

一种自适应量化的分布式次梯度方法^{*}

陈兰兰, 李觉友, 范 静

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:【目的】考虑了网络通讯受宽带限制,提出了一种自适应通讯量化的分布式次梯度方法来求解分布式凸优化问题。【方法】首先设计了自适应量化方法,其次提出了有向网络下的分布式优化次梯度算法。【结果】得到了当目标函数为凸函数和强凸函数时该算法的收敛率。【结论】数值例子表明所提算法能有效降低通讯成本,并能适用于更一般的非平衡有向通讯网络。

关键词:分布式优化;自适应量化;行随机权重矩阵

中图分类号: O221

文献标志码: A

文章编号: 1672-6693(2021)05-0012-09

近些年来,随着科技飞速发展,特别是无人机领域和分布式数据融合等新兴领域的出现,分布式优化理论及应用得到了越来越多学者的重视^[1]。分布式优化是通过多智能体之间的相互合作来共同完成一个全局的优化目标,通常用来解决许多集中式算法难以胜任的大规模复杂优化问题^[2-5]。但是,随着网络规模和问题维数的不断增大,分布式优化将会产生较高的通信成本^[6-8]。

目前,许多文献研究了基于多智能体系统的分布式优化问题^[1-2,6-7],凸优化问题是其中研究的热点之一^[5-8]。为求解分布式凸优化问题,现有的大多数算法是基于分布式网络的通信权重矩阵满足双随机性^[5]或列随机性^[3]来设计的。比如,文献[5]提出了基于双随机权重矩阵假设下的分布式次梯度优化算法。文献[3]研究了基于列随机权重矩阵假设下的分布式优化次梯度算法。但是,这些基于双随机或列随机权重矩阵的算法不能应用到更一般的通讯网络^[8]。文献[8]指出,假设权重矩阵满足双随机性或列随机性这一条件比较苛刻,对于大多数传播基础的通讯网络通常并不满足这个条件。另外,目前基于列随机权重矩阵设计的分布式算法不能用来求解带约束的分布式优化问题。文献[8]设计了一种只需要权重矩阵满足行随机性条件下的分布式次梯度投影方法,其中邻近智能体之间的通信不一定是双向的,并给出了算法的收敛率。然而,这些算法并没有考虑网络通讯受到宽带限制。随着问题规模增大,前面所提到的大多数算法都将会受到通讯宽带的限制,从而增加了通讯成本^[6-8]。

针对分布式优化中经常会遇到通讯成本高的问题,受文献[5,8]的启发,本文提出了一种自适应通讯量化的分布式次梯度优化方法,并分析了算法的收敛性,得到了算法的收敛率。特别地,当目标函数为凸函数时,算法的收敛率为 $\mathcal{O}\left(\frac{\ln(T)}{\sqrt{T}}\right)$;当目标函数为强凸函数时,算法的收敛率为 $\mathcal{O}\left(\frac{\ln(T)}{T}\right)$;其中 T 是总迭代次数。本文的算法收敛率与没有量化的分布式次梯度算法^[8]同阶,但是通过数值仿真对比发现,本文的算法极大地有效节省了通讯成本。另外,本文设计的算法是基于通讯矩阵满足行随机性,这使得本文的算法适用于更一般的非平衡有向通讯网络。

1 问题与算法

1.1 准备工作

如无特殊说明,本文所提到的向量均为列向量。向量 $\mathbf{x}=(x_1, x_2, \dots, x_d)$, $\mathbf{x} \in \mathbf{R}^d$ 表示 d 维列向量; $\mathbf{x}^T$ 表示向量 $\mathbf{x}$ 的转置; $\langle \mathbf{x}, \mathbf{y} \rangle$ 表示向量 $\mathbf{x}$ 和 $\mathbf{y}$ 的内积; $\|\cdot\|$ 表示Euclid范数;如果一个方阵 $\mathbf{A} \in \mathbf{R}^{n \times n}$ 的每一行各个元

* 收稿日期:2021-04-14 修回日期:2021-05-18 网络出版时间:2021-09-13 10:51

资助项目:国家自然科学基金(No. 11971083);重庆市自然科学基金(No. cstc2020jcyj-msxmX0287)。

第一作者简介:陈兰兰,女,研究方向为分布式优化理论与算法,E-mail:1139149863@qq.com;通信作者:李觉友,男,教授,博士,研究方向是最优化理论与算法,E-mail:lijueyou@cqnu.edu.cn

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20210912.1950.002.html

素之和都是1,则称 $\mathbf{A}$ 是行随机矩阵,这里用 $[\mathbf{A}]_{ij}$ 来表示矩阵 $\mathbf{A}$ 的第 i 行第 j 列的元素; $\mathbf{1}=[1, \dots, 1]^T$; $E[\cdot]$ 表示数学期望。对 $\forall \mathbf{x}, \mathbf{y} \in \mathbf{R}^d$ 有 $f(\mathbf{y}) - f(\mathbf{x}) \geq \langle \nabla f(\mathbf{x}), (\mathbf{y} - \mathbf{x}) \rangle + \frac{u}{2} \|\mathbf{y} - \mathbf{x}\|$, 称函数 $f(\mathbf{x})$ 在 $\mathbf{R}^d$ 上是 u -强凸的。 $P_\chi[\mathbf{x}]$ 表示点 $\mathbf{x}$ 在凸集 $\chi \subset \mathbf{R}^d$ 上的Euclid投影。 $\partial f(\mathbf{x}) = \{\mathbf{g} \in \mathbf{R}^d \mid f(\mathbf{y}) - f(\mathbf{x}) \geq \mathbf{g}^T(\mathbf{y} - \mathbf{x}), \forall \mathbf{y} \in \mathbf{R}^d\}$ 表示凸函数 $f: \mathbf{R}^d \rightarrow \mathbf{R}$ 在点 $\mathbf{x}$ 的次梯度。

设 $\mathcal{G}=(V, \epsilon)$ 表示一个有向图,其中 $V=\{1, 2, \dots, n\}$ 表示由 n 个节点(智能体)所构成的集合, $\epsilon \subseteq V \times V$ 表示相邻节点所构成的有向边集,边 $(j, i) \in \epsilon$ 表示节点 i 可以从节点 j 接收数据信息。用 $N_i = \{j \mid (j, i) \in \epsilon\} \cup \{i\}$ 来表示节点 i 的邻居集。

1.2 问题

考虑如下的多智能体分布式优化问题:

$$\min_{\mathbf{x} \in \chi} \left\{ f(\mathbf{x}) \triangleq \sum_{i=1}^n f_i(\mathbf{x}) \right\}. \quad (1)$$

