Systems & Control Letters 107 (2017) 36–43

Contents lists available at ScienceDirect

Systems & Control Letters

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

A fast proximal gradient algorithm for decentralized composite optimization over directed networks☆

Jinshan Zeng *, Tao He, Mingwen Wang

College of Computer Information Engineering, Jiangxi Normal University, Nanchang, 330022, PR China

ARTICLE INFO

Article history:
Received 7 November 2016
Received in revised form 26 June 2017
Accepted 3 July 2017
Available online 8 August 2017

Keywords:
Decentralized optimization
Directed network
Composite objective
Nonconvex
Consensus

ABSTRACT

This paper proposes a fast decentralized algorithm for solving a consensus optimization problem defined in a *directed* networked multi-agent system, where the local objective functions have the smooth+nonsmooth composite form, and are possibly *nonconvex*. Examples of such problems include decentralized compressed sensing and constrained quadratic programming problems, as well as many decentralized regularization problems. We extend the existing algorithms PG-EXTRA and ExtraPush to a new algorithm *PG-ExtraPush* for composite consensus optimization over a *directed* network. This algorithm takes advantage of the proximity operator like in PG-EXTRA to deal with the nonsmooth term, and employs the push-sum protocol like in ExtraPush to tackle the bias introduced by the directed network. With a proper step size, we show that PG-ExtraPush converges to an optimal solution at a linear rate under some regular assumptions. We conduct a series of numerical experiments to show the effectiveness of the proposed algorithm. Specifically, with a proper step size, PG-ExtraPush performs linear rates in most of cases, even in some nonconvex cases, and is significantly faster than Subgradient-Push, even if the latter uses a hand-optimized step size. The established theoretical results are also verified by the numerical results.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

We consider the following consensus optimization problem defined on a *directed*, strongly connected network of n agents:

$$\underset{x\in\mathbb{R}^p}{\text{minimize}}\, f(x) \triangleq \sum_{i=1}^{n} f_i(x), \quad \text{where} \quad f_i(x) = s_i(x) + r_i(x), \tag{1.1}$$

and for every agent i, f_i is a proper, coercive and possibly nonconvex function only known to the agent, s_i is a smooth function, r_i is generally nonsmooth and possibly nonconvex. We say that the objective has the smooth+nonsmooth composite structure.

The smooth+nonsmooth structure of the local objective arises in a large number of signal processing, statistical inference, and machine learning problems. Specific examples include: (i) the geometric median problem in which s_i vanishes and r_i is the ℓ_2-norm [1]; (ii) the compressed sensing problem, where s_i is the data-fidelity term, which is often differentiable, and r_i is a sparsity-promoting regularizer such as the ℓ_q (quasi)-norm with $0 \le q \le 1$ [2,3]; (iii) optimization problems with per-agent constraints, where s_i is a differentiable objective function of agent i and r_i is the indicator function of the constraint set of agent i, that is, $r_i(x) = 0$ if x satisfies the constraint and ∞ otherwise [4,5].

For a stationary network with bi-directional communication, the existing algorithms include the primal–dual domain methods such as the decentralized alternating direction method of multipliers (DADMM) [6,7], and the primal domain methods including the distributed subgradient method (DSM) [8]. Both algorithms do not take advantage of the smooth+nonsmooth structure. While the algorithms that consider smooth+nonsmooth objectives in the form of (1.1) include the following primal-domain methods: the (fast) distributed proximal gradient method (DPGM) [9], the proximal decentralized gradient descent method (Prox-DGD) [10], the distributed iterative soft thresholding algorithm (DISTA) [11], proximal gradient exact first-order algorithm (PG-EXTRA) [12]. All these primal-domain methods consist of a gradient step for the smooth part and a proximal step for the nonsmooth part. Different from DPGM, Prox-DGD and DISTA, PG-EXTRA as an extension of EXTRA [13] has two interlaced sequences of iterates, whereas the proximal-gradient method just inherits the sequence of iterates in the gradient method.

