

基于 Push-Sum 的分布式 Gradient-Free 算法研究^{*}

余淑辉, 李觉友, 杜学武
(重庆师范大学 数学科学学院, 重庆 401331)

摘要:【目的】提出了 push-sum 协议下的分布式 gradient-free 算法来求解一类有向网络的多智能体分布式优化问题。【方法】首先用 Gaussian 光滑化方法来逼近非可微函数,其次采用 push-sum 通讯协议考虑有向网络中的分布式 gradient-free 算法。【结果】分析了算法的收敛性,并得到了算法的收敛率为 $O(\ln \tau / \tau)$,其中 τ 是迭代次数。【结论】数值例子表明了所提出的算法与对应的分布式次梯度算法具有相似的收敛性。

关键词:多智能体网络;分布式优化;gradient-free;push-sum 协议

中图分类号:O224;O236

文献标志码:A

文章编号:1672-6693(2019)02-0011-07

1 背景介绍

当今社会,网络已成为生活中不可或缺的设施系统,如无线网络、电网、交通网络、供应链等。这些系统有一个共同特征即它们均由许多子系统组成,而这些子系统由一群具备一定的感知、通信、计算和执行能力的智能个体通过相互通讯等方式关联成的网络系统,通常称之为多智能体系统(Multi-agent system)^[1-2]。目前有许多文献研究了基于多智能体系统的分布式优化问题^[3-8],凸优化问题是其中研究的重点之一^[5-11],目前大多采用次梯度方法^[5-9]来求解。文献[6]较早讨论了多智能体网络凸优化的分布式次梯度算法,并给出了算法的收敛性分析。对于分布式强凸优化问题,文献[7]提出了分布式随机次梯度算法。但许多实际问题中的目标函数是不可微的^[13],其次梯度往往难以计算,有时甚至不可用^[12,14]。最近,Duchi 等人^[12]提出了一类光滑化方法来求解随机凸优化问题,该方法仅利用函数值信息,不需要梯度信息。但他们的方法是中心化的,并未考虑分布式情形。随后,文献[14]利用 Gaussian 光滑化方法来光滑化非可微目标函数,并提出了一种 gradient-free 分布式优化算法来求解一类非光滑的分布式优化问题。该算法只需计算函数值的信息,而不需要次梯度的信息。但是,文献[14]所提出的算法仅适用于无向网络拓扑。而大多数通讯网络是有向的。最近,文献[7]首次提出了 push-sum 算法,该算法考虑的是时变有向网络拓扑下的分布式优化,但此算法需要计算次梯度,而次梯度信息往往是不容易得到的^[12,14]。

受文献[7,12]的启发,针对强凸优化问题,本文提出了一个有向网络下的分布式 gradient-free 算法,分析了算法的收敛性,并给出了算法的收敛率。通过数值试验,验证了算法的可行性和有效性。与文献[7]的算法相比,本文的算法不需要计算目标函数的次梯度,仅仅需要计算函数值。同时将文献[12]的 gradient-free 方法推广到了分布式情形。与文献[14]所提出的算法不同的是,本文的算法适用于有向网络。

2 问题与算法

2.1 问题

多智能体网络往往需要所有智能体相互协作来极小化几个目标函数之和,其中每个智能体仅知道自身目标函数信息,并与其他智能体局部交换信息来达到全局最优。具体考虑如下类多智能体分布式优化问题:

^{*} 收稿日期:2018-03-07 修回日期:2018-12-26 网络出版时间:2019-01-09 07:38

资助项目:国家自然科学基金(No. 11501070);重庆市自然科学基金(No. cstc2017jcyjAX0253);重庆市教育委员会科学技术研究项目(No. KJQN201800520)

第一作者简介:余淑辉,女,研究方向分布式理论与算法,E-mail:503023849@qq.com;通信作者:李觉友,男,副教授,博士,E-mail:lijueyou@cqnu.edu.cn

网络出版地址: <http://kns.cnki.net/kcms/detail/50.1165.N.20190109.0731.042.html>

$$\min_{\mathbf{z} \in \mathbf{R}^d} F(\mathbf{z}) \triangleq \sum_{i=1}^n f_i(\mathbf{z}). \quad (1)$$

其中,每个智能体 i 仅持有目标函数 $f_i: \mathbf{R}^d \rightarrow \mathbf{R}, i=1, \dots, n, \mathbf{z}$ 是一个全局决策向量。

假设 1 a) 对 $i=1, \dots, n$, 每个 f_i 是非可微的, 并且是 μ_i -强凸^[12] 的, 其中 $\mu_i > 0$; b) 每个 f_i 是 L_i -Lipschitz 连续^[12] 的, 其中 $L_i > 0$ 。记 $L = L_1 + \dots + L_n$ 。

在假设 1 的情形下, 问题(1)存在唯一全局最优解, 记为 $\mathbf{z}^*$, 即 $\mathbf{z}^* = \arg \min_{\mathbf{z} \in \mathbf{R}^d} F(\mathbf{z})$ 。

