第 44 卷 第 11 期 西南师范大学学报(自然科学版) 2019 年 11 月
Vol. 44 No. 11 Journal of Southwest China Normal University (Natural Science Edition) Nov. 2019

DOI:10.13718/j.cnki.xsxb.2019.11.002

一类时变有向图中的 PUSH-SUM 分布式对偶平均优化算法①

周小清[1], 李觉友[2]

1. 重庆龙山中学,重庆 401147;2. 重庆师范大学 数学科学学院,重庆 401331

摘要: 利用 push-sum 通信协议并结合分布式对偶平均方法,在时变有向图中,讨论了一类带有简单约束集的分布式凸优化问题.首先提出了 push-sum 分布式对偶平均算法,然后分析了算法的收敛性,并得到了算法的收敛率为 $O\left(\frac{1}{/T}\right)$,最后用 l_1 线性回归问题的数值结果验证了所提出算法的有效性.对比现有的一些结果,所提出的算法能用于求解带约束的分布式优化问题,并且去掉了网络通讯权矩阵是双随机的限制.

关 键 词: 分布式对偶平均;push-sum 算法;收敛性分析;凸优化;时变网络

中图分类号:O224;O236 **文献标志码:A** **文章编号:** 1000-5471(2019)11-0011-07

分布式优化理论及应用已经成为当前系统和控制科学的重要研究方向之一.在优化理论研究过程中,算法的设计、收敛性分析和复杂性分析是研究的关键点[1].目前已有许多国内外学者研究了分布式优化的理论和算法[2-6].基于次梯度信息,现有的分布式优化算法大致可分为 3 类:primal 分布式算法、dual 分布式算法和 primal-dual 分布式算法.文献[3]最早提出了一种分布式次梯度算法来求解无约束的分布式优化问题,并给出了收敛性分析.随后,文献[4]在固定无向通讯网络中提出了分布式对偶平均次梯度算法,并给出了算法的收敛性分析,得到了收敛率依赖于网络规模和拓扑结构的结论.文献[5]基于 Lagrangian 对偶方法提出了 primal-dual 分布式次梯度算法来求解一类带有等式或不等式约束的分布式凸优化问题.目前分布式方法已广泛应用于各个领域[7-12].

现有的分布式算法大多是基于无向通讯网络设计的,而许多实际问题的通讯网络往往是有向的,比如无线传感器网络、手机通讯网络等[13].而在无向网络中,分布式算法的设计往往需要一个很强的条件,即要求通讯权矩阵是双随机的,这样的要求对于有向通讯网络来说通常是不能满足的[14].最近,文献[13]通过引入 push-sum 协同机制,首次考虑了时间不变的有向通讯网络(假设其通讯权矩阵仅是列随机的)下的分布式优化问题,提出了 push-sum 对偶平均次梯度算法,给出了算法的收敛性分析,但他们考虑的通讯网络是时间不变的.随后,文献[14]考虑了时变有向网络下的分布式优化问题,并得到了算法的收敛率,但作者仅考虑了无约束的分布式优化问题.于是,在时变有向网络情形下,对带约束的分布式问题的研究就显得尤为必要.

受文献[13-14]的启发,本文提出了在时变有向网络下的 push-sum 分布式对偶平均次梯度算法,分析了算法的收敛性,并得到了算法的收敛率为 $\mathrm{O}\left(\frac{1}{/T}\right)$.数值结果验证了所提出算法的有效性.本文将文献

① 收稿日期:2018-05-04

基金项目:国家自然科学基金青年项目(11971083);重庆市自然科学基金项目(cstc2017jcyjAX0253);重庆市教委科学技术研究项目(KJQN201800520).

作者简介:周小清(1993-),女,硕士研究生,主要从事分布式优化理论和算法研究.

通信作者:李觉友,教授.

[13]中的时间不变的有向网络推广到了时变的有向网络中，且能用于求解带约束的分布式优化问题.

1 问题描述

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，并通过与其它智能体交换信息来协同达到整体最优目标值[14]. 本文主要考虑如下分布式约束优化问题[1]：

$$\min_{\boldsymbol{x}\in X} F(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} f_i(\boldsymbol{x}) \tag{1}$$

其中：智能体 i 仅知道它自身的目标函数 $f_i(\boldsymbol{x})$，$i=1,\cdots,n$；所有的智能体受到公共约束集 $X \subset \mathbb{R}^d$ 的约束. 令 X^* 为问题(1)的最优解集.

