Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2014, Article ID 618641, 6 pages
http://dx.doi.org/10.1155/2014/618641

Research Article

Asynchronous Gossip-Based Gradient-Free Method for Multiagent Optimization

Deming Yuan

College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

Correspondence should be addressed to Deming Yuan; dmyuan1012@gmail.com

Received 12 January 2014; Accepted 11 March 2014; Published 2 April 2014

Academic Editor: Hao Shen

This paper considers the constrained multiagent optimization problem. The objective function of the problem is a sum of convex functions, each of which is known by a specific agent only. For solving this problem, we propose an asynchronous distributed method that is based on gradient-free oracles and gossip algorithm. In contrast to the existing work, we do not require that agents be capable of computing the subgradients of their objective functions and coordinating their step size values as well. We prove that with probability 1 the iterates of all agents converge to the same optimal point of the problem, for a diminishing step size.

1. Introduction

In recent years, the problem of solving convex optimization problems over a network has attracted a lot of research attention; see [1–18]. The objective function of the problem is a sum of convex functions, each of which is known by a specific agent only. Such problems arise in many real applications including distributed finite-time optimal rendezvous [2] and distributed regression over sensor networks [5]. The methods that are designed for solving these optimization problems need to be fully distributed; that is, there does not exist a central coordinator.

In this paper, we propose an asynchronous gossip-based gradient-free method for solving the convex optimization problem over a multiagent network. The method is based on the gossip algorithm [19] and the gradient-free oracles [20]. The method is asynchronous in the sense that only one agent communicates at a given time, in contrast to the synchronous methods where all agents communicate simultaneously. Moreover, the method does not rely on the assumption that the information of the subgradients of the objective function is available. As is well known that for a variety of reasons there have been many instances where derivatives of the objective functions are unavailable or computationally expensive to calculate [20, 21].

Literature Review. In [3], the authors study the problem of minimizing a sum of multiple convex functions, each of which is known to one specific agent only. The authors use the average consensus algorithm in the literature on multiagent systems (see, e.g., [19, 22–26]) as a mechanism to develop a distributed subgradient method for solving the optimization problem; the convergence of the method is also given for a constant step size. The authors in [7] further take the global equality and inequality constraints into consideration. The work in [2] proposes a variant of the distributed subgradient method in [3], in which at each iteration several consensus steps are executed, which simplifies the proof of the convergence of the method. Inspired by the work in [2], the authors in [6] further incorporate the global inequality constraints. The aforementioned methods are synchronous because they require that all agents in the network update at the same time. To overcome this limitation, the work in [14] develops an asynchronous distributed algorithm, based on the gossip algorithm. The algorithm is asynchronous in the sense that only one agent communicates at a given time. Moreover, all agents use different step size values and they do not require any coordination of the agents. In [5], the author further removes the need for bidirectional communications of the asynchronous algorithm in [14]; the convergence of the algorithm is also established. The aforementioned methods or algorithms, however, rely on the assumption that the subgradients of the objective functions are available to each agent, respectively.

By comparison to previous work, the main contributions of this paper are twofold: (i) different from the methods or algorithms considered in existing papers, which rely on computing the subgradients of each agent's objective function, we propose the derivative-free method which is based on utilizing the random gradient-free oracles; (ii) the proposed method is asynchronous, in the sense that all agents use different step size values that do not require any coordination of the agents. We prove that with probability 1 the iterates of all agents converge to the same optimal point of the problem, for a diminishing step size.

Notation and Terminology. Let $\mathbb{R}^d$ be the d-dimensional vector space. We denote the standard inner product on $\mathbb{R}^d$ by $\langle a, b\rangle = \sum_{i=1}^d a_i b_i$, for $a, b \in \mathbb{R}^d$. We write $\|x\|$ to denote the Euclidean norm of a vector x and $\Pi_{\mathcal{X}}[x]$ to denote the Euclidean projection of a vector x on $\mathcal{X}$. We use x^{T} to denote the transpose of x. For a matrix P, $[\mathsf{P}]_{ij}$ represents the element in the ith row and jth column of P, and P^{T} represents its transpose. We use $\mathbb{E}[x]$ to denote the expected value of a random variable x. For a function f, its gradient at a point x is represented by $\nabla f(x)$.

2. Problem Formulation

In this section, we start by describing the constrained multiagent optimization problem. Then, we provide some preliminary results on the gossip algorithm that we use in developing the method.

