

# 一种基于条件梯度的加速分布式在线学习算法

吴庆涛<sup>1</sup> 朱军龙<sup>1</sup> 葛泉波<sup>2</sup> 张明川<sup>1</sup>

**摘要** 由于容易实施, 基于投影梯度的分布式在线优化模型逐渐成为一种主流的在线学习方法. 然而, 在处理大数据应用时, 投影步骤成为该方法的计算瓶颈. 近年来, 研究者提出了面向凸代价函数的分布式在线条件梯度算法, 其悔界为  $O(T^{3/4})$ , 其中  $T$  是一个时间范围. 该算法存在两方面的问题, 一是其悔界劣于公认的悔界  $O(\sqrt{T})$ ; 二是没有分析非凸代价函数的收敛性能, 而实际应用中代价函数大部分是非凸函数. 因此, 提出一种基于条件梯度的加速分布式在线学习算法, 使用 Frank-Wolfe 步骤替代投影步骤, 避免昂贵的投影计算. 文中证明当局部代价函数为凸函数时, 所提算法达到公认的悔界  $O(\sqrt{T})$ ; 当局部代价函数为潜在非凸函数时, 所提算法以速率  $O(\sqrt{T})$  收敛到平稳点. 最后, 仿真实验验证了所提算法的性能与理论证明的结论.

**关键词** 条件梯度, 分布式在线学习, 悔界, 收敛速率

**引用格式** 吴庆涛, 朱军龙, 葛泉波, 张明川. 一种基于条件梯度的加速分布式在线学习算法. 自动化学报, 2024, 50(2): 386-402

**DOI** 10.16383/j.aas.c210830

## An Accelerated Distributed Online Learning Algorithm Based on Conditional Gradient

WU Qing-Tao<sup>1</sup> ZHU Jun-Long<sup>1</sup> GE Quan-Bo<sup>2</sup> ZHANG Ming-Chuan<sup>1</sup>

**Abstract** The distributed online optimization model based on projected gradient has become a prominent paradigm for online learning due to its simplicity. However, this paradigm is incapable of handling massive data applications because the projection step becomes the computational bottleneck. Recent studies have presented the distributed online conditional gradient algorithms for convex cost functions, which achieved an  $O(T^{3/4})$  regret bound, where  $T$  is a time horizon. There are two negative problems for those algorithms. The first one is that the regret bound of those algorithms is worse than the well known regret bound  $O(\sqrt{T})$ . The second one is that the convergence performance of the presented algorithms has not been analyzed for non-convex cost functions, which are common in practice. In this paper, we propose an accelerated distributed online learning algorithm based on the conditional gradient method over networks, which avoids the prohibitively expensive projection steps by using Frank-Wolfe step. Moreover, when the local cost functions are convex, we show that the regret bound of  $O(\sqrt{T})$  is achieved. When the local cost functions are potentially non-convex, we also show that the algorithm converges to some stationary points at rate of  $O(\sqrt{T})$ . Finally, the performance of the proposed algorithm and the theoretical results are verified by simulation experiments.

**Key words** Conditional gradient, distributed online learning, regret bound, convergence rate

**Citation** Wu Qing-Tao, Zhu Jun-Long, Ge Quan-Bo, Zhang Ming-Chuan. An accelerated distributed online learning algorithm based on conditional gradient. *Acta Automatica Sinica*, 2024, 50(2): 386-402

近年来, 学术界和工业界对分布式优化产生了

浓厚的兴趣<sup>[1-3]</sup>. 实际应用中的许多经典问题本质上都是分布式优化问题. 例如数据管理问题<sup>[4-5]</sup>、资源分配问题<sup>[6-7]</sup>、目标检测与跟踪问题<sup>[8-9]</sup>、智能电网<sup>[10]</sup>等. 在这些应用中, 数据总量规模庞大, 分散在不同的数据中心; 节点计算能力有限, 分散在不同的物理位置. 为了提高这些系统的工作效率, 均离不开分布式优化算法<sup>[11-13]</sup>.