其中,全局目标函数 f 为各个智能体的目标函数 $f_i: \mathbf{R}^d \rightarrow \mathbf{R}$ 之和, $\chi \subset \mathbf{R}^d$ 是约束集。特别地,每个智能体 $i \in V$ 仅知道它自己所拥有的目标函数 f_i ,邻近智能体之间允许通讯交流局部信息。

假设 1 a) 假设 $\chi \subset \mathbf{R}^d$ 是一个箱子约束集;b) 对 $\forall i \in V$,每个局部目标函数 f_i 在 χ 上是凸的,且其次梯度是一致有界的,即存在常数 $L > 0$ 使得 $\|\mathbf{g}_i\| \leq L, \forall \mathbf{g}_i \in \partial f_i(\mathbf{x}), \forall \mathbf{x} \in \chi$ 。

假设 2 a) 有向图 $\mathcal{G}$ 是强连通的;b) 有向图 $\mathcal{G}$ 的权重矩阵 $\mathbf{A}=[a_{ij}]$ 是行随机的,即满足:如果 $(j, i) \in \epsilon$,则 $a_{ij} > 0$;否则, $a_{ij} = 0$ 。而且还满足: $\sum_{j=1}^n a_{ij} = 1, \forall j \in V$ 。

1.3 自适应量化

为了叙述简单,先考虑一维情形,对高维的情形类似。给定一个区间 $[l, u]$,插入 $B-2$ 个分点 $l = \tau_1 < \dots < \tau_{B-1} = u$ 将这个区间等分为 $B-1$ 个小区间,每个小区间的长度为 $\Delta = \tau_{m+1} - \tau_m = \frac{u-l}{B-1}, m=1, \dots, B-1$,这里 $B > 2$ 。任给实数 $x \in [l, u]$,用 $q = Q(x)$ 表示点 x 在区间 $[l, u]$ 上的量化值。本文定义如下的量化函数:

$$Q(x) \triangleq \arg \min_m |\tau_m - x|, m=1, \dots, B-1. \quad (2)$$

通过(2)式可知,最大量化误差为:

$$|x - q| \leq \Delta = \frac{u-l}{B-1}. \quad (3)$$

本文主要研究智能体间量化通信对算法性能的影响,设计出了一种自适应量化方法,主要思想是逐步缩减量化区间,降低量化误差。现假设量化的区间是时变的,即 $[l(t), u(t)]$ 。特别地,在每一个时刻 $t \geq 0$,每个智能体只允许向邻居发送其量化值和接收其邻居的量化值。用 $\Delta_i(t) \triangleq x_i(t) - q_i(t)$ 表示智能体 i 在时刻 t 的量化误差,这里 $q_i(t) = Q_i(x_i(t))$ 表示点 $x_i(t)$ 在时刻 t 的通过量化函数(2)式量化后得到的值。对于每一个智能体 $i \in V$,考虑如下更新:

$$x_i(t+1) = P_\chi \left(x_i(t) - q_i(t) + \sum_{j \in N_i} a_{ij} q_j(t) - \alpha(t) g_i(t) \right). \quad (4)$$

其中 $\alpha(t)$ 是步长, $g_i(t)$ 是局部目标函数 f_i 在点 $x_i(t)$ 的次梯度。注意量化器 Q_i 依赖于智能体 i 的时变量化区间 $[l_i(t), u_i(t)]$ 和待量化的点 $x_i(t)$,称这种量化方式为自适应量化。

下面给出(4)式自适应量化更新的解释说明。首先,在时刻 $t \geq 0$,每个智能体 i 在量化区间 $[l_i(t), u_i(t)]$ 用量化器 Q_i 量化它的估计值 $x_i(t)$,然后将这个量化值 $q_i(t)$ 发送给它的邻居 $j \in N_i$ 。同时,智能体 i 也会收到邻居 j 发送来的量化值 $q_j(t)$ 。通过和邻居量化信息凸组合并利用自身信息做次梯度下降更新得到 $x_i(t+1)$ 。此外,智能体 i 还将它的量化误差 $x_i(t) - q_i(t)$ 引入到自身的更新中,目的是为了消除因量化误差而产生的网络上的偏差。

令 $\varphi_i(t) = P_\chi(x_i(t) - q_i(t) + \sum_{j=1}^n a_{ij} q_j(t))$,由Euclid投影性质^[7]和假设1,有:

$$|x_i(t+1) - \varphi_i(t)| \leq |\alpha(t) g_i(t)| \leq L\alpha(t),$$

则 $x_i(t+1) \in [\varphi_i(t) - L\alpha(t), \varphi_i(t) + L\alpha(t)]$ 。现在把这个时变量化区间的左右端点分别记为:

$$l_i(t+1)=\varphi_i(t)-L\alpha(t) \text{ 和 } u_i(t+1)=\varphi_i(t)+L\alpha(t),$$

然后在区间 $[l_i(t+1), u_i(t+1)]$ 中插入 $B-2$ 个点, $l_i(t)=\tau_i^1(t)<\tau_i^2(t)<\dots<\tau_i^B(t)=u_i(t)$, 将这个区间等分为 $B-1$ 个小区间, 则每个小区间的长度为:

$$|\tau_i^m(t)-\tau_i^{m+1}(t)|=\frac{2L\alpha(t-1)}{B-1}, l=1, \dots, B-1, i \in V. \quad (5)$$

因为 $x_i(t) \in [l_i(t), u_i(t)]$, 由(2)式可知, $q_i(t) \in [l_i(t), u_i(t)]$, 由此得到量化误差上界为:

$$|x_i(t)-q_i(t)| \leq \frac{2L\alpha(t-1)}{B-1}, t \geq 0, i \in V. \quad (6)$$

与(3)式比较, (6)式的量化误差界依赖时间 t , 并且随着步长 $\alpha(t)$ 减小, 量化误差界也随之缩减。

下面考虑高维的情形, 假设维数 $d>1$ 。基于前面的讨论, 设计了如下的基于自适应通讯量化的分布式次梯度算法来求解约束问题(1)。

算法 1

步骤 1: 初始化 $\mathbf{x}_i(0)=\mathbf{q}_i(0) \in \chi, \mathbf{z}_i(0)=(0, \dots, 1_i, \dots, 0) \in \mathbf{R}^d$, 选择合适的步长 $\{\alpha(t)\}$ 。