2 Push-sum 分布式对偶平均算法

假设考虑的图 $G(t)=(V, E(t))$ 是有向图，其中 $V=\{1, 2, \cdots, n\}$ 为节点或智能体集合，$E(t) \subset V \times V$ 为边集合. 对于节点 i，用 $N_i^{in}(t)=\{j \mid (j, i) \in E(t)\} \cup \{i\}$ 来表示它的内邻居集合；用 $N_i^{out}(t)=\{j \mid (i, j) \in E(t)\} \cup \{i\}$ 来表示它的外邻居集合；用 $d_i(t)=\mid N_i^{out}(t) \mid$ 来表示它的外度. 假设每个智能体 i 在每个时刻 t 仅知道它的外度 $d_i(t)$，定义 $\boldsymbol{A}(t)$ 为通讯权矩阵，其元素为

$$A_{ij}(t)=\begin{cases}\dfrac{1}{d_j(t)}, & j \in N_i^{in}(t)\\ 0, & \text{否则}\end{cases}$$

通常 $\boldsymbol{A}(t)$ 是时变的列随机矩阵[14]. 为了求解问题(1)，提出如下 push-sum 分布式对偶平均算法，简记为算法 PS-DDA：

步 0：初始化 $w_i(0)=1$，$\boldsymbol{z}_i(0)=\boldsymbol{0}$，$\boldsymbol{g}_i(0)=\boldsymbol{0}$，$t=0$；

步 1：对每个智能体 i，$i=1,\cdots,n$，更新

$$w_i(t+1)=\sum_{j=1}^{n} A_{ij}(t) w_j(t) \tag{2}$$

$$\boldsymbol{z}_i(t+1)=\sum_{j=1}^{n} A_{ij}(t) \boldsymbol{z}_j(t)+\boldsymbol{g}_i(t) \tag{3}$$

$$\boldsymbol{y}_i(t+1)=\frac{\boldsymbol{z}_i(t+1)}{w_i(t+1)} \tag{4}$$

$$\boldsymbol{x}_i(t+1)=\Pi_X^{\Psi}(\boldsymbol{y}_i(t+1), a(t)) \tag{5}$$

步 2：令 $t=t+1$，转步 1.

其中步长序列 $\{a(t)\}_{t=0}^{\infty}$ 是非负单调递减的；$\boldsymbol{g}_i(t)$ 是 $f_i(\boldsymbol{x})$ 在 $\boldsymbol{x}=\boldsymbol{x}_i(t)$ 的次梯度；$\Pi_X^{\psi}(\boldsymbol{z}, a(t)) := \arg\min\limits_{\boldsymbol{x}\in \boldsymbol{X}}\left\{\langle \boldsymbol{z}, \boldsymbol{x}\rangle+\dfrac{\psi(\boldsymbol{x})}{a(t)}\right\}$ 是一个非欧几里得投影算子，这里 $\psi: \mathbb{R}^d \longrightarrow \mathbb{R}$ 是一个 1- 强凸函数[4]，并且满足 $\psi(\boldsymbol{x}) \geqslant 0$ 和 $\psi(\boldsymbol{0})=0$.

3 收敛性分析

下面将给出算法 PS-DDA 的收敛性分析.

假设 1 (a) 有向图序列 $\{G(t)\}$ 是 B-强连通，即存在一个正整数 B 使得对于每个 $k \geqslant 0$，$E_B(k)=\bigcup\limits_{i=kB}^{(k+1)B-1} E(i)$ 是强连通的[11].

(b) 对 $i=1,\cdots,n$，$f_i(x): X \longrightarrow \mathbb{R}$ 是凸函数，并且其次梯度 $\boldsymbol{g}_i(\boldsymbol{x})$ 是有界的，即对任意 $\boldsymbol{x} \in \boldsymbol{X}$，有 $\| \boldsymbol{g}_i(\boldsymbol{x}) \|_* \leqslant L$，这里 $\boldsymbol{g}_i(\boldsymbol{x}) \in \partial f_i(\boldsymbol{x})$，$\| \cdot \|_*$ 是 $\| \cdot \|$ 的对偶范数.

引理 1[4] 假设步长序列 $\{a(t)\}_{t=0}^{\infty}$ 是非负单调递减的. 对于 $\forall \boldsymbol{x}^* \in \boldsymbol{X}^*$，有

$$\sum_{t=1}^{T}\langle \hat{\boldsymbol{g}}(t), \boldsymbol{x}(t)-\boldsymbol{x}^*\rangle \leqslant \frac{1}{2}\sum_{t=1}^{T} a(t-1)\|\hat{\boldsymbol{g}}(t)\|_*^2+\frac{\psi(\boldsymbol{x}^*)}{a(T)}$$

这里 $\hat{\boldsymbol{g}}(t)=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t)$，$\boldsymbol{x}(t+1)=\Pi_X^{\psi}(\sum_{r=1}^{t}\hat{\boldsymbol{g}}(r),\ a(t))$.