引理 1^[6] 如果 f_i 是凸的非光滑函数, 令 $\tilde{f}_i(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbf{R}^d} f_i(\mathbf{z} + \sigma_i \boldsymbol{\varphi}) e^{-\frac{1}{2} \|\boldsymbol{\varphi}\|^2} d\boldsymbol{\varphi}$, 其中 σ_i 是一个正的光滑化参数, $\boldsymbol{\varphi} \in \mathbf{R}^d$ 是一个服从标准正太分布的随机向量, 那么:

- $\tilde{f}_i(\mathbf{z})$ 是凸函数, 且满足 $f_i(\mathbf{z}) \leq \tilde{f}_i(\mathbf{z}) \leq f_i(\mathbf{z}) + \sqrt{d} \sigma_i L_i$;
- $\tilde{f}_i(\mathbf{z})$ 是可微函数, 并且它的梯度为 $\nabla \tilde{f}_i(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{\mathbf{R}^d} \frac{f_i(\mathbf{z} + \sigma_i \boldsymbol{\varphi}) - f_i(\mathbf{z})}{\sigma_i} \boldsymbol{\varphi} e^{-\frac{1}{2} \|\boldsymbol{\varphi}\|^2} d\boldsymbol{\varphi}$;
- 令 $\tilde{\mathbf{g}}_i(\mathbf{z}) = \frac{f_i(\mathbf{z} + \sigma_i \boldsymbol{\varphi}) - f_i(\mathbf{z})}{\sigma_i} \boldsymbol{\varphi}$, 则有 $E[\tilde{\mathbf{g}}_i(\mathbf{z})] = \nabla \tilde{f}_i(\mathbf{z})$ 和 $E[\|\tilde{\mathbf{g}}_i(\mathbf{z})\|^2] \leq (d+4)^2 L_i^2$ 。

引理 1 指出: $\tilde{f}_i(\mathbf{z})$ 是 $f_i(\mathbf{z})$ 的 Gaussian 光滑化逼近, 并且具有可微等良好性质。

若将网络中每个智能体看作是一个节点, 则多智能体网络中智能体之间的通讯可以用一个时变的有向图 $G(t) = (V, E(t))$ 来表示, 其中 $V = \{1, 2, \dots, n\}$ 为节点的集合, $E(t) \subseteq V \times V$ 为边的集合。用 $N_i^{\text{in}}(t)$ 和 $N_i^{\text{out}}(t)$ 分别表示在时间 t 时对智能体 i 传入和传出的邻居节点的集合, 记 $N_i^{\text{in}}(t) = \{j | (j, i) \in E(t)\} \cup \{i\}$, $N_i^{\text{out}}(t) = \{j | (i, j) \in E(t)\} \cup \{i\}$ 。用 $d_i(t)$ 表示个体 i 在 t 时刻的外度, 即 $d_i(t) = |N_i^{\text{out}}(t)|$ 。定义通讯权矩阵 $\mathbf{A}(t)$ 如下:

$$[\mathbf{A}(t)]_{ij} = \begin{cases} \frac{1}{d_j(t)}, j \in N_i^{\text{in}}(t) \\ 0, \text{其他} \end{cases}. \quad (2)$$

事实上, 矩阵 $\mathbf{A}(t)$ 是列随机矩阵^[7]。

假设 2 a) 在有向图 $G(t)$ 中, 智能体 i 通过边 (i, j) 传递给智能体 $j \in N_i^{\text{out}}(t)$, 智能体 i 仅知道自身的外度 $d_i(t)$; b) 有向图序列 $\{G(t)\}$ 是 B -强连通的, 即存在一个正整数 B , 使得对于任意的正整数 k , $E_B(k) = E(kB) \cup E(kB+1) \cup \dots \cup E((k+1)B-1)$ 是强连通的。

2.2 分布式 Gradient-free 算法

对于许多非光滑的凸优化问题, 次梯度信息不容易得到, 甚至不可用^[12]。为此, 提出了一种分布式 gradient-free 算法(简称算法 GFD), 同时考虑了有向网络下的分布式优化算法。

算法 GFD 1) 初始化, 对每个智能体 i , 设置 $y_i(0) = 1$, 随机取 $\mathbf{x}_i(0) \in \mathbf{R}^d, i=1, \dots, n$;

2) 对于所有的 $t > 0$ 和对每个智能 $i, i=1, \dots, n$, 按照如下规则更新:

$$\mathbf{w}_i(t+1) = \sum_{j=1}^n [\mathbf{A}(t)]_{ij} \mathbf{x}_j(t), \quad (3)$$

$$y_i(t+1) = \sum_{j=1}^n [\mathbf{A}(t)]_{ij} y_j(t), \quad (4)$$

$$\mathbf{z}_i(t+1) = \frac{\mathbf{w}_i(t+1)}{y_i(t+1)}, \quad (5)$$

$$\mathbf{x}_i(t+1) = \mathbf{w}_i(t+1) - \alpha(t+1) \tilde{\mathbf{g}}_i(t+1). \quad (6)$$

其中, 步长序列 $\{\alpha(t)\}_{t=0}^{\infty}$ 是非负单调递减的, $\tilde{\mathbf{g}}_i(t+1)$ 是按照如下方式计算:

$$\tilde{\mathbf{g}}_i(t+1) = \frac{f_i(\mathbf{z}_i(t+1) + \sigma_i \boldsymbol{\varphi}_i(t+1)) - f_i(\mathbf{z}_i(t+1))}{\sigma_i} \boldsymbol{\varphi}_i(t+1), \quad (7)$$