2.1. Constrained Multiagent Optimization. We consider the following constrained multiagent optimization problem:

$$\min_{x\in\mathcal{X}} f(x) \triangleq \sum_{i=1}^{N} f^i(x), \tag{1}$$

where $x \in \mathbb{R}^d$ is a decision vector; f^i: $\mathbb{R}^d \to \mathbb{R}$ is the convex objective function of agent i known only by agent i, and we assume that f^i is Lipschitz continuous over $\mathcal{X}$ with Lipschitz constant $L(f^i)$; $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty closed convex set. We denote the optimal set of problem (1) by $\mathcal{X}^*$, and we assume that it is nonempty. Note that in problem (1), each function f^i need not be differentiable.

2.2. Gossip Algorithm. The underlying network topology of problem (1) is denoted by $G = (V, E)$, where $V = \{1, \dots, N\}$ is the node set and E is the set of links $\{i, j\}$ with $i \neq j$ and $\{i, j\} \in E$ only if there is a link between agents i and j. We assume that the network G is fixed, undirected, and connected.

In the paper, we utilize gossip algorithm as a mechanism to design the method. To be specific, at each time instant, agent i is chosen with probability $1/N$, and then with some positive probability, agent i communicates with one of its neighbors agent j. The iterations evolve as follows: for $k \geq 0$,

$$x_{k+1}^i = x_{k+1}^j = \frac{1}{2}x_k^i + \frac{1}{2}x_k^j \tag{2}$$

and for agents s that do not belong to $\{i, j\}$, update

$$x_{k+1}^s = x_k^s. \tag{3}$$

3. Gossip-Based Gradient-Free Method

In this section, motivated by the random gradient-free method in [20] and the gossip algorithm in [19], we present an asynchronous gossip-based gradient-free method for solving problem (1). We use $\mathtt{I}_{k+1}$ to denote the index of the agent that is chosen to update at time $k+1$ and J_{k+1} the index of the agent communicating with agent $\mathtt{I}_{k+1}$. The method is given as follows.

Gossip-Based Gradient-Free Method with a Diminishing Step Size

Initialize: choose random $x_0^i \in \mathcal{X}$, $\forall i \in V$.
Iteration ($k \geq 0$):

(i) for $i \in \{\mathtt{I}_{k+1}, \mathsf{J}_{k+1}\}$:

(1) compute $\varphi_{k+1}^i = (1/2)x_k^{\mathtt{I}_{k+1}} + (1/2)x_k^{\mathsf{J}_{k+1}}$;

(2) compute $x_{k+1}^i = \Pi_{\mathcal{X}}[\varphi_{k+1}^i - \sigma_k^i \mathsf{G}_{\mu_k^i}(x_k^i)]$, where $\sigma_k^i = (\Sigma_k^i)^{-1}$, and Σ_k^i denotes the number of updates that agent i has performed until time k, inclusively, and $\mathsf{G}_{\mu^i}(x_k^i)$ is the random gradient-free oracle, given by

$$\mathsf{G}_{\mu_k^i}\left(x_k^i\right) = \frac{f^i\left(x_k^i + \mu_k^i \nu_k^i\right) - f^i\left(x_k^i\right)}{\mu_k^i}\nu_k^i, \tag{4}$$

where $\mu_k^i = \mu\sigma_k^i$, and μ is a positive constant; ν_k^i is a random variable generated locally according to the Gaussian distribution.

(ii) For $i \notin \{\mathtt{I}_{k+1}, \mathsf{J}_{k+1}\}$: $x_{k+1}^i = x_k^i$.

We use $\mathcal{F}_k$ to denote the σ-field generated by the entire history of the random variables to iteration k; that is,

$$\mathcal{F}_k = \left\{x_0^i, i \in V\right\} \cup \left\{\mathtt{I}_{s+1}, \mathsf{J}_{s+1}, \nu_s^{\mathtt{I}_{s+1}}, \nu_s^{\mathsf{J}_{s+1}}; 0 \leq s \leq k-1\right\}, \tag{5}$$

where $\mathcal{F}_0 = \{x_0^i, i \in V\}$.

The method can be presented in a more compact form, by defining the following weight matrix:

$$\mathsf{W}_{k+1} = I - \frac{1}{2}\left(e_{\mathtt{I}_{k+1}} - e_{\mathsf{J}_{k+1}}\right)\left(e_{\mathtt{I}_{k+1}} - e_{\mathsf{J}_{k+1}}\right)^{\mathsf{T}}, \quad k \geq 0, \tag{6}$$

where I is the identity matrix and $e_i \in \mathbb{R}^N$ denotes the ith standard basis vector. It is easy to see that $\mathsf{W}_{k+1} \in \mathbb{R}^{N\times N}$ is doubly stochastic.