☆ The work of J. Zeng is supported in part by the National Natural Science Foundation of China (Grants No. 61603162, 11401462) and the Doctoral start-up foundation of Jiangxi Normal University. The work of T. He has been supported in part by Graduate Innovation Programs of Education Department of Jiangxi Province (No. YC2016-S171). The work of M. Wang is supported in part by the National Natural Science Foundation of China (No. 61462045).

* Corresponding author.
E-mail address: jinshanzeng@jxnu.edu.cn (J. Zeng).

http://dx.doi.org/10.1016/j.sysconle.2017.07.005
0167-6911/© 2017 Elsevier B.V. All rights reserved.

This paper focuses on a *directed* network with *directional* communication, which is pioneered by the works [14–16]. When communication is bi-directional, algorithms can use a symmetric and doubly-stochastic mixing matrix to obtain a consensual solution; however, once the communication is directional, the mixing matrix becomes generally asymmetric and only column-stochastic. In the column-stochastic setting, the push-sum protocol [17] can be used to obtain a stationary distribution for the mixing matrix. Some recent decentralized algorithms over a directed network include Subgradient-Push [18], ExtraPush [19] (also called DEXTRA in [20]) and Push-DIGing [21]. The best rate of Subgradient-Push in the general convex case is $O(\ln t/\sqrt{t})$, where t is the iteration number, and both ExtraPush and Push-DIGing perform linearly convergent in the strongly convex case. However, all of these algorithms do not consider the smooth+nonsmooth structure as well as the nonconvex case as defined in problem (1.1).

In this paper, we extend the algorithms PG-EXTRA and ExtraPush to the composite consensus optimization problem with the smooth+nonsmooth structure, and establish the convergence and linear convergence rate of the proposed PG-ExtraPush algorithm. At each iteration, each agent locally computes a gradient of the smooth part of its objective and a proximal map of the nonsmooth part, and exchanges information with its neighbors, then uses the push-sum protocol [17] to achieve the consensus. When the network is undirected, the proposed PG-ExtraPush reduces to PG-EXTRA, and when $r_i \equiv 0$, PG-ExtraPush reduces to ExtraPush [19]. If the smooth part of objective is Lipschitz differentiable and quasi-strongly convex and the nonsmooth part is convex with bounded subgradient (see Assumption 3), we prove that with a proper step size, the proposed algorithm converges to an optimal solution at a linear rate. We provide a series of numerical experiments including three convex cases and one nonconvex case, to show the effectiveness of the proposed algorithm. Specifically, when applied to the convex cases, PG-ExtraPush performs the linear rates, and is significantly faster than Subgradient-Push, even if the latter uses a hand-optimized step size. While when applied to the nonconvex decentralized ℓ_q regularized least squares regression problems with $0 \leq q < 1$, it can be observed that the proposed algorithm performs an eventual linear convergence rate, that is, PG-ExtraPush performs a linear decay starting from a few iterations but not the initial iteration. This means that if we can fortunately get a good initial guess, the proposed algorithm PG-ExtraPush might decay linearly even in these nonconvex cases.

It should be pointed out that the extension from ExtraPush [19] to PG-ExtraPush is non-trivial. The main differences between the proposed algorithm PG-ExtraPush and ExtraPush [19] can be summarized as follows:

1. **On algorithm development.** Clearly, PG-ExtraPush extends ExtraPush to handle nonsmooth objective terms. This extension is not the same as the extension from the gradient method to the proximal-gradient method, as well as the extension from EXTRA [13] to PG-EXTRA [12]. As the reader will see, PG-ExtraPush will have three interlaced sequences of iterates, whereas the proximal-gradient method just inherits of the sequence of iterates in the gradient method; and PG-ExtraPush uses the proximal maps of a sequence of transformed functions of r_i associated with a positive weight sequence $\{\mathbf{w}^t\}$ essentially introduced by the directed graph, while PG-EXTRA utilizes the proximity operator of r_i.
2. **On convergence analysis.** Although the convergence analysis of this paper is motivated by the existing analysis in [19], there are several new proof techniques. The convergence of many existing algorithms like ExtraPush [19] is established based on a similar inequality of (4.16) as presented in Theorem 3. However, it is difficult to directly prove that such an inequality holds for all iterations of PG-ExtraPush. Instead, we can only establish the inequality (4.16) for a fixed iteration of PG-ExtraPush under the boundedness assumption of the previous two iterates. In order to establish the key inequality for all iterations, an induction technique is used as shown in the proof of Theorem 3. Moreover, the linear convergence rate of the proposed algorithm is established from the key inequality (4.16) via a recursive way. All of these are different from the convergence analysis in [19].