这里 $\sigma_i > 0$ 是光滑化参数, $\boldsymbol{\varphi}_i(t+1) \in \mathbf{R}^d$ 是一个服从标准正态分布的随机向量。

注 1 1) 在算法 GFD 中, 首先智能体 j 发出信息 $\mathbf{x}_j(t)$ 和 $y_j(t)$ 给所有的智能体 $i \in N_j^{\text{out}}(t)$, 然后智能体 i 加权求和所有智能体 $j \in N_i^{\text{out}}(t)$ 的信息, 即(3), (4)式。最后智能体 i 执行局部计算, 即(5), (6)式。

2) 算法 GFD 不需要计算 f_i 的次梯度, 而仅仅计算它的函数值, 即(7)式。这对许多实际工程问题非常有帮助。事实上, 还可以按照如下方式计算 $\tilde{\mathbf{g}}_i(t+1)$:

$$\tilde{\mathbf{g}}_i(t+1) = \frac{f_i(\mathbf{z}_i(t+1) + \sigma_i \boldsymbol{\varphi}_i(t+1)) - f_i(\mathbf{z}_i(t+1) - \sigma_i \boldsymbol{\varphi}_i(t+1))}{2\sigma_i} \boldsymbol{\varphi}_i(t+1)。$$

引理 2^[15] 在假设 1 和 2 成立情形下,并且还假设存在常数 $D>0$,使得 $E[\|\alpha(t)\tilde{\mathbf{g}}_i(t)\|_1] \leq \frac{D}{t}$ 成立,那么对

任意的 $t \geq 1$,有 $E\left[\sum_{t=1}^{\tau} \left\| \mathbf{z}_i(t+1) - \frac{\sum_{j=1}^n \mathbf{x}_j(t)}{n} \right\| \right] \leq \frac{8}{\delta} \frac{\lambda}{1-\lambda} \sum_{j=1}^n \|\mathbf{x}_j(0)\|_1 + \frac{8}{\delta} \frac{Dn}{1-\lambda} (1 + \ln \tau)$, 其中常数 $\delta > 0$ 和 $\lambda \in (0, 1)$,并满足 $\delta \geq \frac{1}{n^{nB}}, \lambda \leq \left(1 - \frac{1}{n^{nB}}\right)^{\frac{1}{nB}}$ 。

定义 σ -代数 $\mathcal{F}_{t+1} = \{(\mathbf{z}_i(k), i=1, 2, \dots, n); (\mathbf{x}_i(0), i=1, 2, \dots, n); 1 \leq k \leq t+1\}$ 。下面给出算法 GFD 的主要收敛性结果。

定理 1 若假设 1 和 2 成立,序列 $\{\mathbf{z}_i(t+1)\}_{i=0}^{\infty}$ 由算法 GFD 迭代产生。对任意的智能体 i 和所有的 $t \geq 0$,有:

$$\begin{aligned} E[F(\mathbf{z}_j(t+1)) \mid \mathcal{F}_{t+1}] - F(\mathbf{z}^*) &\leq \frac{n}{\alpha(t+1)} \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 - \frac{n}{\alpha(t+1)} E[\|\bar{\mathbf{x}}(t+1) - \mathbf{z}^*\|^2 \mid \mathcal{F}_{t+1}] - \\ &\quad \frac{1}{2} \left(\sum_{j=1}^n \mu_j \right) \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 + LE[\|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| \mid \mathcal{F}_{t+1}] + \\ &\quad 2(d+5) \sum_{j=1}^n L_j E[\|\mathbf{z}_j(t+1) - \bar{\mathbf{x}}(t)\| \mid \mathcal{F}_{t+1}] + 2\sqrt{d} \sum_{j=1}^n \sigma_j L_j + \alpha(t+1)(d+4)^2 \sum_{j=1}^n L_j^2, \end{aligned} \quad (8)$$

这里 $\bar{\mathbf{x}}(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j(t)$ 。

证明 注意到 $\mathbf{A}(t)$ 是列随机矩阵,由(2)式中 $\mathbf{w}_i(t+1)$ 的定义可得:

$$\sum_{i=1}^n \mathbf{w}_i(t+1) = \sum_{i=1}^n \sum_{j=1}^n [\mathbf{A}(t)]_{ij} \mathbf{x}_j(t) = \sum_{j=1}^n \sum_{i=1}^n [\mathbf{A}(t)]_{ij} \mathbf{x}_j(t) = \sum_{j=1}^n \mathbf{x}_j(t)。$$

又因为:

$$\bar{\mathbf{x}}(t) = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i(t+1) = \frac{1}{n} \sum_{i=1}^n y_i(t+1) \mathbf{z}_i(t+1)。 \quad (9)$$

结合(5)式和(9)式可得:

$$\bar{\mathbf{x}}(t+1) = \bar{\mathbf{x}}(t) - \frac{\alpha(t+1)}{n} \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1)。 \quad (10)$$