Now we can write the method as follows: for all $k \geq 0$ and any $i \in V$,

$$\varphi_{k+1}^{i} = \sum_{j=1}^{N} [\mathsf{W}_{k+1}]_{ij} x_k^j,$$
$$x_{k+1}^{i} = \varphi_{k+1}^{i} + \left[\Pi_{\mathcal{X}}\left[\varphi_{k+1}^{i} - \sigma_k^i \mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right] - \varphi_{k+1}^{i}\right] \times \mathbf{1}_{\{i\in\{\mathsf{I}_{k+1},\mathsf{J}_{k+1}\}\}}, \tag{7}$$

where $\mathbf{1}_{\{i\in\{\mathsf{I}_{k+1},\mathsf{J}_{k+1}\}\}}$ is the indicator function of the event $\{i \in \{\mathsf{I}_{k+1}, \mathsf{J}_{k+1}\}\}$. For the gradient-free oracle $\mathsf{G}_{\mu_k^i}(x_k^i)$, we have the following lemma, which is adopted from [20].

Lemma 1. *For each $i \in \{\mathsf{I}_{k+1}, \mathsf{J}_{k+1}\}$ and all $k \geq 0$, one has the following:*

(a) $\mathbb{E}[\mathsf{G}_{\mu_k^i}(x_k^i) \mid \mathcal{F}_k, \mathsf{I}_{k+1}, \mathsf{J}_{k+1}] = \nabla f^i_{\mu_k^i}(x_k^i)$, *where* $f^i_{\mu_k^i}(x) = (1/\kappa)\int_{\mathbb{R}^d} f^i(x + \mu_k^i\xi)e^{-(1/2)\|\xi\|^2}d\xi$ *with* $\kappa = \int_{\mathbb{R}^d} e^{-(1/2)\|\xi\|^2}d\xi = (2\pi)^{d/2}$, *and it satisfies:*

$$f^i(x) \leq f^i_{\mu_k^i}(x) \leq f^i(x) + \mu_k^i\sqrt{d}L\left(f^i\right). \tag{8}$$

(b) $\mathbb{E}[\|\mathsf{G}_{\mu_k^i}(x_k^i)\|^2 \mid \mathcal{F}_k, \mathsf{I}_{k+1}, \mathsf{J}_{k+1}] \leq (d+4)^2 L^2(f^i)$.

Remark 2. Note that method (7) is asynchronous, in the sense that to implement the method, each agent need not coordinate its step size with the step sizes of its neighbors; the time-varying parameters μ_k^i ($k \geq 0, i \in V$) share the same feature. In addition, to implement the method (7), the information of subgradients of the objective functions is not needed; however, each agent only needs to make two function evaluations per iteration to get the gradient-free oracle.

Let $\mathcal{E}_k^i = \{i \in \{\mathsf{I}_k, \mathsf{J}_k\}\}$ be the event that agent i updates at time k and π^i the probability of event $\mathcal{E}_k^i$. It is easy to see that

$$\pi^i = \frac{1}{N} + \frac{1}{N}\sum_{j\in N^i}\pi_{ji}, \tag{9}$$

where N^i denotes the set that contains all agents that are neighboring to agent i and $\pi_{ji} > 0$ denotes the probability that agent i is chosen by its neighbor j to communicate. In the paper, we denote $\breve{\pi} = \min_{i\in V}\pi^i$ and $\widehat{\pi} = \max_{i\in V}\pi^i$, respectively. There is an interesting link between the step size $\sigma_k^i = (\Sigma_k^i)^{-1}$ and the probability π^i that agent i updates.

Lemma 3 (see [17]). *Let $\pi_{\min} = \min_{\{i,j\}\in E}\pi_{ij}$. Let $\sigma_k^i = (\Sigma_k^i)^{-1}$ for all $k \geq 1$ and $i \in V$, and also let e be a scalar such that $0 < e < 1/2$. Then, there exists a large enough $\widetilde{k} = \widetilde{k}(e, N)$ such that with probability 1 for all $k \geq \widetilde{k}$ and $i \in V$,*

(a) $\sigma_k^i \leq 2/k\pi^i$;

(b) $(\sigma_k^i)^2 \leq 4N^2/k^2(1+\pi_{\min})^2$;

(c) $|\sigma_k^i - (1/k\pi^i)| \leq 2/k^{3/2-e}(1+\pi_{\min})^2$.