当前, 多数分布式优化算法的局部代价函数是非时变的. 然而, 在许多动态变化环境中, 局部代价函数可能是时变的. 例如, 传感器网络中的估计问题, 由于受干扰和噪声影响, 每个传感器的观察结果随时间变化. 因此, 动态变化环境的分布式优化

收稿日期 2021-09-01 录用日期 2022-03-01

Manuscript received September 1, 2021; accepted March 1, 2022

国家自然科学基金 (62033010, 61871430, 61976243), 中原科技创新领军人才 (214200510012, 224200510004) 资助

Supported by National Natural Science Foundation of China (62033010, 61871430, 61976243) and Leading Talents of Science and Technology in the Central Plain of China (214200510012, 224200510004)

本文责任编辑 秦家虎

Recommended by Associate Editor QIN Jia-Hu

1. 河南科技大学信息工程学院 洛阳 471023 2. 南京信息工程大学自动化学院 南京 210044

1. School of Information Engineering, Henan University of Science and Technology, Luoyang 471023 2. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044

须考虑局部代价函数的时变性, 即为分布式在线优化. 在分布式在线优化中, 智能体仅在每一轮动作结束之后才能获得其局部代价函数值, 即智能体无法获取其未来代价函数. 从这个意义上说, 分布式在线优化本质上不同于分布式优化.

近十年来, 机器学习领域主要关注集中式在线优化问题<sup>[14]</sup>, 定义了一个衡量在线优化算法性能的“悔函数”, 即算法代价与事后最佳行为代价之差<sup>[15]</sup>. 受分布式优化的启发, 一些学者开始研究分布式在线优化算法<sup>[16-21]</sup>. 具体而言, 针对无约束优化问题, 当代价函数是凸函数时, 可达到悔界  $O(\sqrt{T})$ <sup>[16-17]</sup>; 当代价函数是强凸函数时, 可达到悔界  $O((\log T)^2)$ <sup>[18]</sup> 或悔界  $O(\log T)$ <sup>[16-17]</sup>. 但在实际应用中, 大部分优化问题是受约束的<sup>[22]</sup>. 因此针对约束优化问题, 当代价函数是凸函数时, 可达到悔界  $O(\sqrt{T})$ <sup>[19-21]</sup>; 当代价函数是强凸函数时, 可达到悔界  $O(\log T)$ <sup>[21]</sup>.

随着各类智能设备的广泛使用, 许多应用中出现了大数据集, 为了避免在线优化算法投影算子造成的大数据计算瓶颈, 就需要考虑高维约束优化问题<sup>[23]</sup>. 为此, 针对高维约束优化的无投影算法应运而生, 即 Frank-Wolfe 算法<sup>[24]</sup>. 在 Frank-Wolfe 算法中, 使用了一个线性优化步骤替代投影算子. 近年来, Frank-Wolfe 算法<sup>[25-26]</sup> 在许多领域广受关注. 此外, 多种 Frank-Wolfe 算法的变体<sup>[27-31]</sup> 也相继提出以解决各种类似问题. 但是这些算法都是针对集中式场景, 难以适应分布式应用. 为此, 针对非时变的代价函数提出一种分布式 Frank-Wolfe 算法<sup>[32]</sup>. 但在实际中, 代价函数通常是随时间变化的. 针对此问题提出了无投影分布式在线算法<sup>[33-34]</sup>, 可以达到凸局部代价函数的悔界  $O(T^{3/4})$ . 但这个悔界劣于分布式在线优化算法领域公认的悔界  $O(\sqrt{T})$ .

因此, 进一步优化分布式在线学习算法的悔界仍然是一个亟待解决的问题. 本文针对多智能体网络中的高维约束优化问题, 使用 Frank-Wolfe 步骤替代投影算子, 提出了一种分布式条件梯度在线学习算法, 主要贡献如下:

1) 提出了 Frank-Wolfe 在线学习算法, 每个智能体仅利用自己及从邻居接收的信息进行分布式在线学习, 解决了某些场景无法使用集中式学习的应用需求.

2) 分析了所提算法的性能. 当局部代价函数为凸时, 算法的悔界为  $O(\sqrt{T})$ ; 当局部代价函数为非凸时, 算法以速率  $O(\sqrt{T})$  收敛于平稳点.  $O(\sqrt{T})$  是当前同类算法能够达到的最好性能.

3) 通过仿真实验验证了所提出算法的性能和理论分析的结论.

本文其余部分安排如下: 第 1 节给出一些预备知识; 第 2 节对所研究问题进行形式化描述, 并提

出了一种分布式条件梯度在线学习算法; 第 3 节给出了一些假设和主要结果; 第 4 节对主要结果进行详细证明; 第 5 节描述了仿真实验与仿真结果; 最后, 对本文进行了总结.