对任意 $\mathbf{v} \in \mathbf{R}^d$,由(10)式,对所有 $t \geq 0$,有:

$$\|\bar{\mathbf{x}}(t+1) - \mathbf{v}\|^2 = \|\bar{\mathbf{x}}(t) - \mathbf{v}\|^2 - \frac{2\alpha(t+1)}{n} \left\langle \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1), \bar{\mathbf{x}}(t) - \mathbf{v} \right\rangle + \frac{\alpha^2(t+1)}{n} \left\| \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1) \right\|^2。 \quad (11)$$

在(11)式中,取期望有:

$$E[\|\bar{\mathbf{x}}(t+1) - \mathbf{v}\|^2 \mid \mathcal{F}_{t+1}] =$$

$$\|\bar{\mathbf{x}}(t) - \mathbf{v}\|^2 - \frac{2\alpha(t+1)}{n} \sum_{j=1}^n \langle E[\tilde{\mathbf{g}}_j(t+1) \mid \mathcal{F}_{t+1}], \bar{\mathbf{x}}(t) - \mathbf{v} \rangle + \frac{\alpha^2(t+1)}{n} E\left[\left\| \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1) \right\|^2 \mid \mathcal{F}_{t+1}\right]。 \quad (12)$$

由引理 1 结论 c) 可得:

$$\begin{aligned} \sum_{j=1}^n \langle E[\tilde{\mathbf{g}}_j(t+1) \mid \mathcal{F}_{t+1}], \bar{\mathbf{x}}(t) - \mathbf{v} \rangle &= \sum_{j=1}^n \langle \nabla \tilde{f}_j(\mathbf{z}_j(t+1)), \mathbf{z}_j(t+1) - \mathbf{v} \rangle + \\ &\quad \sum_{j=1}^n \langle \nabla \tilde{f}_j(\mathbf{z}_j(t+1)), \bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1) \rangle。 \end{aligned} \quad (13)$$

再由 $\tilde{f}_i$ 的凸性和引理 1 的结论 a) 可得:

$$\langle \nabla \tilde{f}_j(\mathbf{z}_j(t+1)), \mathbf{z}_j(t+1) - \mathbf{v} \rangle \geq \tilde{f}_j(\mathbf{z}_j(t+1)) - \tilde{f}_j(\mathbf{v}) \geq f_j(\mathbf{z}_j(t+1)) - f_j(\mathbf{v}) - \sqrt{d} \sigma_j L_j。$$

又因为:

$$\begin{aligned} f_j(\mathbf{z}_j(t+1)) - f_j(\mathbf{v}) &= f_j(\mathbf{z}_j(t+1)) - f_j(\bar{\mathbf{x}}(t)) + (f_j(\bar{\mathbf{x}}(t)) - f_j(\mathbf{v})) \geq \\ &\quad \nabla f_j(\bar{\mathbf{x}}(t))' (\mathbf{z}_j(t+1) - \bar{\mathbf{x}}(t)) + (f_j(\bar{\mathbf{x}}(t)) - f_j(\mathbf{v}))。 \end{aligned}$$

利用 f_j 的 Lipschitz 连续性,有 $\nabla f_j(\bar{\mathbf{x}}(t))' (\mathbf{z}_j(t+1) - \bar{\mathbf{x}}(t)) \geq -L_j \|\mathbf{z}_j(t+1) - \bar{\mathbf{x}}(t)\|$ 。于是有:

$$\langle \nabla \tilde{f}_j(\mathbf{z}_j(t+1)), \mathbf{z}_j(t+1) - \mathbf{v} \rangle \geq -L_j \|\mathbf{z}_j(t+1) - \bar{\mathbf{x}}(t)\| + f_j(\bar{\mathbf{x}}(t)) - f_j(\mathbf{v}) - \sqrt{d} \sigma_j L_j. \quad (14)$$

由引理 1 的结论 c) 得:

$$\|\nabla \tilde{f}_j(\mathbf{z}_j(t+1))\| = \|E[(\tilde{\mathbf{g}}_j(t+1)) | \mathcal{F}_{t+1}]\| \leq E[\|(\tilde{\mathbf{g}}_j(t+1))\|^2 | \mathcal{F}_{t+1}]^{\frac{1}{2}} \leq (d+4)L_j. \quad (15)$$

考虑(13)式右边的第二项,有:

$$\langle \nabla \tilde{f}_j(\mathbf{z}_j(t+1)), \bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1) \rangle \geq -(d+4)L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\|. \quad (16)$$

结合(13),(14)和(16)式,有:

$$\sum_{j=1}^n \langle E[\tilde{\mathbf{g}}_j(t+1) | \mathcal{F}_{t+1}], \bar{\mathbf{x}}(t) - \mathbf{v} \rangle \geq F(\bar{\mathbf{x}}(t)) - F(\mathbf{v}) - (d+5) \sum_{j=1}^n L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\| - \sqrt{d} \sum_{j=1}^n \sigma_j L_j. \quad (17)$$

考虑(12)式右边的第三项,利用 $\left\| \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1) \right\|^2 \leq n \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1)^2$ 和引理 1 的结论 c),有:

$$\begin{aligned} \frac{\alpha^2(t+1)}{n^2} E\left[\left\| \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1) \right\|^2 | \mathcal{F}_{t+1}\right] &\leq \frac{\alpha^2(t+1)}{n} \sum_{j=1}^n E\left[\left\| \sum_{j=1}^n \tilde{\mathbf{g}}_j(t+1) \right\|^2 | \mathcal{F}_{t+1}\right] \leq \\ &\frac{\alpha^2(t+1)}{n} (d+4)^2 \sum_{j=1}^n L_j^2. \end{aligned} \quad (18)$$