To establish the convergence of method (7), we also make use of the following lemma.

Lemma 4 (see [5]). *Let $\{u_k\}$, $\{v_k\}$, $\{a_k\}$, and $\{w_k\}$ be nonnegative random sequences such that for all $k \geq 1$, $\mathbb{E}[u_{k+1} \mid F_k] \leq (1+a_k)u_k - v_k + w_k$ with probability 1, where $F_k = \{\{u_s, v_s, a_s, w_s\}; 1 \leq s \leq k\}$. If $\sum_{k=1}^{\infty} a_k < \infty$ and $\sum_{k=1}^{\infty} w_k < \infty$ with probability 1, then, with probability 1, the sequence $\{u_k\}$ converges to some random variable and $\sum_{k=1}^{\infty} v_k < \infty$.*

We now present the main result of the paper, which is given in the following theorem.

Theorem 5. *Let $\{x_k^i\}$, $i \in V$, be the sequences generated by method (7) with $\sigma_k^i = (\Sigma_k^i)^{-1}$ and $\mu_k^i = \mu\sigma_k^i$, where μ is some positive constant. Assume that problem (1) has a nonempty optimal set $\mathcal{X}^*$. Also, assume that the sequence $\{\nu_k^i; i \in \{\mathsf{I}_{k+1}, \mathsf{J}_{k+1}\}\}$ is independent and identically distributed. Then the sequences $\{x_k^i\}$, $i \in V$, converge to the same random point in $\mathcal{X}^*$ with probability 1.*

Proof. For $k \geq 0$ and $i \in \{\mathsf{I}_{k+1}, \mathsf{J}_{k+1}\}$, we have for any $x \in \mathcal{X}$,

$$\begin{aligned}
\left\|x_{k+1}^i - x\right\|^2 &= \left\|\Pi_{\mathcal{X}}\left[\varphi_{k+1}^i - \sigma_k^i\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right] - x\right\|^2 \\
&\leq \left\|\varphi_{k+1}^i - \sigma_k^i\mathsf{G}_{\mu_k^i}\left(x_k^i\right) - x\right\|^2 \\
&\leq \left\|\varphi_{k+1}^i - x\right\|^2 + \left(\sigma_k^i\right)^2\left\|\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right\|^2 \\
&\quad - 2\sigma_k^i\left\langle\mathsf{G}_{\mu_k^i}\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle \\
&= \left\|\varphi_{k+1}^i - x\right\|^2 + \left(\sigma_k^i\right)^2\left\|\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right\|^2 \\
&\quad - \frac{2}{k\pi^i}\left\langle\mathsf{G}_{\mu_k^i}\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle \\
&\quad - 2\left(\sigma_k^i - \frac{1}{k\pi^i}\right)\left\langle\mathsf{G}_{\mu_k^i}\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle,
\end{aligned} \tag{10}$$

where the first inequality follows from the nonexpansive property of the projection operation. For $k \geq \widetilde{k}$, by recalling Lemma 3(c), with probability 1 the last term on the right-hand side of (10) can be bounded as follows:

$$\begin{aligned}
&- 2\left(\sigma_k^i - \frac{1}{k\pi^i}\right)\left\langle\mathsf{G}_{\mu_k^i}\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle \\
&\quad \leq \frac{2}{k^{3/2-e}(1+\pi_{\min})^2}\left(\left\|\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right\|^2 + \left\|\varphi_{k+1}^i - x\right\|^2\right).
\end{aligned} \tag{11}$$

Substituting the preceding inequality into (10) gives

$$\begin{aligned}
\left\|x_{k+1}^i - x\right\|^2 &\leq \left(1 + \frac{2}{k^{3/2-e}(1+\pi_{\min})^2}\right)\left\|\varphi_{k+1}^i - x\right\|^2 \\
&\quad + \left(\left(\sigma_k^i\right)^2 + \frac{2}{k^{3/2-e}(1+\pi_{\min})^2}\right)\left\|\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right\|^2 \\
&\quad - \frac{2}{k\pi^i}\left\langle\mathsf{G}_{\mu_k^i}\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle.
\end{aligned} \tag{12}$$