步骤 2: 对 $t \geq 0, i=1, \dots, n$, 更新:

$$\mathbf{x}_i(t+1)=P_\chi\left(\mathbf{x}_i(t)-\mathbf{q}_i(t)+\sum_{j \in N_i} a_{ij} \mathbf{q}_j(t)-\alpha(t) \frac{\tilde{\mathbf{g}}_i(t)}{z_{ii}(t)}\right), \quad (7)$$

$$\mathbf{z}_i(t+1)=\sum_{j \in N_i} a_{ij} \mathbf{z}_j(t). \quad (8)$$

算法 1 考虑了当计算次梯度有噪音的情况, 即 $\tilde{\mathbf{g}}_i(t)=\mathbf{g}_i(t)+\boldsymbol{\varepsilon}_i(t)$, 其中 $\mathbf{g}_i(t) \in \partial f_i\left(\sum_{j=1}^n a_{ij} \mathbf{x}_j(t)\right)$, $\boldsymbol{\varepsilon}_i(t)$ 是随机误差。同时, 还考虑了更一般的通讯网络, 即满足假设 1, 仅要求通讯权重矩阵是行随机的。在(7)式中, $\tilde{\mathbf{g}}_i(t)$ 除以辅助参数 $z_{ii}(t)$, 目的是为了处理网络不平衡性。如果令 $\mathbf{Z}(t)=[z_1(t), \dots, z_n(t)]^T$, 则由(8)式可得 $\mathbf{Z}(t+1)=\mathbf{A}\mathbf{Z}(t)=\dots=\mathbf{A}^t$, 其中 $\mathbf{Z}(0)=\mathbf{I}$, 进而 $z_{ii}(t+1)=[\mathbf{A}^t]_{ii}$, 这意味着不需要量化辅助参数 $z_{ii}(t)$ 。

注 1 算法 1 推广了文献[5]中的算法到更一般的通讯网络, 同时考虑了带有噪音的次梯度。另外, 算法 1 将文献[8]的算法推广到通讯量化的情形下。从数值仿真可以看出, 与文献[8]中提出的没有考虑通讯量化的分布式优化算法相比, 本文提出的算法 1 能极大地节省通讯成本。

令 $\mathcal{F}_t$ 表示算法 1 迭代到 t 时刻产生的所有历史信息。记权重矩阵 $\mathbf{A}$ 的左 Perron 特征向量为 $\boldsymbol{\pi}=[\pi_1, \dots, \pi_n]^T$, 则由文献[8]有 $\mathbf{1}^T \boldsymbol{\pi}=1$ 。

引理 1^[8] 如果假设 2 成立且 $z_{ii}(t)$ 由算法 1 产生, 则对 $\forall i, j \in V$ 和 $\forall t \geq 0$, 存在常数 $C>0, \lambda \in (0, 1)$ 使得不等式 $|\mathbf{A}^t]_{ij}-\pi_i| \leq C\lambda^t$ 和 $|z_{ii}(t)-\pi_i| \leq C\lambda^t$ 成立, 且存在一个常数 $\eta>0$, 使得 $\eta^{-1} \leq z_{ii}(t) \leq 1$ 成立。

假设 3 随机变量 $\boldsymbol{\varepsilon}_i(t)$ 满足均值为零和方差有界, 即对 $\forall i \in V, t \geq 0$, 存在一个实数 $\sigma>0$ 有:

$$E[\boldsymbol{\varepsilon}_i(t) | \mathcal{F}_{t-1}]=0, E[\|\boldsymbol{\varepsilon}_i(t)\|^2 | \mathcal{F}_{t-1}] \leq \sigma^2.$$

由 Jensen 不等式有 $E[\|\boldsymbol{\varepsilon}_i(t)\| | \mathcal{F}_{t-1}] \leq \sqrt{E[\|\boldsymbol{\varepsilon}_i(t)\|^2 | \mathcal{F}_{t-1}]} \leq \sigma$ 。再由三角不等式, 进一步得到:

$$E[\|\mathbf{g}_i(t)+\boldsymbol{\varepsilon}_i(t)\|^2 | \mathcal{F}_{t-1}] \leq G^2, E[\|\mathbf{g}_i(t)+\boldsymbol{\varepsilon}_i(t)\| | \mathcal{F}_{t-1}] \leq G.$$

这里 $G=L+\sigma$ 。另外, 还有 $E[\tilde{\mathbf{g}}_i(t) | \mathcal{F}_{t-1}]=\mathbf{g}_i(t)$ 。

2 主要结果

为后面算法的收敛性做准备, 首先得到以下关于网络化误差的一个引理。

引理 2 若假设 1~3 成立, $\{\mathbf{x}_i(t)\}$ 是由算法 1 迭代产生的序列, $\{\alpha(t)\}$ 是一个非负不增的步长序列。令 $\bar{\mathbf{x}}(t) \triangleq \sum_{i=1}^n \pi_i \mathbf{x}_i(t)$, 对 $t \geq 0, \forall i \in V$ 有: $E[\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\| | \mathcal{F}_{t-1}] \leq M_1 \lambda^t + M_2 \alpha(t-1)$,

其中 $M_1=nC\delta, M_2=\frac{nC(4\sqrt{d}+B)\eta G}{(B-1)(1-\lambda)}$ 。

证明 根据算法 1 有:

$$\|\boldsymbol{\Delta}_i(t)\|=\|\mathbf{x}_i(t)-\mathbf{q}_i(t)\| \leq \frac{2\sqrt{d}\tilde{\mathbf{g}}_i(t)\alpha(t-1)}{(B-1)z_{ii}(t)}. \quad (9)$$

因 $\mathbf{x}_i(t) \in \chi$, 则有 $\sum_{j=1}^n a_{ij} \mathbf{x}_j(t) \in \chi$ 。令 $\mathbf{w}_i(t) = \mathbf{x}_i(t+1) - \sum_{j=1}^n a_{ij} \mathbf{x}_j(t)$, 由 Euclid 投影性质^[7]可知:

$$\|\mathbf{w}_i(t)\| \leq \left\| \boldsymbol{\Delta}_i(t) - \sum_{j=1}^n a_{ij} \boldsymbol{\Delta}_j(t) - \alpha(t) \frac{\tilde{\mathbf{g}}_i(t)}{z_{ii}(t)} \right\| \leq \|\boldsymbol{\Delta}_i(t)\| + \sum_{j=1}^n a_{ij} \|\boldsymbol{\Delta}_j(t)\| + \frac{\alpha(t)}{z_{ii}(t)} \|\tilde{\mathbf{g}}_i(t)\|. \quad (10)$$

另外, 对(9)式两边同时取条件期望有:

$$E[\|\boldsymbol{\Delta}_i(t)\| \mid \mathcal{F}_{t-1}] \leq \frac{2\sqrt{d}\eta G\alpha(t-1)}{B-1}. \quad (11)$$