引理 2[4]　对于任意两个向量 $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^d$，有 $\|\Pi_X^{\psi}(\boldsymbol{u},\ a)-\Pi_X^{\psi}(v,\ a)\|\leqslant a\|\boldsymbol{u}-\boldsymbol{v}\|_*$.

引理 3[14]　假设图序列$\{G(t)\}$是一致强连通的，则

(a) 存在一个随机向量序列$\{\varphi(t)\}_{t=1}^{\infty}$，使得对于 $t\geqslant s\geqslant 0$，有

$$|[A(t:s)]_{ij}-\varphi_i(t)|\leqslant C\lambda^{t-s},\ i,j=1,\cdots,n$$

其中 $A(t:s):=A(t)\cdots A(s)$，$t\geqslant s\geqslant 0$，$C>0$ 为常数，$\lambda\in(0,1)$.

(b) 令 $\delta=\inf\limits_{t=0,1,\cdots}(\min\limits_{1\leqslant i\leqslant n}[A(t)\cdots A(0)\mathbf{1}]_i)$，有 $\delta\geqslant\frac{1}{n^{nB}}$，这里 1 是一个 n 维的全 1 向量.

引进如下记号：$\hat{\boldsymbol{x}}_i(T)=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_i(t)$，$\bar{\boldsymbol{z}}(t)=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i(t)$，$\boldsymbol{y}(t)=\Pi_X^{\psi}(\bar{\boldsymbol{z}}(t),\ a(t-1))$，$\hat{\boldsymbol{y}}(t)=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}(t)$. 为了得到算法 PS-DDA 的收敛性，首先证明如下一个重要引理.

引理 4　假设 1 成立，序列$\{\boldsymbol{z}_i(t)\}_{t=1}^{\infty}$是由算法 PS-DDA 产生的. 对于所有 $t\geqslant 1$，则有

$$\left\|\frac{\boldsymbol{z}_i(t)}{w_i(t)}-\bar{\boldsymbol{z}}(t)\right\|_*\leqslant\frac{2CL}{\delta\lambda(1-\lambda)}+\frac{L}{n\delta}$$

这里$\frac{1}{n^{nB}}\leqslant\delta<1$，$0<\lambda\leqslant\left(1-\frac{1}{n^{nB}}\right)^{\frac{1}{B}}$.

证　由式(2) 递推可得

$$\begin{aligned}w_i(t+1)=&\sum_{j=1}^{n}A_{ij}(t)w_j(t)=\sum_{j=1}^{n}A_{ij}(t)(\sum_{l=1}^{n}A_{jl}(t-1)w_l(t-1))=\\&\sum_{j=1}^{n}A_{ij}(t)(\sum_{l=1}^{n}A_{jl}(t-1)(\sum_{k=1}^{n}A_{lk}(t-2)w_k(t-2)))=\cdots=\\&\sum_{l=1}^{n}[A(t:0)]_{il}w_i(0)=\sum_{l=1}^{n}[A(t:0)]_{il}\end{aligned}$$

由式(3) 递推可得

$$\begin{aligned}&\boldsymbol{z}_i(t+1)=\sum_{j=1}^{n}A_{ij}(t)\boldsymbol{z}_j(t)+\boldsymbol{g}_i(t)=\sum_{j=1}^{n}A_{ij}(t)(\sum_{l=1}^{n}A_{jl}(t-1)\boldsymbol{z}_l(t-1)+\boldsymbol{g}_j(t-1))+\boldsymbol{g}_i(t)=\\&\sum_{j=1}^{n}A_{ij}(t)\sum_{l=1}^{n}A_{jl}(t-1)(\sum_{k=1}^{n}A_{lk}(t-2)\boldsymbol{z}_k(t-2)+\boldsymbol{g}_l(t-2))+\sum_{j=1}^{n}A_{ij}(t)\boldsymbol{g}_j(t-1)+\boldsymbol{g}_i(t)=\\&\sum_{j=1}^{n}A_{ij}(t)\sum_{l=1}^{n}A_{jl}(t-1)\sum_{k=1}^{n}A_{lk}(t-2)\boldsymbol{z}_k(t-2)+\sum_{j=1}^{n}A_{ij}(t)\sum_{l=1}^{n}A_{jl}(t-1)\boldsymbol{g}_l(t-2)+\\&\sum_{j=1}^{n}A_{ij}(t)\boldsymbol{g}_j(t-1)+\boldsymbol{g}_i(t)=\cdots=\\&\sum_{l=1}^{n}[A(t:0)]_{il}\boldsymbol{z}_i(0)+\sum_{s=1}^{t}\sum_{l=1}^{n}[A(t:s)]_{il}\boldsymbol{g}_l(s-1)+\boldsymbol{g}_i(t)\end{aligned}$$

进一步，由以上关系可得

$$\begin{aligned}\bar{\boldsymbol{z}}(t)=&\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i(t)=\frac{1}{n}\sum_{i=1}^{n}(\sum_{j=1}^{n}A_{ij}(t-1)\boldsymbol{z}_j(t-1)+\boldsymbol{g}_i(t-1))=\\&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(t-1)\boldsymbol{z}_j(t-1)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t-1)=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}A_{ij}(t-1)\boldsymbol{z}_j(t-1)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t-1)=\\&\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{z}_j(t-1)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t-1)=\frac{1}{n}\sum_{j=1}^{n}(\sum_{l=1}^{n}A_{jl}(t-2)\boldsymbol{z}_l(t-2)+\boldsymbol{g}_j(t-2))+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t-1)=\\&\frac{1}{n}\sum_{l=1}^{n}\boldsymbol{z}_l(t-2)+\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{g}_j(t-2)+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t-1)=\cdots=\frac{1}{n}\sum_{r=1}^{t}\sum_{j=1}^{n}\boldsymbol{g}_j(r-1)=\frac{1}{n}\sum_{r=1}^{t-1}\sum_{j=1}^{n}\boldsymbol{g}_j(r)\end{aligned}$$

综合以上结果，并注意到 $\boldsymbol{z}_i(0)=\mathbf{0}$，于是，对于 $t\geqslant 0$ 有

$$
\begin{aligned}
&\left\| \frac{\boldsymbol{z}_i(t)}{w_i(t)} - \bar{\boldsymbol{z}}(t) \right\|_* = \\
&\left\| \frac{\sum_{l=1}^{n} [A(t-1:0)]_{il} \boldsymbol{z}_i(0) + \sum_{s=1}^{t} \sum_{l=1}^{n} [A(t-1:s)]_{il} \boldsymbol{g}_l(s-1) + \boldsymbol{g}_i(t-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} - \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} \boldsymbol{g}_l(s-1)}{n} \right\|_* = \\
&\left\| \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} ([A(t-1:s)]_{il} - \varphi_i(t-1) + \varphi_i(t-1)) \boldsymbol{g}_l(s-1) + \boldsymbol{g}_i(t-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} - \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} \boldsymbol{g}_l(s-1)}{n} \right\|_* \leqslant \\
&\left\| \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} ([A(t-1:s)]_{il} - \varphi_i(t-1)) \boldsymbol{g}_l(s-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\| + \left\| \frac{\boldsymbol{g}_i(t-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\|_* + \\
&\left\| \frac{\sum_{l=1}^{n} [A(t-1:0)]_{il} \sum_{s=1}^{t} \sum_{l=1}^{n} \boldsymbol{g}_l(s-1) - n \sum_{s=1}^{t} \sum_{l=1}^{n} \varphi_i(t-1) \boldsymbol{g}_l(s-1)}{n \sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\|_*
\end{aligned}
$$

由文献[14]的推论 2 知，$[A(t-1:0)]_{il} \leqslant \max\limits_{1 \leqslant l \leqslant n} [A(t-1:0)]_{il}$，$i = 1,2,\cdots,n$，进而可得

$$
\begin{aligned}
\left\| \frac{\boldsymbol{z}_i(t)}{w_i(t)} - \bar{\boldsymbol{z}}(t) \right\|_* \leqslant & \left\| \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} ([A(t-1:s)]_{il} - \varphi_i(t-1)) g_l(s-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\|_* + \left\| \frac{g_i(t-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\|_* + \\
& \left\| \frac{n \sum_{s=1}^{t} \sum_{l=1}^{n} A(t-1:0)]_{il} \boldsymbol{g}_l(s-1) - n \sum_{s=1}^{t} \sum_{l=1}^{n} \varphi_i(t-1) \boldsymbol{g}_l(s-1)}{n \sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\|_* \leqslant \\
& \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} | [A(t-1:s)]_{il} - \varphi_i(t-1) | | g_l(s-1) |}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} + \left\| \frac{\boldsymbol{g}_i(t-1)}{\sum_{l=1}^{n} [A(t-1:0)]_{il}} \right\|_* + \\
& \frac{\sum_{s=1}^{t} \sum_{l=1}^{n} | [A(t-1:0)]_{il} - \varphi_i(t-1) | | \boldsymbol{g}_l(s-1) |}{\sum_{l=1}^{n} [A(t-1:0)]_{il}}
\end{aligned}
$$