To simplify the notation, we denote $A_k = 2/k^{3/2-e}(1+\pi_{\min})^2$ and $B_k = 4N^2/k^2(1+\pi_{\min})^2 + A_k$; then from Lemma 3(b) and (12) it follows that with probability 1 for all $k \geq \widetilde{k}$ and $i \in \{\mathtt{I}_{k+1}, \mathtt{J}_{k+1}\}$,

$$\begin{aligned}\left\|x_{k+1}^i - x\right\|^2 \leq{}& (1+A_k)\left\|\varphi_{k+1}^i - x\right\|^2 + B_k\left\|\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right\|^2 \\ & - \frac{2}{k\pi^i}\left\langle \mathsf{G}_{\mu_k^i}\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle.\end{aligned} \tag{13}$$

Taking the conditional expectation on $\mathscr{F}_k$, $\mathtt{I}_{k+1}$ and $\mathtt{J}_{k+1}$ jointly yields

$$\begin{aligned}&\mathbb{E}\left[\left\|x_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k, \mathtt{I}_{k+1}, \mathtt{J}_{k+1}\right] \\ &\quad\leq (1+A_k)\left\|\varphi_{k+1}^i - x\right\|^2 \\ &\qquad + B_k\mathbb{E}\left[\left\|\mathsf{G}_{\mu_k^i}\left(x_k^i\right)\right\|^2 \mid \mathscr{F}_k, \mathtt{I}_{k+1}, \mathtt{J}_{k+1}\right] \\ &\qquad - \frac{2}{k\pi^i}\left\langle \mathbb{E}\left[\mathsf{G}_{\mu_k^i}\left(x_k^i\right) \mid \mathscr{F}_k, \mathtt{I}_{k+1}, \mathtt{J}_{k+1}\right], \varphi_{k+1}^i - x\right\rangle \\ &\quad\leq (1+A_k)\left\|\varphi_{k+1}^i - x\right\|^2 + B_k(d+4)^2L^2\left(f^i\right) \\ &\qquad - \frac{2}{k\pi^i}\left\langle \nabla f_{\mu_k^i}^i\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle,\end{aligned} \tag{14}$$

where the last inequality follows from using Lemma 1. For the last term on the right-hand side of the preceding inequality, we can derive

$$\begin{aligned}&-\frac{2}{k\pi^i}\left\langle \nabla f_{\mu_k^i}^i\left(x_k^i\right), \varphi_{k+1}^i - x\right\rangle \\ &\quad= -\frac{2}{k\pi^i}\left\langle \nabla f_{\mu_k^i}^i\left(x_k^i\right), x_k^i - x\right\rangle \\ &\qquad - \frac{2}{k\pi^i}\left\langle \nabla f_{\mu_k^i}^i\left(x_k^i\right), \varphi_{k+1}^i - x_k^i\right\rangle \\ &\quad\leq -\frac{2}{k\pi^i}\left[f_{\mu_k^i}^i\left(x_k^i\right) - f_{\mu_k^i}^i(x)\right] \\ &\qquad + \frac{2}{k\pi^i}\left\|\nabla f_{\mu_k^i}^i\left(x_k^i\right)\right\|\left\|\varphi_{k+1}^i - x_k^i\right\| \\ &\quad\leq -\frac{2}{k\pi^i}\left[f_{\mu_k^i}^i\left(x_k^i\right) - f_{\mu_k^i}^i(x)\right] \\ &\qquad + \frac{2}{k\pi^i}(d+4)L\left(f^i\right)\left\|\varphi_{k+1}^i - x_k^i\right\|,\end{aligned} \tag{15}$$

where in the last inequality we have use the bound $\|\nabla f_{\mu_k^i}^i(x_k^i)\| \leq (d+4)L(f^i)$, according to Lemma 1. Hence, substituting (15) into (14) yields

$$\begin{aligned}&\mathbb{E}\left[\left\|x_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k, \mathtt{I}_{k+1}, \mathtt{J}_{k+1}\right] \\ &\quad\leq (1+A_k)\left\|\varphi_{k+1}^i - x\right\|^2 + B_k(d+4)^2L^2\left(f^i\right) \\ &\qquad - \frac{2}{k\pi^i}\left[f^i\left(x_k^i\right) - f^i(x)\right] + \frac{2}{k\pi^i}\mu_k^i\sqrt{d}L\left(f^i\right) \\ &\qquad + \frac{2}{k\pi^i}(d+4)L\left(f^i\right)\left\|\varphi_{k+1}^i - x_k^i\right\|,\end{aligned} \tag{16}$$

