Systems & Control Letters 114 (2018) 44–51

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Distributed resource allocation over random networks based on stochastic approximation ☆

Peng Yi [a], Jinlong Lei [b], Yiguang Hong [c,*]

[a] Department of Electrical and Systems Engineering, Washington University in St. Louis, MO, USA
[b] Department of Industrial and Manufacturing Engineering, Pennsylvania State University, University Park, PA, USA
[c] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form 27 November 2017
Accepted 10 February 2018

Keywords:
Distributed optimization
Resource allocation
Stochastic approximation
Random graph
Demand response

ABSTRACT

In this paper, we study a resource allocation problem in which a group of agents cooperatively optimize a separable optimization problem with a linear network resource constraint and allocation feasibility constraints, where the global objective function is the sum of agents' local objective functions. Each agent can only get noisy observations of its local gradient function and its local resource, which cannot be shared by other agents or transmitted to a center. There also exist communication uncertainties such as time-varying topologies (described by random graphs) and additive channel noises. To solve the resource allocation with uncertainties, we propose a stochastic approximation based distributed algorithm, and prove that agents can collaboratively achieve the optimal allocation with probability one by virtue of the ordinary differential equation (ODE) method for stochastic approximation. Finally, simulations related to the demand response management in power systems verify the effectiveness of the proposed algorithm.

1. Introduction

Resource allocation problem is to allocate the network resource among a group of agents while optimizing certain performance index. It has drawn much research attention in many areas, such as the media access control in communication networks [1], signal processing in [2], and load demand management in [3]. Hence, various resource allocation models and algorithms have been proposed (see [1–6] and the references therein). However, most of existing algorithms need a center to collect the data over networks or to coordinate computation processes among all agents.

In fact, distributed optimization has attracted more and more research attention in recent years [7–13]. In various network optimization problems, the optimal decisions are made based on the whole network data, which, however, are collected and stored by each individual agent in the network. Distributed optimization keeps the data stored by network agents when seeking the optimal decision, and hence eliminates the "one-to-all" communication burden and protects agents' privacy. Distributed optimization also endows each agent with autonomy and reactivity by allowing it to formulate its local objective function and constraints with its local data. From the network viewpoint, the robustness to single point failure and the network scalability can be enhanced.

Following the seminal work [5] of resource allocation in large-scale networks along with distributed optimization in [7–13], various distributed algorithms for resource allocation have been proposed recently in [14–17]. Nevertheless, those works have not considered various stochastic uncertainties related to information sharing or data observations in distributed resource allocation. In this distributed problem, each agent needs to share its local information with other agents through a communication network, maybe with various uncertainties. Firstly, the topology of communication networks may be variable due to packet loss, media access control, or energy constraint. Secondly, the information shared through the network may not be accurate due to quantization errors or may be corrupted by random noises due to channel fading (referring to [13,18] and [19]). Thirdly, agents may not get the exact local gradient or resource information due to measurement or observation noises. In practice, noisy gradient was discussed in the zero-order distributed optimization as in [20], and randomized data sample was considered to reduce the computational complexity, also leading to noisy gradients [21].

Stochastic approximation has been adopted in distributed optimization to address various uncertainties. In [8], a distributed algorithm was proposed when each agent can only get the noisy observations of its local gradient, extending the centralized stochastic approximation (see [22]) to distributed settings. In [23], a stochastic approximation algorithm was given for root seeking

☆ This work was supported by National Key Research and Development Program of China (2016YFB0901902) and NSFC (61733018, 61333001, 61573344).
* Corresponding author.
E-mail addresses: yipeng@wustl.edu (P. Yi), jxl800@psu.edu (J. Lei), yghong@iss.ac.cn (Y. Hong).

https://doi.org/10.1016/j.sysconle.2018.02.009
0167-6911/

of a vector function as a sum of local functions. Besides, stochastic approximation was adopted in distributed optimization to handle uncertainties in communication systems in [9,10], and [18].