在(10)式两边同时取条件期望并联立(9)式, 可得:

$$E[\|\mathbf{w}_i(t)\| \mid \mathcal{F}_{t-1}] \leq \frac{4\eta\sqrt{d}G\alpha(t-1)}{B-1} + \eta G\alpha(t) \leq \frac{(4\sqrt{d}+B)\eta G\alpha(t-1)}{B-1}. \quad (12)$$

利用 $\mathbf{w}_i(t) = \mathbf{x}_i(t+1) - \sum_{j=1}^n a_{ij} \mathbf{x}_j(t)$, 可得:

$$\mathbf{x}_i(t) = \sum_{j=1}^n a_{ij} \mathbf{x}_j(t-1) + \mathbf{w}_i(t-1) = \sum_{j=1}^n [\mathbf{A}^t]_{ij} \mathbf{x}_j(0) + \sum_{s=0}^{t-1} \sum_{j=1}^n [\mathbf{A}^{t-1-s}]_{ij} \mathbf{w}_j(s). \quad (13)$$

在(13)式两边先同乘 π_i 然后求和, 并注意到 $\boldsymbol{\pi}^T \mathbf{A} = \boldsymbol{\pi}^T$, 可得:

$$\bar{\mathbf{x}}(t) = \sum_{j=1}^n \pi_j \mathbf{x}_j(0) + \sum_{s=0}^{t-1} \sum_{j=1}^n \pi_j \mathbf{w}_j(s). \quad (14)$$

将(13), (14)式相减后取范数, 可得:

$$\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| \leq \sum_{j=1}^n |[\mathbf{A}^t]_{ij} - \pi_j| \|\mathbf{x}_j(0)\| + \sum_{s=0}^{t-1} \sum_{j=1}^n |[\mathbf{A}^{t-1-s}]_{ij} - \pi_j| \|\mathbf{w}_j(s)\|. \quad (15)$$

在(15)式两边同时取条件期望, 并根据(12)式, 可得:

$$E[\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| \mid \mathcal{F}_{t-1}] \leq nC\delta\lambda^t + \frac{nC(4\sqrt{d}+B)\eta G}{B-1} \sum_{s=0}^{t-1} \lambda^{t-1-s} \alpha(s-1).$$

由文献[8]可得 $\sum_{s=0}^{t-1} \lambda^{t-s} \alpha(s-1) \leq \frac{\alpha(t-1)}{1-\lambda}$ 。令 $\delta = \max_{j \in V} \{\|\mathbf{x}_j(0)\|\}$, 由上式可证得结论。

引理 3 若假设 1~3 成立, $\{\mathbf{x}_i(t)\}$ 是由算法 1 产生的序列。对 $t \geq 0$ 和 $\forall i \in V$, 则有:

$$E\left[\sum_{i=1}^n \pi_i \|\mathbf{x}_i(t+1) - \mathbf{x}^*\|^2 \mid \mathcal{F}_t\right] \leq E\left[\sum_{i=1}^n \pi_i \|\mathbf{x}_i(t) - \mathbf{x}^*\|^2\right] - 2\alpha(t)E[f(\mathbf{x}_i(t)) - f(\mathbf{x}^*)] + (2L(n+\eta)M_1 + 2n\eta CL\rho)\lambda^t \alpha(t) + (2L(n+\eta)M_2 + 8M_3 + \eta^2 G^2)\alpha^2(t-1),$$

其中 $\rho = \sup_{\mathbf{u}, \mathbf{v} \in \chi} \|\mathbf{u} - \mathbf{v}\|$, $M_3 = \left(\frac{2\sqrt{d}\eta G}{B-1}\right)^2$ 。

证明 由 $\boldsymbol{\Delta}_i(t) = \mathbf{x}_i(t) - \mathbf{q}_i(t)$, 有 $\mathbf{x}_i(t+1) = P_\chi\left(\sum_{j=1}^n a_{ij} \mathbf{x}_j(t) + \boldsymbol{\Delta}_i(t) - \sum_{j=1}^n a_{ij} \boldsymbol{\Delta}_j(t) - \alpha(t) \frac{\tilde{\mathbf{g}}_i(t)}{z_{ii}(t)}\right)$ 。

令 $\mathbf{y}_i(t) = \sum_{j=1}^n a_{ij} \mathbf{x}_j(t)$, 根据 Euclid 投影性质^[7], 有:

$$\begin{aligned} \|\mathbf{x}_i(t+1) - \mathbf{x}^*\|^2 &\leq \left\| \mathbf{y}_i(t) - \mathbf{x}^* + \boldsymbol{\Delta}_i(t) - \sum_{j=1}^n a_{ij} \boldsymbol{\Delta}_j(t) - \alpha(t) \frac{\tilde{\mathbf{g}}_i(t)}{z_{ii}(t)} \right\|^2 = \\ &\quad \|\mathbf{y}_i(t) - \mathbf{x}^*\|^2 - 2\langle \alpha(t) \frac{\tilde{\mathbf{g}}_i(t)}{z_{ii}(t)}, \mathbf{y}_i(t) - \mathbf{x}^* \rangle + \\ &\quad 2\langle \boldsymbol{\Delta}_i(t) - \sum_{j=1}^n a_{ij} \boldsymbol{\Delta}_j(t), \mathbf{y}_i(t) - \mathbf{x}^* \rangle + \left\| \boldsymbol{\Delta}_i(t) - \sum_{j=1}^n a_{ij} \boldsymbol{\Delta}_j(t) - \alpha(t) \frac{\tilde{\mathbf{g}}_i(t)}{z_{ii}(t)} \right\|^2. \end{aligned} \quad (16)$$

由 $\boldsymbol{\pi}^T \mathbf{A} = \boldsymbol{\pi}^T$, 有:

$$\sum_{i=1}^n \pi_i \langle \boldsymbol{\Delta}_i(t) - \sum_{j=1}^n a_{ij} \boldsymbol{\Delta}_j(t), \mathbf{y}_i(t) - \mathbf{x}^* \rangle = 0 \quad (17)$$

以及:

$$\sum_{i=1}^n \pi_i \|\mathbf{y}_i(t) - \mathbf{x}^*\|^2 = \sum_{i=1}^n \pi_i \left\| \sum_{j=1}^n a_{ij} \mathbf{x}_j(t) - \mathbf{x}^* \right\|^2 \leq \sum_{i=1}^n \pi_i \sum_{j=1}^n a_{ij} \|\mathbf{x}_j(t) - \mathbf{x}^*\|^2 = \sum_{i=1}^n \pi_i \|\mathbf{x}_i(t) - \mathbf{x}^*\|^2. \quad (18)$$