where we have used the inequalities $f_{\mu_k^i}^i(x_k^i) \geq f^i(x_k^i)$ and $f_{\mu_k^i}^i(x) \leq f^i(x) + \mu_k^i\sqrt{d}L(f^i)$, based on Lemma 1(a). Using the fact that $\mu_k^i = \mu\sigma_k^i$ and Lemma 3(a), we obtain

$$\frac{2}{k\pi^i}\mu_k^i\sqrt{d}L\left(f^i\right) \leq \frac{4\mu}{k^2(\pi^i)^2}\sqrt{d}L\left(f^i\right) \tag{17}$$

which implies

$$\begin{aligned}&\mathbb{E}\left[\left\|x_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k, \mathtt{I}_{k+1}, \mathtt{J}_{k+1}\right] \\ &\quad\leq (1+A_k)\left\|\varphi_{k+1}^i - x\right\|^2 + B_k(d+4)^2L^2\left(f^i\right) \\ &\qquad + \frac{4\mu}{k^2(\pi^i)^2}\sqrt{d}L\left(f^i\right) - \frac{2}{k\pi^i}\left[f^i\left(x_k^i\right) - f^i(x)\right] \\ &\qquad + \frac{2}{k\pi^i}(d+4)L\left(f^i\right)\left\|\varphi_{k+1}^i - x_k^i\right\|.\end{aligned} \tag{18}$$

Taking the expectation with respect to $\mathscr{F}_k$ and using the fact the preceding inequality holds with probability π^i, and $x_{k+1}^i = \varphi_{k+1}^i$ with probability $1-\pi^i$, we obtain with probability 1 for all $k \geq \widetilde{k}$ and $i \in V$,

$$\begin{aligned}&\mathbb{E}\left[\left\|x_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k\right] \\ &\quad\leq \left(1+\pi^iA_k\right)\mathbb{E}\left[\left\|\varphi_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k\right] \\ &\qquad + \pi^iB_k(d+4)^2L^2\left(f^i\right) \\ &\qquad + \frac{4\mu}{k^2\pi^i}\sqrt{d}L\left(f^i\right) - \frac{2}{k}\left[f^i\left(x_k^i\right) - f^i(x)\right] \\ &\qquad + \frac{2}{k}(d+4)L\left(f^i\right)\mathbb{E}\left[\left\|\varphi_{k+1}^i - x_k^i\right\| \mid \mathscr{F}_k\right].\end{aligned} \tag{19}$$

Summing the above inequality for $i = 1, \ldots, N$, and noting that $\breve{\pi} = \min_{i\in V}\pi^i$, $\widehat{\pi} = \max_{i\in V}\pi^i$ and denoting $\widehat{L}(f) = \max_{i\in V}L(f^i)$, we obtain with probability 1 for all $k \geq \widetilde{k}$ and $i \in V$,

$$\begin{aligned}&\sum_{i=1}^N\mathbb{E}\left[\left\|x_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k\right] \\ &\quad\leq (1+\widehat{\pi}A_k)\sum_{i=1}^N\mathbb{E}\left[\left\|\varphi_{k+1}^i - x\right\|^2 \mid \mathscr{F}_k\right] \\ &\qquad + N\widehat{\pi}B_k(d+4)^2\widehat{L}^2(f) + \frac{4N\mu}{k^2\breve{\pi}}\sqrt{d}\widehat{L}(f) \\ &\qquad - \frac{2}{k}\left[f(\overline{x}_k) - f(x)\right] + \frac{2}{k}\widehat{L}(f)\sum_{i=1}^N\left\|x_k^i - \overline{x}_k\right\| \\ &\qquad + \frac{2}{k}(d+4)\widehat{L}(f)\sum_{i=1}^N\mathbb{E}\left[\left\|\varphi_{k+1}^i - x_k^i\right\| \mid \mathscr{F}_k\right],\end{aligned} \tag{20}$$

where $\overline{x}_k = (1/N)\sum_{i=1}^{N} x_k^i$ and we have used the following inequality:

$$\begin{aligned}\sum_{i=1}^{N}\left[f^i\left(x_k^i\right)-f^i\left(\overline{x}_k\right)\right] &\geq -\sum_{i=1}^{N} L\left(f^i\right)\left\|x_k^i-\overline{x}_k\right\| \\ &\geq -\widehat{L}\left(f\right)\sum_{i=1}^{N}\left\|x_k^i-\overline{x}_k\right\|.\end{aligned} \tag{21}$$