## 1 符号及定义

为了方便描述, 本节介绍一些符号约定和数学基础. 在本文中, 所有向量都是列向量, 符号  $\mathbf{R}$  和  $\mathbf{R}^d$  分别表示实数集和  $d$  维实欧几里得空间; 符号  $\mathbf{R}^+$  和  $\mathbf{Z}^+$  分别表示正实数集和正整数集; 符号  $\langle \cdot, \cdot \rangle$  表示实欧几里得空间的内积; 符号  $\|\cdot\|_2$  表示标准欧几里得范数. 设  $\mathbf{1}$  表示向量中所有元素为 1 的列向量;  $D$  表示约束集  $X$  相对于标准欧几里得范数  $\|\cdot\|_2$  的直径, 例如  $D = \sup_{\mathbf{x}, \mathbf{x}' \in X} \|\mathbf{x} - \mathbf{x}'\|_2$ .  $E[X]$  表示随机变量  $X$  的期望值. 设凸紧集  $X$  是  $\mathbf{R}^d$  的一个子集. 此外, 与函数  $f$  相关的一些定义表述如下.

**定义 1**<sup>[35]</sup>. 对于任意的  $\mathbf{x}, \mathbf{y} \in X$ ,  $\alpha \in [0, 1]$ , 若有  $f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})$ , 则称函数  $f: X \mapsto \mathbf{R}$  为凸的.

**定义 2**<sup>[32]</sup>. 对于任意的  $\mathbf{x}, \mathbf{y} \in X$ , 若有

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\mu}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \quad (1)$$

则称函数  $f: X \mapsto \mathbf{R}$  为  $\mu$ -强凸的, 其中  $\mu$  是一个非负常数.

**定义 3**<sup>[32]</sup>. 对于任意的  $\mathbf{x}, \mathbf{y} \in X$ , 若有

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{\beta}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \quad (2)$$

则称函数  $f: X \mapsto \mathbf{R}$  为  $\beta$ -光滑, 其中  $\beta$  是一个正常数. 另外, 式 (2) 等价于以下关系式

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq \beta \|\mathbf{x} - \mathbf{y}\|_2, \quad \forall \mathbf{x}, \mathbf{y} \in X \quad (3)$$

**定义 4**<sup>[32]</sup>. 对于任意的  $\mathbf{x}, \mathbf{y} \in X$ , 若有

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L \|\mathbf{x} - \mathbf{y}\|_2 \quad (4)$$

则称函数  $f: X \mapsto \mathbf{R}$  为  $L$ -Lipchitz, 其中  $L$  是一个正常数.

如果函数  $f$  是强凸的并且  $\mathbf{x}^* = \arg \min_{\mathbf{x} \in X} f(\mathbf{x})$ , 则有

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \mu \|\mathbf{x} - \mathbf{x}^*\|_2^2 \quad (5)$$

此时, 若式 (1) 中的  $\mu = 0$ , 则称函数  $f$  为凸函数.

## 2 问题形式化和算法

本文考虑一个由  $N$  个智能体组成的网络, 表示为图  $G(V, E)$ , 其中,  $V = \{1, \dots, N\}$  表示智能体的集合,  $E \subseteq V \times V$  表示边的集合. 假设图是固定无向图, 符号  $(i, j) \in E$  表示对于所有的  $i, j \in V$ , 智能体  $j$  可以直接向智能体  $i$  发送信息. 若两个智能

体可以直接交换信息, 则它们是邻居.  $N_j$  表示包括智能体  $j$  本身的  $j$  的邻居集合. 此外, 使用一个  $N$ -by- $N$  邻接矩阵  $\mathbf{A}$  表示通信模式, 并假设  $\mathbf{A}$  为双随机, 定义为  $\mathbf{A} = [a_{ij}]$ ,  $i, j = 1, \dots, N$ .

本文主要研究分布式在线优化问题, 即每个智能体的局部代价函数随时间变化. 在每轮时隙  $t = 1, \dots, T$  中, 智能体  $i \in \{1, \dots, N\}$  必须从凸紧集  $X \subset \mathbf{R}^d$  中选择一个点  $\mathbf{x}_i(t)$ . 此时环境生成一个代价函数  $f_t^i: X \mapsto \mathbf{R}$  作为回报, 并且智能体  $i$  产生了代价  $f_t^i(\mathbf{x}_i(t))$ . 这样, 就转化为在一个时间范围  $T$  内协作优化全局代价函数

$$\min_{\mathbf{x} \in X} f(\mathbf{x}) = \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N f_t^i(\mathbf{x}) \quad (6)$$

其中, 代价函数  $f_t^i: X \mapsto \mathbf{R}$  是凸的或是潜在非凸的, 并对于智能体  $i \in V$  在  $t \in \{1, \dots, N\}$  轮选择动作  $\mathbf{x}_i(t)$  后依然是有效的. 由于每个智能体只能利用其本地信息及其与邻居交换的信息, 因此需要设计分布式在线学习算法解决在线优化问题 (6).