于是由(12),(17)和(18)式,可以得到:

$$\begin{aligned} E[\|\bar{\mathbf{x}}(t+1) - \mathbf{v}\|^2 | \mathcal{F}_{t+1}] &\leq \|\bar{\mathbf{x}}(t) - \mathbf{v}\|^2 - \frac{2\alpha(t+1)}{n} F(\bar{\mathbf{x}}(t)) - F(\mathbf{v}) + \\ &\frac{2\alpha(t+1)}{n} (d+5) \sum_{j=1}^n L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\| + \frac{2\alpha(t+1)}{n} \sqrt{d} \sum_{j=1}^n \sigma_j L_j + \frac{\alpha^2(t+1)}{n} (d+4)^2 \sum_{j=1}^n L_j^2. \end{aligned} \quad (19)$$

在(19)式中,令 $\mathbf{v} = \mathbf{z}^*$, 对所有的 $t \geq 0$, 有:

$$\begin{aligned} E[\|\bar{\mathbf{x}}(t+1) - \mathbf{z}^*\|^2 | \mathcal{F}_{t+1}] &\leq \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 - \frac{2\alpha(t+1)}{n} F(\bar{\mathbf{x}}(t)) - F(\mathbf{z}^*) + \\ &\frac{2\alpha(t+1)}{n} (d+5) \sum_{j=1}^n L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t)\| + \frac{2\alpha(t+1)}{n} \sqrt{d} \sum_{j=1}^n \sigma_j L_j + \frac{\alpha^2(t+1)}{n} (d+4)^2 \sum_{j=1}^n L_j^2. \end{aligned} \quad (20)$$

一方面,由假设 1 的结论 a) 知, $F(\mathbf{x})$ 是强凸函数, 因此有:

$$F(\bar{\mathbf{x}}(t)) - F(\mathbf{z}^*) \geq \frac{1}{2} \left(\sum_{j=1}^n \mu_j \right) \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2. \quad (21)$$

另一方面,由假设 1 的结论 b) 知 F 是 L -Lipschitz 连续的, 有:

$$\begin{aligned} F(\bar{\mathbf{x}}(t)) - F(\mathbf{z}^*) &= F(\bar{\mathbf{x}}(t)) - F(\mathbf{z}_i(t+1)) + F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*) \geq \\ &-L \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| + F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*). \end{aligned} \quad (22)$$

(21)式与(22)式相加可得:

$$2[F(\bar{\mathbf{x}}(t)) - F(\mathbf{z}^*)] \geq \frac{1}{2} \left(\sum_{j=1}^n \mu_j \right) \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 - L \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| + F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*). \quad (23)$$

又由(20)式与(23)式,可得:

$$\begin{aligned} E[\|\bar{\mathbf{x}}(t+1) - \mathbf{z}^*\|^2 | \mathcal{F}_{t+1}] &\leq \\ &\|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 - \frac{\alpha(t+1)}{2n} \left(\sum_{j=1}^n \mu_j \right) \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 + \frac{\alpha(t+1)}{n} L \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| - \\ &\frac{\alpha(t+1)}{n} (F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*)) + \frac{2\alpha(t+1)}{n} (d+5) \sum_{j=1}^n L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\| + \\ &\frac{2\alpha(t+1)}{n} \sqrt{d} \sum_{j=1}^n \sigma_j L_j + \frac{\alpha^2(t+1)}{n} (d+4)^2 \sum_{j=1}^n L_j^2. \end{aligned} \quad (24)$$

对(24)式两边取期望,即得定理 1 的结果。

证毕

下面将给出算法 GFD 的收敛率。

定理 2 在定理 1 的条件下,令:

$$\alpha(t) = \frac{p}{t}, t \geq 1, \quad (25)$$

其中常数 p 满足:

$$p \frac{\sum_{i=1}^n \mu_i}{n} \geq 4. \quad (26)$$

对所有 $\tau \geq 2$, 有 $E[F(\hat{z}_i(\tau))] - F(\mathbf{z}^*) \leq \frac{K_1 \ln \tau + K_2}{\tau} + 4\sqrt{d} \sum_{j=1}^n \sigma_j L_j$ 成立。其中:

$$\hat{z}_i(t) = \frac{\sum_{s=1}^t (s-1) \mathbf{z}_i(s)}{\frac{t(t-1)}{2}}, t \geq 2, K_1 = \frac{16npL(2d+11)(d+4) \max_j L_j}{\delta(1-\lambda)},$$

$$K_2 = \frac{16L\lambda(2d+11)}{\delta(1-\lambda)} \sum_{j=1}^n \|\mathbf{x}_j(0)\|_1 + \frac{16L(2d+11)}{\delta} \frac{np(d+4) \max_j L_j}{1-\lambda} + 2p(d+4)^2 \sum_{j=1}^n L_j^2.$$