Now by using the definition of the weight matrix W_{k+1} and the convexity of the squared norm it follows that

$$\begin{aligned}\sum_{i=1}^{N}\mathbb{E}\left[\left\|\varphi_{k+1}^i-x\right\|^2 \mid \mathscr{F}_k\right] &\leq \sum_{i=1}^{N}\sum_{j=1}^{N}[\mathsf{W}_{k+1}]_{ij}\left\|x_k^j-x\right\|^2 \\ &= \sum_{j=1}^{N}\left\|x_k^j-x\right\|^2\end{aligned} \tag{22}$$

which yields the final bound for all $k \geq \widetilde{k}$ and $i \in V$ with probability 1:

$$\begin{aligned}&\sum_{i=1}^{N}\mathbb{E}\left[\left\|x_{k+1}^i-x^*\right\|^2 \mid \mathscr{F}_k\right] \\ &\quad\leq \left(1+\widehat{\pi}A_k\right)\sum_{i=1}^{N}\left\|x_k^i-x^*\right\|^2 + N\widehat{\pi}B_k(d+4)^2\widehat{L}^2\left(f\right) \\ &\qquad+\frac{4N\mu}{k^2\check{\pi}}\sqrt{d}\widehat{L}\left(f\right)-\frac{2}{k}\left[f\left(\overline{x}_k\right)-f\left(x^*\right)\right] \\ &\qquad+\frac{2}{k}\left(2d+9\right)\widehat{L}\left(f\right)N\max_{i\in V}\left\|x_k^i-\overline{x}_k\right\|,\end{aligned} \tag{23}$$

where $x^* \in \mathscr{X}^*$ and we have used the following inequality:

$$\begin{aligned}&\sum_{i=1}^{N}\mathbb{E}\left[\left\|\varphi_{k+1}^i-x_k^i\right\| \mid \mathscr{F}_k\right] \\ &\quad\leq \sum_{i=1}^{N}\sum_{j=1}^{N}[\mathsf{W}_{k+1}]_{ij}\left\|x_k^j-x_k^i\right\| \leq 2N\max_{i\in V}\left\|x_k^i-\overline{x}_k\right\|.\end{aligned} \tag{24}$$

Now we are ready to establish the convergence of the method. First, note that

$$\begin{aligned}&\sum_{k=1}^{\infty}\widehat{\pi}A_k<\infty, \\ &\sum_{k=1}^{\infty}N\widehat{\pi}B_k(d+4)^2\widehat{L}^2\left(f\right)+\sum_{k=1}^{\infty}\frac{4N\mu}{k^2\check{\pi}}\sqrt{d}\widehat{L}\left(f\right)<\infty\end{aligned} \tag{25}$$

which can be easily seen from the explicit expressions for A_k and B_k. For the term $\max_{i\in V}\|x_k^i-\overline{x}_k\|$, we can follow an argument similar to the proof of Lemma 4 in [5] and derive that for each $i \in V$, $\sum_{k=1}^{\infty}(1/k)\|x_k^i-\overline{x}_k\| < \infty$ and $\lim_{k\to\infty}\|x_k^i-\overline{x}_k\| = 0$, which gives

$$\sum_{k=1}^{\infty}\frac{2}{k}\left(2d+9\right)\widehat{L}\left(f\right)N\max_{i\in V}\left\|x_k^i-\overline{x}_k\right\|<\infty. \tag{26}$$

Hence, combining the preceding fact with Lemma 4, which we can obtain with probability 1, the sequence $\{\mathbb{E}[\|x_k^i-x^*\|^2]\}$ converges for any $x^* \in \mathscr{X}^*$, and $\sum_{k=1}^{\infty}(1/k)[f(\overline{x}_k)-f(x^*)] < \infty$ (note that $\overline{x}_k \in \mathscr{X}$, and hence $f(\overline{x}_k)-f(x^*) \geq 0$), which implies

$$\liminf_{k\to\infty} f\left(\overline{x}_k\right)=f\left(x^*\right). \tag{27}$$

This, along with the fact that the sequence $\{\mathbb{E}[\|x_k^i-x^*\|^2]\}$ converges for any $x^* \in \mathscr{X}^*$ and $\lim_{k\to\infty}\|x_k^i-\overline{x}_k\| = 0$, gives our final statement, that is, $\lim_{k\to\infty}x_k^i = x^*$ for all $i \in V$ with probability 1. □

Remark 6. Note that other choices of the parameters μ_k^i ($k \geq 0, i \in V$) are possible. For example, we can set $\mu_k^i = \mu\sqrt{\sigma_k^i}$, for all $k \geq 0$ and any $i \in V$, under which case the convergence of the method (7) can also be established.

