $x_*$. Figures 2 and 3 show the objective function $(1/5)\sum_{i=1}^{5} f^i(\xi^i_k)$ and the constraints $g^i(\xi^i_k)$, respectively. We can see that the objective function approaches to $f_*$ and constraints are nearly satisfied. In fact, $g^3(x) \le 0$ and $g^5(x) \le 0$ are the active constraints of this problem and, for $k \in [9000, 10000]$, $g^3(\xi^3_k) \in [-0.0071, 0.0007]$ and $g^5(\xi^5_k) \in [-0.0025, 0.0018]$. In the same span of $k$, we have $(1/5)\sum_{i=1}^{5} f^i(\xi^i_k) \in [2.0019, 2.0068]$.

We also show results of the protocol in Section IV in Figures 4–6. The trajectories approaches to the optimal point, with $g^3(\xi^3_k) \in [-0.0020, 0.0004]$, $g^5(\xi^5_k) \in [-0.0004, 0.0010]$ and $(1/5)\sum_{i=1}^{5} f^i(\xi^i_k) \in [2.0163, 2.0188]$ for $k \in [9000, 10000]$.

## VI. Conclusion

In this paper, we showed protocols to solve a distributed optimization problem to minimize the average of objective functions of the agents in the network with satisfying local constraints of each agent. Sharing information by passing $m$-bit data on fulfillment of constraints, the protocols generate sequences of decision variables that are proved to make consensus between agents and converge in the sense of limit infimum to the optimal value of the problem.

For the sake of conciseness, this paper did not consider time-varying topology of networks and projection to known constraints. However, based on the fruitful results on distributed multi-agent optimizations (e.g., [3]), the results of this paper are expected to be extended to problems with these issues.

## References


[1] A. Jadbabaie, J. Lin and A. S. Morse: Coordination of groups of mobile autonomous agents using nearest neighbor rules, *IEEE Transactions on Automatic Control*, Vol. 48, No. 6, pp. 988-1001 (2003)

[2] R. Olfati-Saber and R. M. Murray: Consensus problems in networks of agents with switching topology and time-delays: *IEEE Transactions on Automatic Control*, Vol. 49, No. 9, pp. 1520-1533 (2004)

[3] A. Nedić and A. Ozdaglar: Distributed subgradient methods for multi-agent optimization, *IEEE Transactions on Automatic Control*, Vol. 54, No. 1, pp. 48-61 (2009)

[4] A. Nedić, A. Ozdaglar and P. A. Parrilo: Constrained consensus and optimization in multi-agent networks, *IEEE Transactions on Automatic Control*, Vol. 55, No. 4, pp. 922-938 (2010)

[5] K. Srivastava, A. Nedić and D. Stipanović: Distributed min-max optimization in networks, *Proceedings of the 17th Digital Signal Processing Conference* (2011)

[6] K. Srivastava, A. Nedić and D. Stipanović: Distributed Bregman-distance algorithms for min-max optimization, *Agent-Based Optimization*, Studies in Computational Intelligence, Vol. 456, Springer, pp. 143-174 (2013)

[7] M. Zhu, S. Martínez: On distributed convex optimization under inequality and equality constraints, *IEEE Transactions on Automatic Control*, Vol. 57, No. 1, pp. 151-164 (2012)

[8] R. Morita and Y. Fujisaki: Distributed minimax optimization via multi-agent systems, *Proceedings of the SICE Annual Conference 2013*, MoAT2.3 (2013)

[9] D. P. Bertsekas: *Nonlinear Programming*, Athena Scientific (1999)

[10] S. Boyd and A. Mutapcic: Subgradient methods, Notes for EE364b, Stanford University, http://www.stanford.edu/class/ee364b/notes/subgrad_method_notes.pdf (2007)


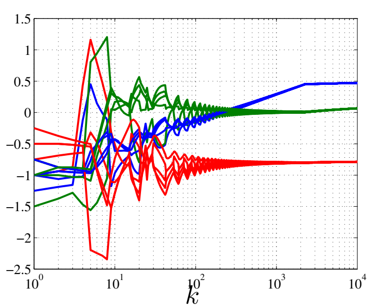

Fig. 1. $x_k^i$ / protocol of Section III

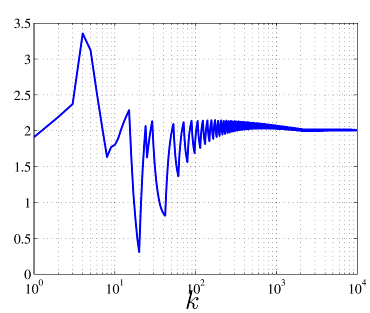

Fig. 2. Objective func. / protocol of Section III

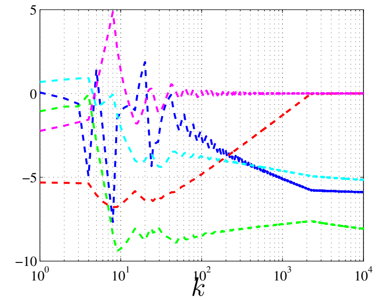

Fig. 3. Constraints / protocol of Section III

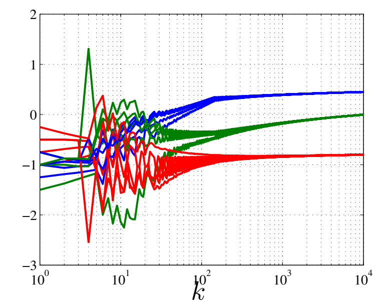

Fig. 4. $x_k^i$ / protocol of Section IV

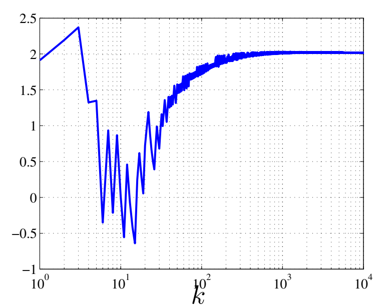

Fig. 5. Objective func. / protocol of Section IV

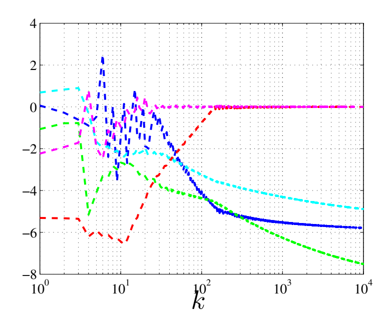

Fig. 6. Constraints / protocol of Section IV