证明 将(25),(26)式代入(24)式中可得:

$$E[\|\bar{\mathbf{x}}(t+1) - \mathbf{z}^*\|^2 | \mathcal{F}_{t+1}] \leq \left(1 - \frac{2}{t+1}\right) \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 - \frac{p}{n(t+1)} (F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*)) + \frac{pL}{n(t+1)} \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| + \frac{2p(d+5)}{n(t+1)} \sum_{j=1}^n L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\| + \frac{2p\sqrt{d}}{n(t+1)} \sum_{j=1}^n \sigma_j L_j + \frac{p^2(d+4)^2}{n(t+1)^2} \sum_{j=1}^n L_j^2.$$

将以上不等式两边同时乘以 $t(t+1)$, 有:

$$t(t+1)E[\|\bar{\mathbf{x}}(t+1) - \mathbf{z}^*\|^2 | \mathcal{F}_{t+1}] \leq t(t-1) \|\bar{\mathbf{x}}(t) - \mathbf{z}^*\|^2 - \frac{pt}{n} (F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*)) + \frac{pLt}{n} \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\| + \frac{2pt\sqrt{d}}{n} \sum_{j=1}^n \sigma_j L_j + \frac{2pt(d+5)}{n} \sum_{j=1}^n L_j \|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\| + \frac{p^2t(d+4)^2}{n(t+1)} \sum_{j=1}^n L_j^2.$$

在上式两边, 首先对 t 从 1 到 $\tau-1$ 求和, 然后两边同时取期望有:

$$\tau(\tau-1)E[\|\bar{\mathbf{x}}(\tau) - \mathbf{z}^*\|^2] \leq -\frac{p}{n} \sum_{t=1}^{\tau-1} tE[F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*)] + \frac{pL}{n} \sum_{t=1}^{\tau-1} tE[\|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t)\|] + \frac{2p(d+5)}{n} \sum_{t=1}^{\tau-1} t \sum_{j=1}^n L_j E[\|\bar{\mathbf{x}}(t) - \mathbf{z}_j(t+1)\|] + \frac{2p\sqrt{d}}{n} \sum_{t=1}^{\tau-1} t \sum_{j=1}^n \sigma_j L_j + \frac{p^2(d+4)^2}{n} \sum_{t=1}^{\tau-1} \frac{t}{t+1} \sum_{j=1}^n L_j^2. \quad (27)$$

由(15)式知, $E[\|\alpha(t)\tilde{\mathbf{g}}_i(t)\|_1] \leq \frac{p(d+4)L_i}{t}$, 从而满足引理 2 的条件。因此有:

$$E\left[\sum_{t=1}^{\tau-1} \left\| \mathbf{z}_i(t+1) - \frac{\sum_{j=1}^n \mathbf{x}_j(t)}{n} \right\|\right] \leq \frac{8}{\delta} \frac{\lambda}{1-\lambda} \sum_{j=1}^n \|\mathbf{x}_j(0)\|_1 + \frac{8}{\delta} \frac{np(d+4) \max_j L_j}{1-\lambda} (1 + \ln(\tau-1)). \quad (28)$$

将(28)式代入(27)式中, 两边同时除以 $\tau(\tau-1)$, 有:

$$\frac{p}{n\tau(\tau-1)} \sum_{t=1}^{\tau-1} tE[F(\mathbf{z}_i(t+1)) - F(\mathbf{z}^*)] \leq \frac{pL}{n\tau} \frac{8}{\delta} \left[\frac{\lambda}{1-\lambda} \sum_{j=1}^n \|\mathbf{x}_j(0)\|_1 + \frac{np(d+4) \max_j L_j}{1-\lambda} (1 + \ln(\tau-1)) \right] + \frac{2pL(d+5)}{n\tau} \frac{8}{\delta} \left[\frac{\lambda}{1-\lambda} \sum_{j=1}^n \|\mathbf{x}_j(0)\|_1 + \frac{np(d+4) \max_j L_j}{1-\lambda} (1 + \ln(\tau-1)) \right] + \frac{2p\sqrt{d}}{n} \sum_{j=1}^n \sigma_j L_j + \frac{p^2(d+4)^2}{n\tau} \sum_{j=1}^n L_j^2.$$

最后, 由 F 的凸性即可证得定理 2 的结果。

证毕

注 2 定理 2 的收敛结果分成两部分, 其中第一部分表示优化误差项, 第二部分表示光滑化误差项。定理 2

表明, 当 τ 逐渐增大时, $E[F(\hat{z}_i(\tau))]$ 逐渐收敛到 $F(\mathbf{z}^*)$, 收敛率是 $O(\ln \tau / \tau)$, 并且带有光滑化误差项 $4\sqrt{d} \sum_{j=1}^n \sigma_j L_j$ 。

特别地, 当所有的光滑化参数 σ_j 取得足够小的时候, 光滑化误差项 $4\sqrt{d} \sum_{j=1}^n \sigma_j L_j$ 也足够小, 从而 $E[F(\hat{z}_i(\tau))] - F(\mathbf{z}^*)$ 也足够小。对比文献[7]中的 push-sum 次梯度算法(简称算法 SGD), 算法 GFD 多了光滑化误差项, 收敛率略差于算法 SGD, 这是因为本文的算法 GFD 仅仅使用了函数值信息。