In this paper, a stochastic-approximation-based distributed algorithm is proposed to handle the resource allocation problem with uncertainties, where each agent only utilizes noisy observations of its local gradient and resource information, and noisy neighboring information shared over randomly switching networks. The convergence of the proposed algorithm to the optimal allocation with probability one is shown with help of the ODE method for stochastic approximation, and then the proposed model and algorithm are applied to distributed multi-periods demand response management in power systems, along with simulations for illustration. Note that the proposed algorithm takes a decomposition method different from existing ones on distributed optimization or distributed stochastic approximation in [23,8,7,10], since our problem cannot be converted to finding zeros of a sum of vector functions. In fact, we have each agent only take care of its local variable instead of reaching consensus on a global decision variable.

The remainder of the paper is organized as follows. The resource allocation problem is formulated and a distributed algorithm is proposed in Section 2. Then the convergence result is established in Section 3, while simulation studies are shown in Section 4. Finally, concluding remarks are given in Section 5.

Notations: Denote $\mathbf{1}_m = (1, \ldots, 1)^T \in \mathbf{R}^m$ and $\mathbf{0}_m = (0, \ldots, 0)^T \in \mathbf{R}^m$. $col\{x_1, \ldots, x_n\} = (x_1^T, \ldots, x_n^T)^T$ stacks the vectors $x_1, \ldots, x_n$. I_n denotes the identity matrix in $\mathbf{R}^{n\times n}$. For a matrix $A = [a_{ij}]$, a_{ij} or A_{ij} stands for the matrix entry in the ith row and jth column of A. $\otimes$ denotes the Kronecker product. Denote $ker\{A\}$ and $range\{A\}$ as the null space and range space of matrix A, respectively. For a closed convex set $\Omega \subset \mathbf{R}^m$ and point $x \in \mathbf{R}^m$, denote $P_\Omega(x)$ as the point in Ω that is closest to x, which is called the projection of x on Ω with the nonexpansive property as $\|P_\Omega(x) - P_\Omega(y)\| \le \|x - y\|\ \forall x, y \in \mathbf{R}^m$. For a convex set $\Omega \subset \mathbf{R}^m$ and $x \in \Omega$, define the normal cone to Ω at x as $N_\Omega(x) \triangleq \{v \in \mathbf{R}^m : \langle v, y - x\rangle \le 0\ \forall y \in \Omega\}$.

2. Problem formulation and distributed algorithm

We first formulate the resource allocation problem with the data observation and communication network models, and then propose the distributed algorithm.

2.1. Problem formulation

Consider a group of agents $\mathcal{N} = \{1, \ldots, n\}$ that cooperatively decide the optimal network resource allocation formulated as follows:

$$\min_{x_i\in\mathbf{R}^m, i\in\mathcal{N}} \quad \sum_{i\in\mathcal{N}} f_i(x_i),\ s.t., \sum_{i\in\mathcal{N}} x_i = \sum_{i\in\mathcal{N}} d_i, \quad x_i \in \Omega_i, i \in \mathcal{N}. \quad (1)$$

The local allocation variable $x_i \in \mathbf{R}^m$ is decided by agent i, which is also associated with a local objective function $f_i(x_i)$. d_i is the local resource data, and can only be observed by agent i. The resource of the whole network is the sum of all local resources, i.e., $\sum_{i\in\mathcal{N}} d_i$. Ω_i is the local allocation feasibility constraint of agent i, and cannot be known by other agents. Furthermore, Ω_i is determined by p_i inequality constraints: $\Omega_i = \{x \in \mathbf{R}^m : q_{ij}(x) \le 0,\ \forall j = 1, \ldots, p_i\}$, where $q_{ij}(\cdot),\ j = 1, \ldots, p_i$ are continuously differentiable convex functions on $\mathbf{R}^m$. Therefore, resource allocation problem (1) is to find an allocation that minimizes the sum of local objective functions while satisfying the network resource constraint and the allocation feasibility constraints. The following assumptions can also be found in [1–6].

Assumption 1. Problem (1) is feasible and has a finite optimal solution. For any $i \in \mathcal{N}$, $f_i(x_i)$ is a differentiable and strictly convex function, and moreover, it has a Lipschitz continuous gradient over Ω_i, i.e., $\exists l_c > 0$ such that $\|\nabla f_i(x) - \nabla f_i(y)\| \le l_c\|x - y\|$, $\forall x, y \in \Omega_i$.

The following constraint qualification assumption can be found in [24].