对(16)式两边同时乘 π_i 并求和,联立(17)和(18)式,可得:

$$\begin{aligned} \sum_{i=1}^n \pi_i \|x_i(t+1) - x^*\|^2 &\leq \sum_{i=1}^n \pi_i \|x_i(t) - x^*\|^2 - 2 \sum_{i=1}^n \frac{\pi_i \alpha(t)}{z_{ii}(t)} \langle \tilde{g}_i(t), y_i(t) - x^* \rangle + \\ &\quad \sum_{i=1}^n \pi_i \left\| \Delta_i(t) - \sum_{j=1}^n a_{ij} \Delta_j(t) - \alpha(t) \frac{\tilde{g}_i(t)}{z_{ii}(t)} \right\|^2. \end{aligned} \quad (19)$$

首先估计(19)式右边第 2 项,可得:

$$\begin{aligned} \sum_{i=1}^n \frac{-2\alpha(t)\pi_i}{z_{ii}(t)} \langle \tilde{g}_i(t), y_i(t) - x^* \rangle &\leq \sum_{i=1}^n \frac{-2\alpha(t)\pi_i}{z_{ii}(t)} (f_i(y_i(t)) - f_i(x^*)) \leq \\ \sum_{i=1}^n \frac{2\alpha(t)\pi_i}{z_{ii}(t)} (L \sum_{j=1}^n a_{ij} \|x_j(t) - \bar{x}(t)\|) - \sum_{i=1}^n \frac{2\alpha(t)\pi_i}{z_{ii}(t)} (f_i(\bar{x}(t)) - f_i(x^*)) &\leq \\ 2L\eta\alpha(t) \sum_{i=1}^n \pi_i \|x_i(t) - \bar{x}(t)\| - \sum_{i=1}^n \frac{2\alpha(t)\pi_i}{z_{ii}(t)} (f_i(\bar{x}(t)) - f_i(x^*)). \end{aligned} \quad (20)$$

上式中的第 1 个不等式是由 f_i 的凸性可得,第 2 个不等式是由假设 1 中的 b)可得,第 3 个不等式是由引理 1 和 $\pi^T A = \pi^T$ 可得。由 $f(x) = \sum_{i=1}^n f_i(x)$, 有:

$$\begin{aligned} -2 \sum_{i=1}^n \pi_i \frac{\alpha(t)}{z_{ii}(t)} (f_i(\bar{x}(t)) - f_i(x^*)) &\leq \\ -2\alpha(t) [f(\bar{x}(t)) - f(x^*)] + 2\alpha(t) \sum_{i=1}^n \left| \frac{z_{ii}(t) - \pi_i}{z_{ii}(t)} \right| |f_i(\bar{x}(t)) - f_i(x^*)| &\leq \\ -2\alpha(t) [f(\bar{x}(t)) - f(x^*)] + 2n\alpha(t) \eta C \lambda' L \rho, \end{aligned} \quad (21)$$

这里 $\rho = \sup_{u,v \in \mathcal{X}} |u - v|$ 。由 f 的凸性有:

$$f(x_i(t)) - f(x^*) \leq nL [\|x_i(t) - \bar{x}(t)\|] + [f(\bar{x}(t)) - f(x^*)], \quad (22)$$

由假设 3 可知, $E[\tilde{g}_i(t) | \mathcal{F}_{t-1}] = \nabla f_i(x_i(t))$, 联立(20), (21)和(22)式,并取条件期望,可得:

$$\begin{aligned} -2 \sum_{i=1}^n \frac{\pi_i \alpha(t)}{z_{ii}(t)} E[\langle \tilde{g}_i(t), x_i(t) - x^* \rangle | \mathcal{F}_{t-1}] &\leq \\ 2L(n + \eta)\alpha(t)(M_1 \lambda' + M_2 \alpha(t-1)) - 2\alpha(t) E[f(x(t)) - f(x^*)] + 2n\alpha(t) \eta C \lambda' L \rho. \end{aligned} \quad (23)$$

下面估计(19)式右边第 3 项,由 $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$, 有:

$$\sum_{i=1}^n \pi_i \left\| \Delta_i(t) - \sum_{j=1}^n a_{ij} \Delta_j(t) - \alpha(t) \frac{\tilde{g}_i(t)}{z_{ii}(t)} \right\|^2 \leq 2 \sum_{i=1}^n \pi_i \left\| \Delta_i(t) - \sum_{j=1}^n a_{ij} \Delta_j(t) \right\|^2 + 2 \frac{\alpha^2(t)}{z_{ii}^2(t)} \sum_{i=1}^n \pi_i \|\tilde{g}_i(t)\|^2, \quad (24)$$

其中:

$$\left\| \Delta_i(t) - \sum_{j=1}^n a_{ij} \Delta_j(t) \right\|^2 \leq \sum_{j=1}^n a_{ij} \|\Delta_i(t) - \Delta_j(t)\|^2 \leq 4 \sum_{j=1}^n a_{ij} \|\Delta_i(t)\|^2. \quad (25)$$

在(9)式两边平方后取条件期望,并由假设 3 和引理 1, 有:

$$E[\|\Delta_i(t)\|^2 | \mathcal{F}_{t-1}] = E[\|x_i(t) - q_i(t)\|^2 | \mathcal{F}_{t-1}] \leq \left(\frac{2\sqrt{d}\eta G}{B-1} \right)^2 \alpha^2(t-1) = M_3 \alpha^2(t-1). \quad (26)$$

将(25)式代入(24)式并取条件期望,再结合(26)和 $\pi^T A = \pi^T$, 有:

$$E \left[\sum_{i=1}^n \pi_i \left\| \Delta_i(t) - \sum_{j=1}^n a_{ij} \Delta_j(t) - \alpha(t) \frac{\tilde{g}_i(t)}{z_{ii}(t)} \right\|^2 \middle| \mathcal{F}_{t-1} \right] \leq 8M_3 \alpha^2(t-1) + \eta^2 G^2 \alpha^2(t). \quad (27)$$

最后,结合(19), (23)和(27)式,结论成立。

下面给出了本文的第 1 个主要结果,即当函数 f_i 是凸的时,算法 1 的收敛速率。

定理 1 若假设 1~3 成立, $\{x_i(t)\}$ 是由算法 1 产生的序列。对 $t \geq 0$, 取步长 $\alpha(t) = \frac{1}{\sqrt{t+1}}$ 和令 $z_i(T) =$

$$\frac{\sum_{t=0}^T \alpha(t) x_i(t)}{\sum_{t=0}^T \alpha(t)}, \forall i \in V. \text{ 则有: } E[f(z_i(T))] - f(x^*) \leq C_1 \frac{1}{\sqrt{T+1}} + C_2 \frac{(\ln(T+1)+1)}{\sqrt{T+1}}, \text{ 其中: } C_1 =$$

$$\left(\frac{\rho}{2} + LM_1(n+\eta) + n\eta CL\rho\right) \frac{1}{1-\lambda}, C_2 = LM_2(n+\eta) + 4M_3 + \frac{1}{2}\eta^2 G^2.$$