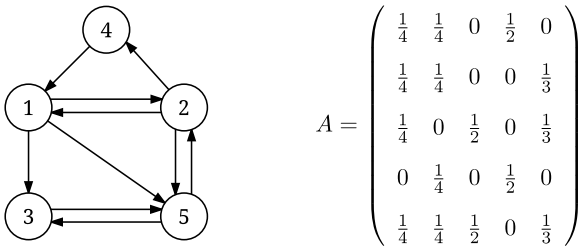

Fig. 1. A directed graph $\mathcal{G}$ (left) and its mixing matrix A (right) [19].

The rest of paper is organized as follows. Section 2 introduces the problem setup. Section 3 develops the proposed algorithm. Section 4 establishes its convergence and convergence rate. Section 5 presents our numerical results. We conclude this paper in Section 6.

Notation: Let $\mathbf{I}_n$ denote an identity matrix with the size $n \times n$. We use $\mathbf{1}_n \in \mathbb{R}^n$ as a vector of all 1's. For any *vector* x, we let x_i denote its ith component and $\mathbf{diag}(x)$ denote the diagonal matrix generated by x. For any matrix X, X^T denotes its transpose, X_{ij} denotes its (i,j) th component, and $\|X\| \triangleq \sqrt{\langle X, X\rangle} = \sqrt{\sum_{i,j} X_{ij}^2}$ denotes its Frobenius norm. The largest and smallest eigenvalues of matrix X are denoted as $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$, respectively. For any matrix $B \in \mathbb{R}^{m\times n}$, $\mathbf{null}(B) \triangleq \{x \in \mathbb{R}^n | Bx = 0\}$ is the null space of B. **Given a matrix** $B \in \mathbb{R}^{m\times n}$, **by** $Z \in \mathbf{null}(B)$, **we mean that each column of** Z **lies in null**(B). The smallest *nonzero* eigenvalue of a symmetric positive semidefinite matrix $X \neq \mathbf{0}$ is denoted as $\tilde{\lambda}_{\min}(X)$, which is strictly positive. For any positive semidefinite matrix $G \in \mathbb{R}^{n\times n}$ (not necessarily symmetric in this paper), we use the notion $\|X\|_G^2 \triangleq \langle X, GX\rangle$ for a matrix $X \in \mathbb{R}^{n\times p}$.

2. Problem reformulation

2.1. Network

Consider a *directed* network $\mathcal{G} = \{V, E\}$, where V is the vertex set and E is the edge set. Any edge $(i,j) \in E$ represents a directed arc from node i to node j. The sets of in-neighbors and out-neighbors of node i are

$$\mathcal{N}_i^{\text{in}} \triangleq \{j : (j,i) \in E\} \cup \{i\},\ \mathcal{N}_i^{\text{out}} \triangleq \{j : (i,j) \in E\} \cup \{i\},$$

respectively. Let $d_i \triangleq |\mathcal{N}_i^{\text{out}}|$ be the out-degree of node i. In $\mathcal{G}$, each node i can only send information to its out-neighbors, *not* vice versa.

To illustrate a mixing matrix for a directed network, consider $A \in \mathbb{R}^{n\times n}$ where

$$\begin{cases} A_{ij} > 0, & \text{if } j \in \mathcal{N}_i^{\text{in}} \\ A_{ij} = 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

The entries A_{ij} satisfy that, for each node j, $\sum_{i\in V} A_{ij} = 1$. An example is the following mixing matrix

$$A_{ij} = \begin{cases} 1/d_j, & \text{if } j \in \mathcal{N}_i^{in} \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات