Automatica 77 (2017) 230–238

Contents lists available at ScienceDirect

Automatica

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

Brief paper

A distributed hierarchical algorithm for multi-cluster constrained optimization ☆

Fanghong Guo [a,b], Changyun Wen [b], Jianfeng Mao [c], Guoqi Li [d], Yong-Duan Song [e]

[a] Energy Research Institute@NTU (ERI@N), Interdisciplinary Graduate School, Nanyang Technological University, Singapore
[b] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[c] School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China
[d] Center for Brain Inspired Computing Research, Department of Precision Instrument, Tsinghua University, Beijing, China
[e] Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, and School of Automation, Chongqing University, Chongqing, China

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form
18 July 2016
Accepted 7 November 2016

Keywords:
Distributed optimization
Projected gradient
Sequential communication
Virtual agent

ABSTRACT

In this paper, we consider a constrained optimization problem for a large-scale multi-cluster agent system, in which a number of clusters already exist as a priori. The aim is to minimize a global objective function being the sum of multi-cluster local agents' cost functions subject to certain global constraints. To solve this problem, a novel distributed hierarchical algorithm based on projected gradient method is proposed by using synchronous and sequential communication strategies. We firstly assign one agent as leader agent in each cluster, which can communicate with the leaders of its neighboring clusters. The agents in the same cluster conduct local optimization and communicate with their neighboring agents synchronously while the leader agents of different clusters exchange information in a sequential way. Then a scheme is proposed for each agent to iteratively estimate a solution of the optimization problem in a distributed manner. It is theoretically proved that the estimated solutions of all the agents reach consensus of the optimal solution asymptotically when the chosen stepsizes are diminishing. Numerical examples are provided to validate the proposed method.

1. Introduction

Due to potential applications in power system (Lam, Zhang, & Tse, 2012; Xue, Li, & Nahrstedt, 2006), wireless network system (Wang, Jiang, & He, 2014) and sensor network systems (Wan & Lemmon, 2009), there has been growing research interests in distributed optimization, where local agents cooperatively minimize a global objective cost function which is the sum of local objective functions (Bianchi & Jakubowicz, 2013; Droge, Kawashima, & Egerstedt, 2014; Duchi, Agarwal, & Wainwright, 2012; Gharesifard & Cortes, 2012a,b; Kia, Cortes, & Martinez, 2015; Lin, Ren, & Song, 2016; Nedic & Bertsekas, 2001; Nedic & Ozdaglar, 2009; Nedic, Ozdaglar, & Parrilo, 2010; Yi, Hong, & Liu, 2015; Zhu & Martinez, 2012). Different from the conventional centralized optimization where information of all the agents is available and collected in one central node, distributed optimization decomposes the central node into several sub-nodes (agents) for reasons including privacy concern, computational and communication burden. Generally speaking, according to the form of formulated problems, there are mainly two kinds of algorithms to solve distributed optimization, i.e., continuous-time algorithm (Droge et al., 2014; Gharesifard & Cortes, 2012a,b; Kia et al., 2015; Yi et al., 2015) and discrete-time algorithm (Bianchi & Jakubowicz, 2013; Duchi et al., 2012; Johansson, Rabi, & Johansson, 2009; Lin et al., 2016; Nedic & Bertsekas, 2001; Nedic & Ozdaglar, 2009; Nedic et al., 2010; Zhu & Martinez, 2012).

The continuous-time algorithm is mainly developed on the basis of well-developed control theory (Gharesifard & Cortes, 2012b).

☆ This work was supported in part by the Major State Basic Research Development Program 973 under Grant 2012CB215202 and in part by the National Natural Science Foundation of China under Grant 61134001. This work was also jointly supported by Singapore's National Research Foundation and the Energy Research Institute at NTU (ERI@N) under Grant NRF-CRP8-2011-03. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Giancarlo Ferrari-Trecate under the direction of Editor Ian R. Petersen.

E-mail addresses: fguo003@e.ntu.edu.sg (F. Guo), ecywen@ntu.edu.sg (C. Wen), jfmao@cuhk.edu.cn (J. Mao), liguoqi@mail.tsinghua.edu.cn (G. Li), ydsong@cqu.edu.cn (Y.-D. Song).

http://dx.doi.org/10.1016/j.automatica.2016.11.029
0005-1098/

In Droge et al. (2014), a continuous-time proportional–integral distributed optimization method is proposed, where dual decomposition and consensus-based method are used. The convergence of continuous-time distributed optimization over directed networks is analyzed in Gharesifard and Cortes (2012a). An event-triggered based continuous-time distributed coordination algorithm is proposed in Kia et al. (2015). However, very few continuous-time algorithms have addressed constrained optimization problems except for Yi et al. (2015), where the KKT condition and Lagrangian multiplier methods are cooperatively used.

Apart from continuous-time algorithms, discrete-time algorithms have also been well studied in distributed optimization. Among them, gradient or subgradient based methods are most popular due to their simplicity and ease of implementation. In Zhu and Martinez (2012), two distributed primal–dual subgradient algorithms are developed under inequality and equality constraints respectively. The convergence rate of distributed dual subgradient averaging method is analyzed in Duchi et al. (2012). A stochastic gradient algorithm is proposed in Bianchi and Jakubowicz (2013) to solve a distributed non-convex optimization problem. Constrained consensus and constrained optimization problems are considered in Nedic et al. (2010), where a distributed projected consensus algorithm and a projected gradient method are proposed to solve such problems respectively. It is worthy to point out that all these distributed optimizations are based on synchronous communication, i.e., all the agents conduct the optimization and communicate in a synchronous way.