根据引理 3 有

$$
\begin{aligned}
\left\| \frac{\boldsymbol{z}_i(t)}{w_i(t)} - \bar{\boldsymbol{z}}(t) \right\|_* \leqslant & \sum_{s=1}^{t} \frac{CL\lambda^{t-1-s}}{\delta} + \sum_{s=1}^{t} \frac{CL\lambda^{t-1}}{\delta} + \frac{L}{n\delta} \leqslant \sum_{s=1}^{t} \frac{CL\lambda^{t-1-s}}{\delta} + \sum_{s=1}^{t} \frac{CL\lambda^{t-1-s}}{\delta} + \frac{L}{n\delta} \leqslant \\
& 2 \sum_{s=1}^{t} \frac{CL\lambda^{t-1-s}}{\delta} + \frac{L}{n\delta} \leqslant \frac{2CL}{\delta\lambda(1-\lambda)} + \frac{L}{n\delta}
\end{aligned}
$$

证毕.

下面将陈述本文的主要结果.

定理 1 假设 1 成立，步长序列 $\{a(t)\}_{t=0}^{\infty}$ 是非负单调递减的. 对任意 $T \geqslant 1$ 和 $\boldsymbol{x}^* \in \boldsymbol{X}^*$，有

$$
F(\hat{\boldsymbol{x}}_j(T)) - F(\boldsymbol{x}^*) \leqslant \frac{L^2}{2T} \sum_{t=1}^{T} a(t-1) + \frac{\psi(\boldsymbol{x}^*)}{Ta(T)} + \frac{3L}{T} \left(\frac{2CL}{\delta\lambda(1-\lambda)} + \frac{L}{n\delta} \right) \sum_{t=1}^{T} a(t-1)
$$

证 由 $F(\boldsymbol{x})$ 的凸性和假设 1(b) 可得

$$
F(\hat{\boldsymbol{x}}_j(T)) - F(\boldsymbol{x}^*) = F(\hat{\boldsymbol{y}}(T)) - F(\boldsymbol{x}^*) + F(\hat{\boldsymbol{x}}_j(T)) - F(\hat{\boldsymbol{y}}(T)) \leqslant
$$

$$F(\hat{\boldsymbol{y}}(T))-F(\boldsymbol{x}^*)+L\|\hat{\boldsymbol{x}}_j(T)-\hat{\boldsymbol{y}}(T)\|\leqslant\frac{1}{T}\sum_{t=1}^{T}F(\boldsymbol{y}(t))-F(\boldsymbol{x}^*)+\frac{L}{T}\sum_{t=1}^{T}\|\boldsymbol{x}_j(t)-\boldsymbol{y}(t)\|\leqslant$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}(f_i(\boldsymbol{y}(t))-f_i(\boldsymbol{x}_i(t)))+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}(f_i(\boldsymbol{x}_i(t))-f_i(\boldsymbol{x}^*))+\frac{L}{T}\sum_{t=1}^{T}\|\boldsymbol{x}_j(t)-\boldsymbol{y}(t)\|\leqslant$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}L\|\boldsymbol{y}(t)-\boldsymbol{x}_i(t)\|+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\langle\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{x}^*\rangle+\frac{L}{T}\sum_{t=1}^{T}\|\boldsymbol{x}_j(t)-\boldsymbol{y}(t)\| \tag{6}$$

考虑式(6)中的第二项，并结合引理 1 可得

$$\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\langle\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{x}^*\rangle=\sum_{t=1}^{T}\langle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t),\boldsymbol{y}(t)-\boldsymbol{x}^*\rangle+\sum_{t=1}^{T}\langle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle=$$

$$\sum_{t=1}^{T}\langle\hat{\boldsymbol{g}}(t),\Pi_X^{\psi}(\bar{\boldsymbol{z}}(t),a(t-1))-\boldsymbol{x}^*\rangle+\sum_{t=1}^{T}\langle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle=$$

$$\sum_{t=1}^{T}\langle\hat{\boldsymbol{g}}(t),\Pi_X^{\psi}\Big(\frac{1}{n}\sum_{r=1}^{t-1}\sum_{j=1}^{n}\boldsymbol{g}_j(r),a(t-1)\Big)-\boldsymbol{x}^*\rangle+\sum_{t=1}^{T}\langle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle=$$

$$\sum_{t=1}^{T}\langle\hat{\boldsymbol{g}}(t),\Pi_X^{\psi}\Big(\sum_{r=1}^{t-1}\hat{\boldsymbol{g}}(r),a(t-1)\Big)-\boldsymbol{x}^*\rangle+\sum_{t=1}^{T}\langle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle=$$

$$\sum_{t=1}^{T}\langle\hat{\boldsymbol{g}}(t),\boldsymbol{x}(t)-\boldsymbol{x}^*\rangle+\sum_{t=1}^{T}\langle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle\leqslant$$

$$\frac{L^2}{2}\sum_{t=1}^{T}a(t-1)+\frac{\psi(\boldsymbol{x}^*)}{a(T)}+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\langle\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle \tag{7}$$