悔函数是衡量在线算法性能的重要指标, 定义为<sup>[21]</sup>

$$R(T) = \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N f_t^i(\bar{\mathbf{x}}(t)) - \min_{\mathbf{x} \in X} f(\mathbf{x}) \quad (7)$$

这样, 研究目标转化为设计一个能够解决问题 (6) 的分布式在线算法, 可以在  $T$  时间达到次线性悔界, 即  $\limsup_{T \rightarrow \infty} R(T)/T = 0$ .

本文借鉴集中式 Frank-Wolfe 在线学习算法<sup>[28]</sup>, 提出了分布式无投影在线学习算法. 在该算法中, 每一个智能体  $i \in \{1, \dots, N\}$  线性地组合它自己的估计值和它从邻居接收的估计值, 然后计算一个能够替代局部梯度的聚合梯度. 最后, 每个智能体  $i$  执行一个 Frank-Wolfe 步骤更新它的估计值. 在该算法中, 每个智能体只知道自己的局部信息, 并只采取局部计算和局部通信. 具体算法描述如算法 1 所示. 相关的更新过程见式 (8) ~ (12).

### 算法 1. 基于条件梯度的加速分布式在线学习算法

1) 输入. 起始点  $\mathbf{x}_i(1)$ ,  $i = 1, \dots, N$ ; 最大时间范围  $T$ ; 双随机矩阵  $\mathbf{A} = [a_{ij}] \in \mathbf{R}^{N \times N}$ ; 智能体个数  $N$ .

2) for  $t = 1, \dots, T$  do

3) for 每个智能体  $i = 1, \dots, N$  do

4) 一致性步骤: 与邻居节点  $\mathbf{x}_i(t)$  交互, 即执行式 (8).

5) 聚合步骤: 计算聚合梯度代替平均梯度, 即执行式 (9) 和式 (10).

6) Frank-Wolfe 步骤: 计算式 (11), 并根据式

(12) 更新估计值  $\mathbf{x}_i(t)$ .

7) end for

8) end for

9) 输出. 对于所有  $i \in V$ ,  $\{\mathbf{x}_i(t) : 1 \leq t \leq T\}$ .

算法 1 中, 一致性步骤为

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

聚合步骤为

$$\begin{aligned} \mathbf{s}_t^i(t) &= \sum_{j \in N_i} a_{ij} \mathbf{s}_t^j(t-1) + \\ &\quad \nabla f_t^i(\mathbf{z}_i(t)) - \nabla f_t^i(\mathbf{z}_i(t-1)) \end{aligned} \quad (9)$$

$$\mathbf{S}_t^i(t) = \frac{1}{t} \sum_{\tau=1}^t \sum_{j=1}^N a_{ij} \mathbf{s}_\tau^j(t) \quad (10)$$

Frank-Wolfe 步骤为

$$\mathbf{v}_i(t) = \arg \min_{\mathbf{v} \in X} \langle \mathbf{v}, \mathbf{S}_t^i(t) \rangle \quad (11)$$

$$\mathbf{x}_i(t+1) = (1 - \gamma(t)) \mathbf{z}_i(t) + \gamma(t) \mathbf{v}_i(t) \quad (12)$$

其中,  $\gamma(t) \in (0, 1)$  是学习速率.

## 3 假设和主要结果

本节首先给出一些算法的假设, 然后描述本文的主要结果. 假设邻接矩阵  $\mathbf{A} = [a_{ij}] \in \mathbf{R}^{N \times N}$  满足以下假设.

**假设 1.** 图  $G$  是强连通图, 对于任意的  $i, j \in \{1, \dots, N\}$ , 有  $a_{ii} > 0$ ; 如果  $(i, j) \in E$ , 则  $a_{ij} > 0$ , 否则  $a_{ij} = 0$ . 并假设矩阵  $\mathbf{A}$  是双随机矩阵, 即对于所有的  $i, j \in V$ , 有  $\sum_{j=1}^N a_{ij} = 1$  和  $\sum_{i=1}^N a_{ij} = 1$ .