However, when the system becomes larger and larger especially for the situation that the agents are sparsely distributed with different clusters, distributed optimization with synchronous communication may not be proper. For example, for a large-scale sensor network system, the sensors are located sparsely, then their communication latencies are quite different when using the synchronous communication strategy. In this paper, a constrained optimization problem for a large-scale multi-cluster agent system is considered. Motivated by the sequential communication strategy (also known as incremental communication) proposed in Johansson et al. (2009) and Nedic and Bertsekas (2001), a distributed algorithm is proposed to solve the above problem, which inherits the advantages of both synchronous and sequential communication.

The main ideas and contributions of this paper are summarized as follows. Different from related works in Nedic and Bertsekas (2001) and Nedic et al. (2010), we consider a constrained optimization problem in multi-cluster case, where clusters exist as a priori. Firstly we assign a leader to each cluster. Then within the same cluster, each agent conducts the optimization based on its local cost functions. Next it exchanges its estimate with its corresponding neighboring agents by using finite-time average consensus algorithm. Once the exchanged estimates of all the agents in the cluster reach consensus, a cluster estimate is obtained. Subsequently, the leader agent conducts an extra manipulation on the cluster estimate, and passes the modified cluster estimate to a leader agent in its neighboring cluster based on a sequential communication protocol. It is worthy to point out that such extra manipulation can be regarded as the effect of *virtual agents*, a new idea proposed to achieve convergence property. With the help of such virtual agents, it is theoretically proved that the estimated solutions of all the agents reach consensus of the optimal solution asymptomatically if the stepsize is chosen according to a guideline given in Theorem 1.

2. Problem formulation

Motivated by the optimization problems in practical systems such as power system (Lam et al., 2012; Xue et al., 2006), wireless network system (Wan & Lemmon, 2009; Wang et al., 2014) and so on, a large-scale multi-agent system with $m = |\mathcal{M}|$ clusters is considered, $\mathcal{M}$ is the set of clusters. Each cluster has $n_i = |\mathcal{A}_i|$ agents, with $\mathcal{A}_i, \forall i = 1, \ldots, m$ denoting the set of agents in the ith cluster. Each agent j in cluster i has its local objective function $f_i^j(x)$, which is only known to agent j itself and cannot be shared with other agents. Also a global constraint set X is imposed and known to all the agents. The goal of the agents is to cooperatively solve the constrained optimization problem

$$\begin{cases} \min\limits_{x} \sum_{i=1}^{m} \sum_{j=1}^{n_i} f_i^j(x) \\ \text{s.t. } x \in X \end{cases} \tag{1}$$

where $f_i^j(x) : \mathbb{R}^n \to \mathbb{R}$ and X are a convex function and a compact convex set respectively.

Denote the optimal solution of (1) as $x^\star \in X$, it is easy to conclude that $x^\star$ exists according to the well-known *extreme value theorem* (Nedic et al., 2010). But it is unknown to each agent. Our idea is to let each agent estimate the optimal solution by using the available information of its neighboring agents and itself iteratively. Denote the estimate of the agent j in cluster i at iteration l as $\hat{x}_i^j(l),\ i = 1, \ldots, m,\ j = 1, \ldots, n_i$. Then our objective is to propose an algorithm to ensure all these estimates reach consensus of the optimal solution as iterations increase, i.e., $\lim_{l\to\infty} \hat{x}_i^j(l) = x^\star$, for all $i = 1, \ldots, m,\ j = 1, \ldots, n_i$.

To achieve this, we make the following assumption.

Assumption 1. Function f_i^j is convex and differentiable.

Suppose ∇f_i^j is the gradient function of f_i^j, then from Johansson et al. (2009) and Nedic and Bertsekas (2001), Assumption 1 ensures that it is bounded over the set X, i.e., there exists a scalar $L > 0$, such that

$$\left\| \nabla f_i^j(x) \right\| \le L, \quad \forall x \in X \tag{2}$$

if the set X is compact (Nedic & Ozdaglar, 2009).

Remark 1. In this paper, we assume f_i^j is convex and differentiable, so that its gradient $\nabla f_i^j(x)$ exists for any $x \in \mathbb{R}^n$. However, this assumption can be relaxed to that f_i^j is only convex and allowed to be non-differentiable at some points. In this case, a *subgradient* exists and can be used in the role of a gradient (Nedic et al., 2010).

3. Distributed optimization algorithm

In this section, a distributed optimization algorithm is proposed to solve the problem formulated in (1). As preliminaries which will be useful in our later convergence analysis, we first briefly review and introduce some properties on projection operation followed by reviewing some knowledge on finite-time average consensus. After that, we present our distributed optimization algorithm.

3.1. Preliminaries

3.1.1. Projection

The projection of a vector $\bar{x}$ onto a closed convex set X is defined as

$$P_X[\bar{x}] = arg \min_{x\in X} \|\bar{x} - x\| \tag{3}$$

where $\|x\|$ denotes the Euclidean norm, i.e., $\|x\| = \sqrt{x^T x}$, x^T denotes the transpose of vector x.

One important property of the projection operation is summarized as following lemma.