再由 Cauchy 不等式和引理 2 可得

$$\sum_{i=1}^{n}\langle\boldsymbol{g}_i(t),\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\rangle\leqslant\sum_{i=1}^{n}\|\boldsymbol{g}_i(t)\|_*\|\boldsymbol{x}_i(t)-\boldsymbol{y}(t)\|\leqslant$$

$$\sum_{i=1}^{n}L\|\Pi_X^{\psi}(\boldsymbol{y}_i(t),a(t-1))-\Pi_X^{\psi}(\bar{\boldsymbol{z}}(t),a(t-1))\|\leqslant$$

$$\sum_{i=1}^{n}La(t-1)\|\bar{\boldsymbol{z}}(t)-\boldsymbol{y}_i(t)\|_*\leqslant\sum_{i=1}^{n}La(t-1)\left\|\bar{\boldsymbol{z}}(t)-\frac{\boldsymbol{z}_i(t)}{w_i(t)}\right\|_* \tag{8}$$

现在估计式(6)中第三项

$$L\sum_{t=1}^{T}\|\boldsymbol{x}_j(t)-\boldsymbol{y}(t)\|=\sum_{t=1}^{T}L\|\Pi_X^{\psi}(\boldsymbol{y}_j(t),a(t-1))-\Pi_X^{\psi}(\bar{\boldsymbol{z}}(t),a(t-1))\|\leqslant$$

$$\sum_{t=1}^{T}La(t-1)\|\boldsymbol{y}_j(t)-\bar{\boldsymbol{z}}(t)\|_*\leqslant\sum_{t=1}^{T}La(t-1)\left\|\frac{\boldsymbol{z}_j(t)}{w_j(t)}-\bar{\boldsymbol{z}}(t)\right\|_* \tag{9}$$

综合式(7),(8) 和(9)，则有

$$F(\hat{\boldsymbol{x}}_j(T))-F(\boldsymbol{x}^*)\leqslant\frac{L^2}{2T}\sum_{t=1}^{T}a(t-1)+\frac{\psi(x^*)}{Ta(T)}+\frac{2L}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}a(t-1)\left\|\bar{\boldsymbol{z}}(t)-\frac{\boldsymbol{z}_i(t)}{w_i(t)}\right\|_*+$$
$$\frac{L}{T}\sum_{t=1}^{T}a(t-1)\left\|\frac{\boldsymbol{z}_j(t)}{w_j(t)}-\bar{\boldsymbol{z}}(t)\right\|_*$$

最后利用引理 4，即可证得定理 1 的结论. 证毕

定理 2　假设定理 1 的条件成立. 令 $\psi(\boldsymbol{x}^*)\leqslant R^2$，选择步长 $a(t)=\dfrac{p}{\sqrt{t}}(t\geqslant 1)$，这里的 p 是一个正常数. 对任意 $T\geqslant 1$ 和 $\boldsymbol{x}^*\in\boldsymbol{X}^*$，有

$$F(\hat{\boldsymbol{x}}_j(T))-F(\boldsymbol{x}^*)\leqslant\frac{L^2p}{\sqrt{T}}+\frac{1}{\sqrt{T}}\frac{R^2}{p}+\frac{6L^2p}{\sqrt{T}}\Big(\frac{2C}{\delta\lambda(1-\lambda)}+\frac{1}{n\delta}\Big)$$

证　利用定理 1 的结果和 $\psi(\boldsymbol{x}^*)\leqslant R^2$，则有

$$F(\hat{\boldsymbol{x}}_j(T))-F(\boldsymbol{x}^*)\leqslant\frac{L^2}{2T}\sum_{t=1}^{T}a(t-1)+\frac{R^2}{Ta(T)}+\frac{3L}{T}\Big(\frac{2CL}{\delta\lambda(1-\lambda)}+\frac{L}{n\delta}\Big)\sum_{t=1}^{T}a(t-1)$$

注意到 $a(t)=\frac{p}{\sqrt{t}}(t\geqslant 1)$，有 $\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\leqslant 2\sqrt{T}$. 根据以上结果，结论得证.