Remark 7. In contrast to the subgradient-based methods in [1–3], the implementation of the proposed method does not need the information of subgradients but only the function values. This makes our method suitable for the cases where explicit gradient calculations are computationally infeasible or expensive. In contrast to the gradient-free method in [13], the proposed method is asynchronous and the step sizes do not require any coordination of the agents.

4. Conclusion

In this paper, we have considered the constrained multiagent optimization problem. We present an asynchronous method that is based on the gossip algorithm and the gradient-free oracles for solving the problem. The proposed method removes the need for synchronous communications and the information of the subgradients as well. Finally, we prove that with probability 1 the iterates of all agents converge to the same optimal point of the problem, for a diminishing step size. There are several interesting questions that remain to be explored. For instance, it would be interesting to study the case of constant step size; it would be also interesting to study the effects of message quantization on the proposed method.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant no. 61304042, the Natural Science Foundation of Jiangsu Province under Grant no. BK20130856, the Jiangsu Planned Projects for Postdoctoral Research Funds under Grant no. 1302003A, and Nanjing University of Posts and Telecommunications under Grant no. NY213041.

References

[1] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: convergence analysis and network scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, 2012.

[2] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proceedings of the 47th IEEE Conference on Decision and Control (CDC '08)*, pp. 4185–4190, Cancun, Mexico, December 2008.

[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] A. Nedić, "Asynchronous broadcast-based convex optimization over a network," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1337–1351, 2011.

[6] D. Yuan, S. Xu, and H. Zhao, "Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 6, pp. 1715–1724, 2011.

[7] M. Zhu and 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] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[9] M. G. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*, pp. 20–27, Berkeley, Calif, USA, 2004.

[10] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, "Gossip algorithms for convex consensus optimization over networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2917–2923, 2011.

[11] D. Yuan, S. Xu, H. Zhao, and L. Rong, "Distributed dual averaging method for multi-agent optimization with quantized communication," *Systems & Control Letters*, vol. 61, no. 11, pp. 1053–1061, 2012.

[12] D. Yuan, S. Xu, B. Zhang, and L. Rong, "Distributed primal-dual stochastic subgradient algorithms for multi-agent optimization under inequality constraints," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 16, pp. 1846–1868, 2013.

[13] D. Yuan, S. Xu, and J. Lu, "Gradient-free method for distributed multi-agent optimization via push-sum algorithms," *International Journal of Robust and Nonlinear Control*. In press.

[14] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Asynchronous gossip algorithms for stochastic optimization," in *Proceedings of the 48th IEEE Conference on Decision and Control Held Jointly with the 28th Chinese Control Conference (CDC/CCC '09)*, pp. 3581–3586, Shanghai, China, December 2009.

[15] K. Srivastava and A. Nedić, "Distributed asynchronous constrained stochastic optimization," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772–790, 2011.

[16] S. Lee and A. Nedic, "Distributed random projection algorithm for convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 221–229, 2013.

[17] S. Lee and A. Nedich, "Asynchronous gossip-based random projection algorithms over networks," submitted to *IEEE Transactions on Automatic Control*.

[18] S. S. Ram, A. Nedic, and V. V. Veeravalli, "Asynchronous gossip algorithms for stochastic optimization: constant stepsize analysis," in *Recent Advances in Optimization and Its Applications in Engineering*, M. Diehl, F. Glineur F, E. Jarlebring, and W. Michiels, Eds., pp. 51–60, Springer, Berlin, Germany, 2010.

[19] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[20] Yu. Nesterov, "Random gradient-free minimization of convex functions," CORE Discussion Paper 2011/1, 2011.

[21] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*, vol. 8 of *MPS/SIAM Series on Optimization*, SIAM, Philadelphia, Pa, USA, 2009.

[22] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Transactions on Automatic Control*, vol. 50, no. 5, pp. 655–661, 2005.

[23] 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.

[24] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[25] D. Yuan, S. Xu, H. Zhao, and Y. Chu, "Distributed average consensus via gossip algorithm with real-valued and quantized data for $0 < q < 1$," *Systems & Control Letters*, vol. 59, no. 9, pp. 536–542, 2010.

[26] J. Lu, D. W. C. Ho, and J. Kurths, "Consensus over directed static networks with arbitrary finite communication delays," *Physical Review E*, vol. 80, no. 6, Article ID 066121, 2009.