3 数值实验

考虑一个正则化的 LASSO 问题^[5]:

$$\min_{z \in \mathbf{R}^d} F(\mathbf{z}) \triangleq \frac{1}{2} \sum_{i=1}^n (\langle \mathbf{a}_i, \mathbf{z} \rangle - b_i)^2 + \rho \|\mathbf{z}\|_1,$$

这里 ρ 是一个正的权参数, $\mathbf{a}_i \in \mathbf{R}^d, b_i \in \mathbf{R}, \|\cdot\|_1$ 表示 l_1 范数。

对任意的 $i=1,2,\cdots,n$, 假设点对 $(\mathbf{a}_i, b_i)$ 是在区间 $(0,1)$ 上随机产生, 且服从均匀分布。取步长 $\alpha(t)$ 满足定理 2 的条件。取 $\rho=0.1$, 取光滑化参数 $\sigma_i=10^{-6}, i=1,2,\cdots,n$ 。对于时变的权矩阵, 首先随机生成 50 个满足假设 2 的矩阵; 然后每次迭代从中选取一个。考虑最大误差 $\max_{\text{error}} = \max_{i \in V} [F(\hat{\mathbf{z}}_i(\tau)) - F(\mathbf{z}^*)]$ 。将本文的算法 GFD 与算法 SGD^[7] 进行对比。

图 1 表示智能体总个数 n 为 100, 问题维数 d 分别为 1 和 3 时, 最大误差 $\max_{\text{error}}$ 的演化图。图 2 表示智能体总个数 n 为 200, 问题维数 d 分别为 1 和 3 时, 最大误差 $\max_{\text{error}}$ 的演化图。由于算法 GFD 仅仅使用了函数值信息, 在理论上算法 GFD 的收敛率略差于算法 SGD, 但从数值试验中可以看出, 算法 GFD 与算法 SGD 具有相似的收敛效果。另外, 对比图 1a, b 发现, 随着问题维数的增加, 算法 GFD 的收敛误差也随之增大。图 2 也展示了相似的结果。

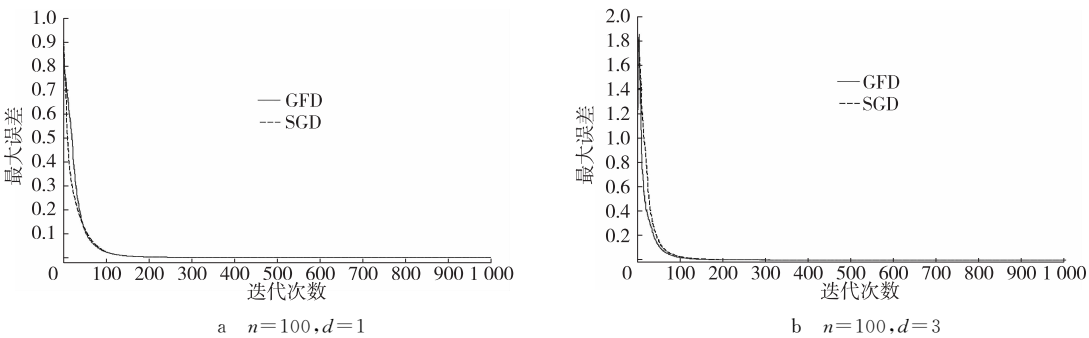


图 1 智能体总个数为 100 的最大误差对比
Fig. 1 Max error versus number of iterations with

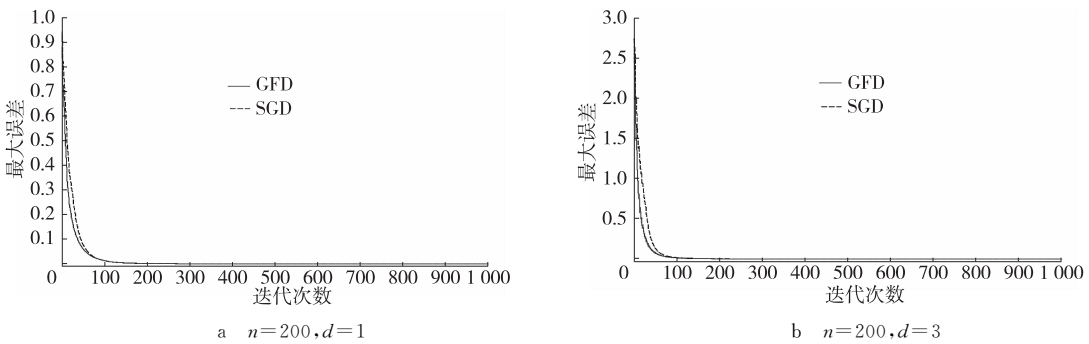


图 2 智能体总个数为 200 的最大误差对比
Fig. 2 Max error versus number of iterations with $n=200$

4 结语

针对分布式优化问题, 许多实际问题中的目标函数是不可微的, 其次梯度往往难以计算, 有时甚至不可用。基于此, 本文提出了一种分布式 gradient-free 算法, 分析了算法的收敛性, 并给出了算法的收敛率。通过数值试验, 验证了算法的可行性和有效性。与文献[7]的分布式次梯度算法相比, 本文的算法不需要计算目标函数的次梯度, 仅仅需要计算函数值。而与文献[14]的分布式 gradient-free 算法不同的是, 本文的算法适用于有向网络。

参考文献:

[1] 洪奕光, 张艳琼. 分布式优化: 算法设计和收敛性分析[J]. 控制理论与应用, 2014, 31(7): 850-857.