注 定理 2 的前两项给出了优化误差项，最后一项给出了网络误差项，这里的 λ 刻画了网络的连通性，δ 刻画了网络的平衡性. 同时，定理 2 表明

$$F(\hat{\boldsymbol{x}}_j(T))-F(\boldsymbol{x}^*)=O\left(\frac{1}{\sqrt{T}}\right)$$

它刻画了 $F(\hat{\boldsymbol{x}}_j(T))$ 收敛到最优值 $F(\boldsymbol{x}^*)$ 的速率是 $O\left(\frac{1}{\sqrt{T}}\right)$，其中常数依赖于参数 λ 和 δ.

4 数值实验

考虑一个 l_1 回归问题[6]：给出 n 个点对 $(\boldsymbol{a}_i, b_i)\in\mathbb{R}^d\times\mathbb{R}$，需要估计未知参数 $\boldsymbol{x}\in\mathbb{R}^d$，使得 $\boldsymbol{a}_i^T\boldsymbol{x}\approx b_i$，即求解下述优化问题：

$$\min_{x\in\mathbb{R}^d}F(\boldsymbol{x}):=\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}_i^T\boldsymbol{x}-b_i|,\ s.t.\ \|\boldsymbol{x}\|\leqslant D$$

显然 $F(\boldsymbol{x})$ 是凸的，其次梯度是有界的，上界为 $L=\max_i\|\boldsymbol{a}_i\|$.

对 $i=1,\cdots,n$，假设点对 $(\boldsymbol{a}_i, b_i)$ 是随机产生的，且服从标准正态分布. 取步长 $a(t)=\frac{0.1}{\sqrt{t}}$，$D=10$. 考虑最大误差 $\max\ \mathrm{error}=\max_{i\in V}[F(\hat{\boldsymbol{x}}_i(T))-F(\boldsymbol{x}^*)]$.

图 1(a),(b) 描述了节点个数 $Node=100$，维数分别为 $Dim=2$ 和 $Dim=4$ 的最大误差的收敛曲线. 从图 1(a) 和(b) 可以看出，算法是收敛的. 对比图 1(a) 和(b) 知，随着问题维数增加，其收敛误差也随之增大.

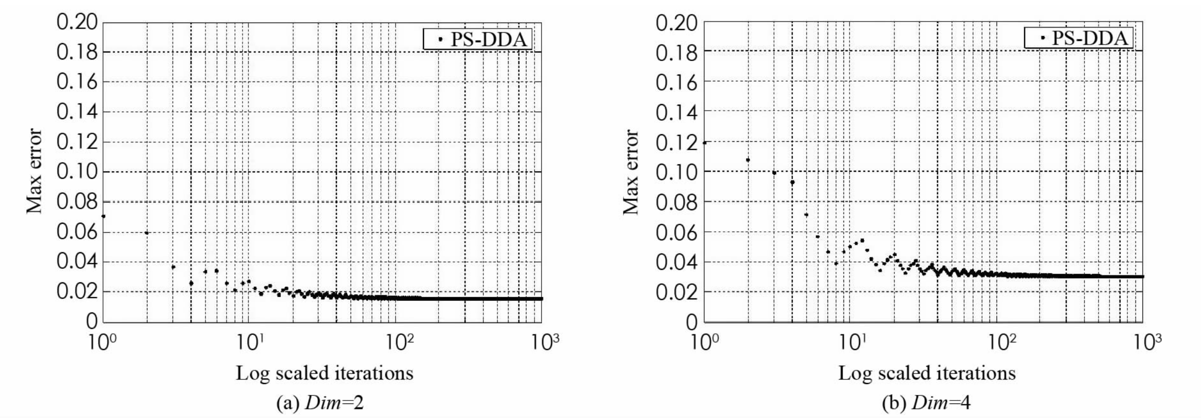

图 1 节点数为 100，维数为 2 和 4 的最大误差收敛曲线

将本文 PS-DDA 算法与文献[11]中的分布式 push-sum 次梯度(PS-SG)算法进行对比. 图 2(a),(b) 分别描述了节点个数 $Node=200$ 和 $Node=400$，维数为 $Dim=2$ 的最大误差值的收敛曲线. 通过对比发现，算法 PS-DDA 的收敛精度均略优于算法 PS-SG.