证明 对引理3两边同时从 $t=0$ 到 T 求和,可得:

$$2 \sum_{t=0}^T \alpha(t) E[f(\mathbf{x}_i(t)) - f(\mathbf{x}^*)] \leq E \left[\sum_{i=1}^n \pi_i \|\mathbf{x}_i(0) - \mathbf{x}^*\|^2 \right] + (2L(n+\eta)M_1 + 2n\eta CL\rho\alpha) \sum_{t=0}^T \lambda^t \alpha(t) + (2L(n+\eta)M_2 + 8M_3 + \eta^2 G^2) \sum_{t=0}^T \alpha^2(t-1). \quad (28)$$

由 $\alpha(t) = \frac{1}{\sqrt{t+1}}$, 有:

$$\sum_{t=0}^T \alpha(t) \lambda^t \leq \sum_{t=0}^T \alpha(0) \lambda^t \leq \frac{1}{1-\lambda}, \sum_{t=0}^T \alpha^2(t-1) = \sum_{t=0}^{T-1} \frac{1}{t+1} \leq 1 + \int_0^T \frac{1}{t+1} dt \leq \ln(T+1) + 1. \quad (29)$$

联立(28)和(29)式,可得:

$$2 \sum_{t=0}^T \alpha(t) E[f(\mathbf{x}_i(t)) - f(\mathbf{x}^*)] \leq (\rho + 2L(n+\eta)M_2 + 2n\eta CL\rho\alpha) \frac{1}{1-\lambda} + (2L(n+\eta)M_2 + 8M_3 + \eta^2 G^2)(\ln(T+1) + 1). \quad (30)$$

在(30)式两边同时除以 $2 \sum_{t=0}^T \alpha(t)$, 由 f 的凸性和 Jensen 不等式,并注意到: $\sum_{t=0}^T \alpha(t) \geq \sqrt{T+1}$, 则(30)式可得:

$$E[f(\mathbf{z}_i(T))] - f(\mathbf{x}^*) \leq \left(\frac{\rho}{2} + L(n+\eta)M_1 + n\eta CL\rho\right) \frac{1}{(1-\lambda)\sqrt{T+1}} + \left(L(n+\eta)M_2 + 4M_3 + \frac{1}{2}\eta^2 G^2\right) \frac{\ln(T+1) + 1}{\sqrt{T+1}}.$$

令 $C_1 = \frac{\frac{\rho}{2} + LM_1(n+\eta) + n\eta CL\rho}{1-\lambda}, C_2 = LM_2(n+\eta) + 4M_3 + \frac{1}{2}\eta^2 G^2$, 结论成立。

下面给出本文的第2个主要结果,即当函数 f_i 是 u_i -强凸函数时,算法1的收敛速率。

定理2 若假设1~3成立,并假设函数 f_i 是 u_i -强凸函数, $\{x_i(t)\}$ 是由算法1产生的序列。令 $\alpha(t) = \frac{a}{t+1}$,

$\alpha(0) = a \geq \frac{1}{u}$, 其中 $u = \min_{i \in V} u_i$ 。对 $i \in V$ 且 $T \geq 1$, 有: $\|E[\hat{\mathbf{z}}_i(T)] - \mathbf{x}^*\| \leq C_3 \frac{1}{T+1} + C_4 (\ln(T+1) + 1) \frac{1}{T+1}$,

其中: $C_3 = \frac{a(2L(n+3\eta)M_1 + 2n\eta CL\rho)}{1-\lambda}, C_4 = 2a(2L(n+3\eta)M_2 + 8M_3 + \eta^2 G^2), \hat{\mathbf{z}}_i(T) = \frac{\sum_{t=0}^T \mathbf{x}_i(t)}{T+1}, \forall i \in V$ 。

证明 因为:

$$\sum_{i=1}^n \frac{-2\alpha(t)\pi_i}{z_{ii}(t)} \langle \mathbf{g}_i(t), \mathbf{y}_i(t) - \mathbf{x}^* \rangle = - \sum_{i=1}^n \frac{2\alpha(t)\pi_i}{z_{ii}(t)} \langle \mathbf{g}_i(t), \mathbf{x}_i(t) - \mathbf{x}^* \rangle - \sum_{i=1}^n \frac{2\alpha(t)\pi_i}{z_{ii}(t)} \langle \mathbf{x}_i(t), \mathbf{y}_i(t) - \mathbf{x}_i(t) \rangle. \quad (31)$$

首先估计(31)式右边第1项。由(20)式和 f 是 u -强凸的,有:

$$\sum_{i=1}^n \frac{-2\alpha(t)\pi_i}{z_{ii}(t)} \langle \mathbf{g}_i(t), \mathbf{x}_i(t) - \mathbf{x}^* \rangle \leq \sum_{i=1}^n \frac{-2\alpha(t)\pi_i}{z_{ii}(t)} \left(f_i(\mathbf{x}_i(t)) - f_i(\mathbf{x}^*) + \frac{u_i}{2} \|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \right) \leq 2\eta L\alpha(t) \sum_{i=1}^n \pi_i \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| - \sum_{i=1}^n \frac{2\alpha(t)\pi_i}{z_{ii}(t)} (f_i(\bar{\mathbf{x}}(t)) - f_i(\mathbf{x}^*)) - u\alpha(t) \sum_{i=1}^n \pi_i \|\mathbf{x}_i(t) - \mathbf{x}^*\|^2. \quad (32)$$

上式最后一个不等式是由 $-\sum_{i=1}^n \pi_i \frac{u_i \alpha(t)}{z_{ii}(t)} \|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \leq -u\alpha(t) \sum_{i=1}^n \pi_i \|\mathbf{x}_i(t) - \mathbf{x}^*\|^2$ 得到的。

下面估计(31)式右边第二项,由假设1和 $\boldsymbol{\pi}^T \mathbf{A} = \boldsymbol{\pi}^T$, 有:

$$\sum_{i=1}^n \frac{-2\alpha(t)\pi_i}{z_{ii}(t)} \langle \mathbf{g}_i(t), \mathbf{y}_i(t) - \mathbf{x}_i(t) \rangle \leq 2L\eta\alpha(t) \sum_{i=1}^n \pi_i \sum_{j=1}^n a_{ij} \|\mathbf{x}_j(t) - \bar{\mathbf{x}}(t)\| + 2L\eta\alpha(t) \sum_{i=1}^n \pi_i \sum_{j=1}^n a_{ij} \|\bar{\mathbf{x}}(t) - \mathbf{x}_i(t)\| =$$

$$4L\eta\alpha(t)\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|. \quad (33)$$