Assumption 2. For any $i \in \mathcal{N}$, Ω_i is a closed convex set and has nonempty interior points, and $\{\nabla q_{ij}(x),\ j \in \mathcal{I}_i(x)\}$ is linearly independent, where $\mathcal{I}_i(x) = \{j : q_{ij}(x) = 0\}$.

The data observation model for agent i at time k is given as follows: agent i can get the noisy observation of its gradient $\nabla f_i(x_i)$ at a given testing point $x_i(k)$ corrupted with noise $\nu_i(k)$ (that is, $\nabla f_i(x_i(k)) + \nu_i(k)$) and the noisy local resource information corrupted with noise $\delta_i(k)$ (that is, $d_i + \delta_i(k)$). The stochastic gradient model should be taken into consideration in the following three cases:

(i) Stochastic optimization: Agent i's local objective function takes the expectation-valued form as $f_i(x_i) = E_{\phi_i}[g(x_i, \phi_i)] = \int_{\Delta_i} g(x_i, \phi_i)d\mathbb{P}(\phi_i)$, where ϕ_i is a random variable supported on set $\Delta_i \in \mathbf{R}^d$ with distribution $\mathbb{P}$, and $g_i : \mathbf{R}^m \times \Delta_i \to \mathbf{R}$. Since the exact gradient of the expectation-valued function $E_{\phi_i}[g(x_i, \phi_i)]$ is generally unavailable, it is practical to utilize the noisy sampled gradient $\nabla g_i(x_i, \phi_i)$. The stochastic approximation algorithms in [22] and [8] considered this type of gradient noises.

(ii) Zero-order optimization: When agent i can only get the value of $f_i(x_i)$ at a given testing point $x_i(k)$, the gradient estimation methods, such as the Kiefer–Wolfowitz method in [25] and the randomized coordinate estimation in [20], can lead to noisy gradient observations.

(iii) Randomized data sample: If the local objective functions are constructed with "big data", a noisy gradient based on randomly sampled data is an alternative to the exact gradient, which may reduce the computational complexity (see [21]).

Given the local data observations, it is important and practical to solve (1) in a distributed way, where the agents need to share the local information with neighbors through switching networks and noisy channels.

As we know, switching communication networks can be modeled by random graphs, e.g., [9,10]. Denote a realization of the random graph at time k as $\mathcal{G}(k) = (\mathcal{N}, \mathcal{E}(k))$, where $\mathcal{E}(k) \subset \mathcal{N} \times \mathcal{N}$ is the edge set at time k. If agent i can get information from agent j at time k, then $(j, i) \in \mathcal{E}(k)$ and agent j belongs to agent i's neighbor set $\mathcal{N}_i(k) = \{j|(j, i) \in \mathcal{E}(k)\}$ at time k. Define the adjacency matrix $A(k) = [a_{ij}(k)]$ of $\mathcal{G}(k)$ with $a_{ij}(k) = 1$ if $j \in \mathcal{N}_i(k)$, and $a_{ij}(k) = 0$ otherwise. Denote by $Deg(k) = diag\{\sum_{j=1}^n a_{1j}(k), \ldots, \sum_{j=1}^n a_{nj}(k)\}$ the degree matrix, and by $L(k) = Deg(k) - A(k)$ the Laplacian matrix of $\mathcal{G}(k)$.

The following assumption is given for the random graphs $\{\mathcal{G}(k)\}_{k\ge1}$ (referring to [9,10,26,19]).

Assumption 3. $\{L(k)\}$ is an i.i.d. sequence with its mean denoted by $\bar{L} = E[L(k)]$. Besides, $\bar{L}$ is symmetric with $s_2(\bar{L}) > 0$, where $s_2(\bar{L})$ denotes the secondly smallest eigenvalue of $\bar{L}$.

Remark 1. Note that Assumption 3 does not require the communication graph to be connected or undirected at any time instance. Only the mean graph is required to be undirected and connected, which ensures that the local information can reach any other agents in the average sense. The gossip model in [26] and the broadcast model in [10] are consistent with Assumption 3.

2.2. Distributed algorithm

With Assumption 1, the KKT condition of (1) is $\exists\ \lambda^* \in \mathbf{R}^m$ such that