- HONG Y G, ZHANG Y Q. Distributed optimization; algorithm design and convergence analysis[J]. Control Theory and Applications, 2014, 31(7): 850-857.
- [2] 洪奕光, 翟超. 多智能体系统动态协调与分布式控制设计[J]. 控制理论与应用, 2011, 28(10): 1506-1512.
- HONG Y G, ZHAI C. Dynamic coordination and distributed control design of multi-agent systems[J]. Control Theory & Applications, 2011, 28(10): 1506-1512.
- [3] 李德权, 陈平. 基于概率量化的分布式无梯度优化算法研究[J]. 皖西学院报, 2015(5): 25-28.
- LI D Q, CHEN P. Study on distributed free-gradient optimization algorithm based on probabilistic quantization[J]. Journal of West Anhui University, 2015(5): 25-28.
- [4] YUAN D M, XU S Y, LU J W. Gradient-free method for distributed multi-agent optimization via push-sum algorithms[J]. Robust Nonlinear Control, 2015, 25(10): 1569-1580.
- [5] NEDIC A, OLSHEVSKY A, OZDAGLAR A, et al. On distributed averaging algorithms and quantization effects[J]. IEEE Transactions on Automatic Control, 2009, 54(11): 2506-2517.
- [6] NEDIC A, OZDAGLAR A. Distributed subgradient methods for multi-agent optimization[J]. IEEE Transactions on Automatic Control, 2009, 54(1): 48-61.
- [7] NEDIC A, OLSHEVSKY A. Distributed optimization over time-varying directed graphs[J]. IEEE Transactions on Automatic Control, 2014, 60(3): 601-615.
- [8] 唐朝君. 切换拓扑下离散时间多智能体系统的包含控制[J]. 重庆理工大学学报(自然科学), 2017, 31(5): 143-147.
- TANG C J. Containment control of discrete-time multi-agent systems with switching topologies[J]. Journal of Chongqing University of Technology (Natural Science), 2017, 31(5): 143-147.
- [9] 龙强, 李觉友. 次梯度法在求解非光滑最优化问题时的计算效果研究[J]. 重庆师范大学学报(自然科学版), 2013, 30(6): 25-30.
- LONG Q, LI J Y. Numerical performance of subgradient methods in solving nonsmooth optimization problems[J]. Journal of Chongqing Normal University (Natural Science), 2013, 30(6): 25-30.
- [10] 龙强. 基因算法在求解非光滑优化问题中的应用[J]. 重庆师范大学学报(自然科学版), 2013, 30(1): 12-15.
- LONG Q. The application of genetic algorithm in solving nonsmooth optimization problems[J]. Journal of Chongqing Normal University (Natural Science), 2013, 30(1): 12-15.
- [11] 万轩, 赵克全. 非光滑 B-预不变凸优化问题的解集刻画[J]. 重庆师范大学学报(自然科学版), 2012, 29(6): 1-5.
- WAN X, ZHAO K Q. On characterizing solution sets of nonsmooth B-preinvex optimization problems[J]. Journal of Chongqing Normal University (Natural Science), 2012, 29(6): 1-5.
- [12] DUCHI J C, JORDAN M I, WAINWRIGHT M J, et al. Optimal rates for zero-order convex optimization: the power of two function evaluations[J]. IEEE Transactions on Information Theory, 2015, 61(5): 2788-2806.
- [13] 赵克全, 唐莉萍. 一类非可微多目标分式规划问题的混合对偶[J]. 重庆师范大学学报(自然科学版), 2011, 28(3): 1-3.
- ZHAO K Q, TANG L P. The mixed duality of a kind of non-differentiable multi objective fractional[J]. Journal of Chongqing Normal University (Natural Science), 2011, 28(3): 1-3.
- [14] LI J Y, WU C Z, WU Z Y. Gradient-free method for non-smooth distributed optimization[J]. Journal of Global Optimization, 2017, 13(2): 21-21.
- [15] NEDIC A, OLSHEVSKY A. Stochastic gradient-push for strongly convex functions on time-varying directed graphs[J]. IEEE Transactions on Automatic Control, 2016, 61(12): 36-47.

Operations Research and Cybernetics

Gradient-Free Push-Sum Method for Strongly Convex Distributed Optimization

YU Shuhui, LI Jueyou, DU Xuewu

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] To solve the distributed convex optimization problem in multi-agent systems, a gradient-free algorithm with push-sum protocol is proposed. [Methods] Based on push-sum subgradient algorithm, the subgradient is replaced with gradient-free by using Gaussian smoothing, and prove the convergence of the proposed gradient-free push-sum algorithm. [Findings] The convergence rate of the proposed gradient-free push-sum algorithm is order of $\frac{1}{\sqrt{k}}$, where k is the number of iterations. [Conclusions] The numerical example shows that the proposed algorithm has the similar convergence as the subgradient algorithm.

Keywords: multi-agent system; distributed convex optimization; gradient-free; push-sum