将(32),(33)式代入(31)式,联立(21)和(22)式并取条件期望,由假设 3 和引理 2,可得:

$$\begin{aligned} & \sum_{i=1}^n \frac{2\pi_i\alpha(t)}{z_{ii}(t)} E[\langle \tilde{\mathbf{g}}_i(t), \mathbf{x}_i(t) - \mathbf{x}^* \rangle | \mathcal{F}_{t-1}] \leqslant \\ & 2(n+3\eta)L\alpha(t)(M_1\lambda^t + M_2\alpha(t-1)) - 2\alpha(t)E[f(\mathbf{x}(t)) - f(\mathbf{x}^*)] + \\ & 2n\eta CL\rho\lambda^t\alpha(t) - u\alpha(t)\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2. \end{aligned} \quad (34)$$

联立(19),(27)和(34)式,有:

$$\begin{aligned} E\left[\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t+1) - \mathbf{x}^*\|^2 | \mathcal{F}_t\right] & \leqslant (1 - u\alpha(t))E\left[\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2\right] - 2\alpha(t)E[f(\mathbf{x}_i(t)) - f(\mathbf{x}^*)] + \\ & (2(n+3\eta)LM_1 + 2n\eta CL\rho)\lambda^t\alpha(t) + (2L(n+3\eta)M_2 + 8M_3 + \eta^2 G^2)\alpha^2(t-1). \end{aligned}$$

由 $\alpha(t) = \frac{a}{t+1}$, $\alpha(0) = a \geqslant \frac{1}{u}$, 则 $u\alpha(t) \geqslant \frac{1}{t+1}$, 有:

$$\begin{aligned} E\left[\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t+1) - \mathbf{x}^*\|^2 | \mathcal{F}_t\right] & \leqslant \left(1 - \frac{1}{t+1}\right)E\left[\sum_{i=1}^n\pi_i\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2\right] - 2\alpha(t)E[f(\mathbf{x}_i(t)) - f(\mathbf{x}^*)] + \\ & (2L(n+3\eta)M_1 + 2n\eta CL\rho)\lambda^t\alpha(t) + (2L(n+3\eta)M_2 + 8M_3 + \eta^2 G^2)\alpha^2(t-1). \end{aligned} \quad (35)$$

在(35)式两边同时乘以 $(t+1)$, 并利用 $\frac{t+1}{t} \leqslant 2$, 两边同时从 $t=0$ 到 T 求和, 整理可得:

$$\begin{aligned} & 2a\sum_{t=0}^T E[f(\mathbf{x}_i(t)) - f(\mathbf{x}^*)] + (T+1)E\left[\sum_{i=1}^n\pi_i\|\mathbf{x}_i(T+1) - \mathbf{x}^*\|^2\right] \leqslant \\ & \frac{a(2L(n+3\eta)M_1 + 2n\eta CL\rho)}{1-\lambda} + 2a(2L(n+3\eta)M_2 + 8M_3 + \eta^2 G^2)(\ln(T+1) + 1). \end{aligned} \quad (36)$$

由 $\alpha(t) = \frac{a}{t+1}$, 有 $\sum_{t=0}^T \alpha(t-1) = \sum_{t=0}^{T-1} \frac{a}{t+1} \leqslant a\left(1 + \int_0^T \frac{1}{t+1} dt\right) \leqslant a(\ln(T+1) + 1)$.

(36)式两边同时除以 $(T+1)$, 并利用 Jensen 不等式, 可得:

$$\begin{aligned} 2aE[f(\hat{\mathbf{z}}_i(t)) - f(\mathbf{x}^*)] & \leqslant a(2L(n+3\eta)M_1 + 2n\eta CL\rho)\frac{1}{1-\lambda}\frac{1}{T+1} + \\ & 2a(2L(n+3\eta)M_2 + 8M_3 + \eta^2 G^2)(\ln(T+1) + 1)\frac{1}{T+1}. \end{aligned}$$

利用最优性条件, 结合 f 的 u -强凸性和 $a \geqslant \frac{1}{u}$, 有: $\|E[\hat{\mathbf{z}}_i(T)] - \mathbf{x}^*\|^2 \leqslant \frac{2}{u}[E[f(\hat{\mathbf{z}}_i(T))] - f(\mathbf{x}^*)]$. 令

$$C_3 = \frac{a(2L(n+3\eta)M_1 + 2n\eta CL\rho)}{1-\lambda}, C_4 = 2a(2L(n+3\eta)M_2 + 8M_3 + \eta^2 G^2),$$

则结论成立。

注 2 1) 定理 1 表明, 当 f_i 是凸函数, $\mathbf{z}_i(t)$ 逐渐增大时, $E[f(\mathbf{z}_i(T))]$ 能收敛到最优值 $f(\mathbf{x}^*)$, 收敛率为 $\mathcal{O}\left(\frac{\ln(T)}{\sqrt{T}}\right)$ 。通过对比, 算法 1 与文献[8]提出的没有量化的分布式梯度算法具有同阶的收敛速率, 但定理 1 的收敛性结果是建立在更一般的非平衡有向图通讯网络上。

2) 定理 2 表明, 当 f_i 是强凸函数, $\mathbf{z}_i(t)$ 逐渐增大时, $\hat{\mathbf{z}}_i(T)$ 能收敛到最优解 $\mathbf{x}^*$, 收敛率为 $\mathcal{O}\left(\frac{\ln(T)}{T}\right)$ 。通过对比, 算法 1 与文献[8]提出的没有量化的分布式梯度算法具有同阶的收敛速率。

3 数值实验

在无线传感器(sensor)网络上, 考虑以下参数估计问题^[2]:

$$\min_{\mathbf{x} \in \chi} \left\{ f(\mathbf{x}) \triangleq \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \|\mathbf{x} - \xi_{ij}\|^2 \right\},$$

其中: $\chi \subset \mathbf{R}^d$ 是约束集, d 是维数, n 是 sensor 的个数, $(\xi_{i1}, \dots, \xi_{im})$ 是第 i 个 sensor 可访问的数据点, m 是数据点