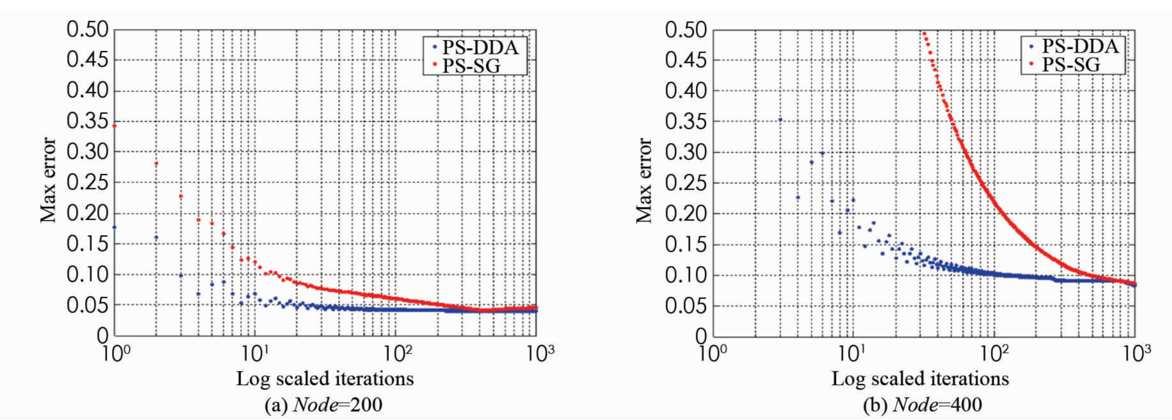

图 2 算法 PS-DDA 与算法 PS-SG 对比的最大误差收敛曲线

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.
[2] 衣　鹏，洪奕光. 分布式合作优化及其应用 [J]. 中国科学：数学，2016，46(10)：1547-1564.
[3] NEDIC A，OZDAGLAR A. Distributed Subgradient Methods for Multi-Agent Optimization [J]. IEEE Transactions on Automatic Control，2009，54(1)：48-61.
[4] DUCHI J C，AGARWAL A，WAINWRIGHT M J. Dual Averaging for Distributed Optimization：Convergence Analysis and Network Scaling [J]. IEEE Transactions on Automatic Control，2012，57(3)：592-606.
[5] ZHU M，MARTINEZ S. On Distributed Convex Optimization under Inequality and Equality Constraints [J]. IEEE Transactions on Automatic Control，2012，57(1)：151-164.
[6] LI J Y，WU C Z，WU Z Y，et al. Gradient-Free Method for Nonsmooth Distributed Optimization [J]. Journal of Global Optimization，2015，61(2)：325-340.
[7] 闫　兵. 基于分布式数据库的图书馆自动管理系统设计 [J]. 西南师范大学学报(自然科学版)，2016，41(2)：147-153.
[8] 张　豪，韩易言，吕庆国，等. 时变网络拓扑图下智能电网中基于优化算法的分布式调度响应 [J]. 西南大学学报(自然科学版)，2018，40(7)：177-180.
[9] 张鼎兴. 一种面向智能电网的无线传感器网络簇路由算法 [J]. 贵州师范大学学报(自然科学版)，2016，34(5)：93-97.
[10] 赵　钢. 基于分布式多引擎架构的网格工作流管理系统 [J]. 西南大学学报(自然科学版)，2012，34(11)：100-107.
[11] 吴其林，方　周，张正金，等. 多信息流协作自组织网络资源优化策略研究综述 [J]. 重庆工商大学学报(自然科学版)，2017，34(3)：39-48.
[12] 黄庆东，闫乔乔，孙　晴. 基于 Fiedler 矢量的分布式自适应分簇算法 [J]. 重庆邮电大学学报(自然科学版)，2017，29(3)：301-306.
[13] TSIANOS K I，LAWLOR S，RABBAT M G. Push-Sum Distributed Dual Averaging for Convex Optimization [C]//51st IEEE Conference on Decision and Control. New York：IEEE Press，2012.
[14] NEDIC A，OLSHEVSKY A. Distributed Optimization over Time-Varying Directed Graphs [J]. IEEE Transactions on Automatic Control，2015，60(3)：601-615.

Distributed Push-Sum Dual Averaging for Convex Optimizationover Time-Varying Directed Graphs

ZHOU Xiao-qing[1]，　LI Jue-you[2]

1. *Chongqing Longshan Middle School，Chongqing 401147，China*；
2. *School of Mathematical Sciences，Chongqing Normal University，Chongqing 401331，China*

Abstract：A distributed convex optimization problem with simple constraints over time-varying directed graphs has been investigated in this paper. By combining the push-sum scheme and distributed dual averaging method，a distributed dual averaging algorithm with push-sum protocolhas first beenproposed. Then，the convergenceof the proposed method has beenestablished，and the explicit convergence rate with order of $O\left(\frac{1}{\sqrt{T}}\right)$ been obtained. Finally，a numerical example on linear regression problem has beenused to demonstrate the effectiveness of the proposed algorithm. Compared with existing works，our algorithm is not required the doublestochasticity of the communication weight matrices. Meanwhile，the proposed algorithm can be used to solve a large class of constrained convex optimization.

Key words：distributed dual averaging；push-sum algorithm；convergence；convex optimization；time-varying network

责任编辑　张　枸