的个数。

在仿真中,令 $\chi = [-100, 100]^d$, $m = n = 10$ 和 $d = 100$ 。对所有 $i \in V$ 和 $j = 1, \dots, m$, 独立地产生一组数据 $\xi_{ij} \sim \xi_{jx} + \mathcal{N}(0, I)$, 其中 ξ_{jx} 是在 $[0, 100]$ 均匀随机产生的数, $\mathcal{N}(0, I)$ 是标准正态分布。取步长为 $\alpha(t) = \frac{0.5}{t+1}$ 。按照文献[8]的方法来产生满足假设 2 的通讯网络, 并使用最大化误差 $E_{\max} = \max_i \{ \|\bar{\mathbf{x}}_i(T) - \mathbf{x}^*\| \}$ 来度量算法的性能, 这里 $\bar{\mathbf{x}}_i(T) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_i(t)$, $\mathbf{x}^*$ 是问题的最优解。

实验的结果如图 1 所示。图 1a 给出了文献[8]中提出的算法 DPSG 和算法 1 在不同量化级别下的数值对比结果。从图 1a 可以看出, 两种算法都能收敛。算法 DPSG 的收敛精度略好于算法 1, 但算法 1 仅仅使用了量化的信息, 而算法 DPSG 使用的是精确信息, 这也验证了定理 1 的结果。图 1b 给出了算法 DPSG 和算法 1 在不同量化级别下关于最大化误差和通讯字节数的数值对比结果。从图 1b 中观察到, 在达到相同的最大化误差条件下, 不同的量化水平需要不同的字节数。通过比较图 1b 中的 3 种情况, 不难发现, 没有量化的算法 DPSG 需要最多的字节数, 而使用 8 字节量化的算法 1 则需要最少的字节数。通过对比, 使用 8 字节量化的算法 1 比不量化的算法 DPSG 节省约近 6 倍的字节数, 这极大地节约了通讯成本。

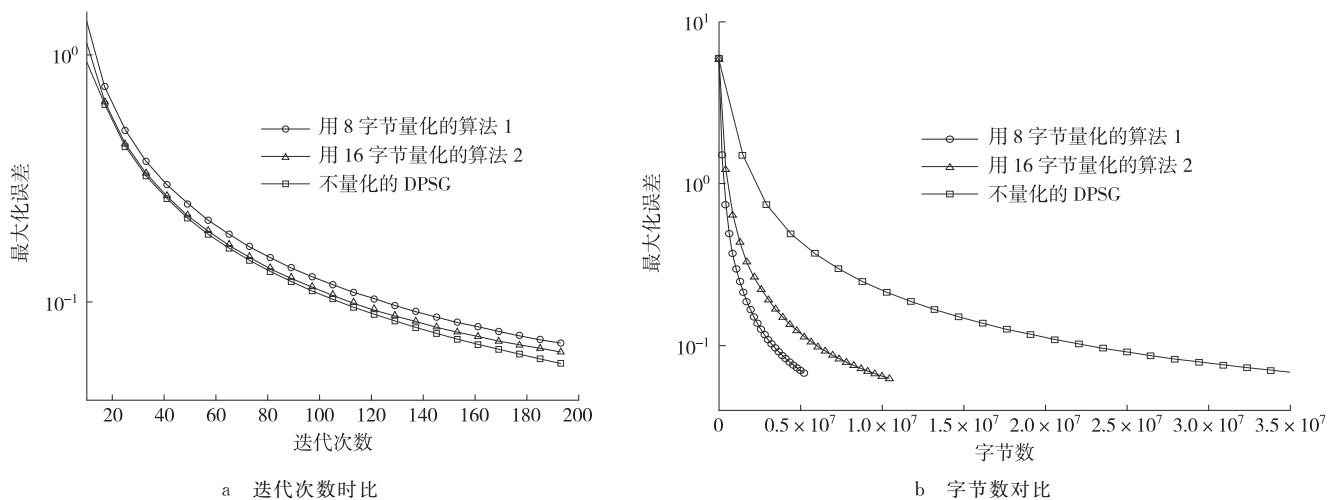


图 1 算法 1 和算法 DPSG 的数值对比

Fig. 1 Comparisons between algorithm 1 and DPSG

4 结语

针对分布式优化中会经常遇到带宽受限的问题, 本文提出了一种自适应通讯量化的分布式次梯度方法, 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[8]的分布式次梯度算法相比, 本文的算法极大地节省了通讯成本。与文献[5]的算法不同的是, 本文的算法适用于更一般的非平衡有向通讯网络。

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 and Applications, 2011, 28(10): 1506-1512.
- [3] 余淑辉, 李觉友, 杜学武. 基于 Push-Sum 的分布式 Gradient-Free 算法研究[J]. 重庆师范大学学报(自然科学版), 2019, 36(2): 11-17.
YU S H, LI J Y, DU X W. Research on distributed Gradient-free algorithm based on Push-sum [J]. Journal of Chongqing Normal

University (Natural Science), 2019, 36(2): 11-17.

- [4] 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 sub-gradient methods in solving non-smooth optimization problems[J]. Journal of Chongqing Normal University (Natural Science), 2013, 30(6): 25-30.
- [5] DOAN T T, MAGULURI S T, ROMBERG J. Fast convergence rates of distributed sub-gradient methods with adaptive quantization[J]. IEEE Transactions on Automatic Control, 2020, 66(5): 2191-2205.
- [6] RAM S S, NEDIĆ A, VEERAVALLI V V. Distributed stochastic subgradient projection algorithms for convex optimization[J]. Journal of Optimization Theory and Applications, 2010, 147(3): 516-545.
- [7] NEDIĆ A, OZDAGLAR A. Distributed subgradient methods for multi-agent[J]. IEEE Transactions on Automatic Control, 2009, 54(1): 48-61.
- [8] MAI V S, ABED E H. Distributed optimization over directed graphs with row stochasticity and constraint regularity[J]. Automatic, 2019, 102: 94-104.

Operations Research and Cybernetics

Distributed Subgradient Method with Adaptive Quantization

CHEN Lanlan, LI Jueyou, FAN Jing

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

Abstract: [Purposes] Considering that the network communication subject to the limited bandwidth, a distributed subgradient method with adaptive communication quantization is proposed to solve a class of distributed convex optimization problem. [Methods] Firstly, an adaptive communication quantization scheme is designed, and secondly, a distributed subgradient algorithm with quantization is proposed over more general directed network only requiring the weighed matrix row stochastic. [Findings] Get the convergence rate of the algorithm when the objective function is convex function and strong convex function. [Conclusions] Numerical simulations show that the proposed algorithm can effectively reduce the communication cost, apply to more general unbalanced directed networks.

Keywords: distributed optimization; adaptive quantization; row stochastic matrix

(责任编辑 陈 乔)