假设 1 是为了保证智能体  $i$  的信息能直接或间接传输至其他智能体. 该假设在分布式优化领域中是一个标准假设<sup>[21, 32, 36]</sup>. 需要注意  $\rho(\mathbf{A} - (1/N)\mathbf{1}\mathbf{1}^T) < 1$ , 其中  $\rho(\cdot)$  是矩阵的谱半径. 因此,  $\exists \lambda \in (0, 1)$ , 对于任意  $\mathbf{x} \in \mathbf{R}^N$ , 有

$$\begin{aligned} \|\mathbf{A}\mathbf{x} - \mathbf{1}\bar{x}\|_2 &= \left\| \left( \mathbf{A} - \frac{1}{N}\mathbf{1}\mathbf{1}^T \right) (\mathbf{x} - \mathbf{1}\bar{x}) \right\|_2 \leq \\ &\quad \lambda \|\mathbf{x} - \mathbf{1}\bar{x}\|_2 \end{aligned} \quad (13)$$

其中,  $\bar{x} = (1/N)\mathbf{1}^T\mathbf{x}$ . 当  $\exists t_0 \in \mathbf{Z}^+$  且  $\kappa \in (0, 1]$  时,  $\lambda$  的上界满足

$$\lambda \leq \frac{t_0^\kappa}{1 + t_0^\kappa} \times \left( \frac{t_0}{1 + t_0} \right)^\kappa \quad (14)$$

同时, 从式 (14) 可以得出

$$t_0 \geq \left\lceil \frac{1}{\lambda^{-\frac{1}{1+\kappa}} - 1} \right\rceil \quad (15)$$

为了分析算法 1 的收敛性能, 本文分别给出假

设 2 和假设 3<sup>[28, 32]</sup>.

**假设 2.** 约束集  $X$  是凸紧集, 最优集  $X^*$  是非空集.  $\mathbf{x}^* \in X^*$  是函数  $f$  的最小值, 其中  $\mathbf{x}^*$  是  $X$  的内部点, 即  $\eta = \inf_{\mathbf{x} \in \partial X} \|\mathbf{x} - \mathbf{x}^*\|_2 > 0$ , 其中  $\partial X$  表示约束集  $X$  的边界集.

假设 2 确保可以有效地求解一个线性函数的最小值.

**假设 3.** 对于所有的  $i \in \{1, \dots, N\}$  和  $t \in \{1, \dots, T\}$ , 局部代价函数  $f_t^i$  是  $\beta$ -光滑和  $L$ -Lipschitz 的.

在 Frank-Wolfe 算法中, 假设 3 是一个标准假设<sup>[28, 30, 32-34]</sup>, 便于理论分析算法的性能. 因此, 本文也采用该假设.

定义  $\Delta(t) = F_t(\bar{\mathbf{x}}(t)) - F_t(\mathbf{x}^*(t))$ ,  $F_t = (1/t) \times \sum_{\tau=1}^t f_\tau$ ,  $f_t = (1/N) \sum_{i=1}^N f_t^i$ ,  $\mathbf{x}^*(t) \in \arg \min_{\mathbf{x} \in X} F_t(\mathbf{x})$ . 这样, 可以通过获取  $\Delta(t)$  的上界得到悔界.

**定理 1.** 基于假设 1~3, 令步长  $\gamma(t) = 2/(t+2)$ , 对于任意的智能体  $i \in \{1, \dots, N\}$  和  $t \in \{1, \dots, T\}$ , 假设每个代价函数  $f_t^i$  都是一个带正常数  $\mu$  的  $\mu$ -强凸函数. 则  $\exists t_0 \in \mathbf{Z}^+$ , 对于所有的  $t \geq t^*$ , 其中  $t^*$  定义为

$$t^* = \inf \left\{ t \geq 1 : \frac{2C}{\sqrt{t+1}} + \frac{2C}{9}(1-\theta) + (4DC_2 + 2\beta D^2 + 4D\beta C_1)(1-\theta) \leq 0 \right\} \quad (16)$$

则有

$$\Delta(t) \leq \max \left\{ \frac{9}{2}(\beta D^2 + 2DC_2 + 2D\beta C_1), \frac{\frac{2\eta^2\mu}{\theta^2} - \frac{4C'}{9} + \sqrt{\left(\frac{2\eta^2\mu}{\theta^2}\right)^2 - \frac{16}{9} \times \frac{\eta^2\mu}{\theta^2} C'}}{8/81} \right\} \frac{1}{t+1} \quad (17)$$

其中,  $C_2 = \sqrt{N} \left\{ \left[ \delta + \frac{\beta(D+2C_1)}{1-\kappa} \right] + 2\beta(D+2C_1) \right\} t_0^\kappa$ , 且  $\kappa \in (0, 1)$  和  $\delta \in \mathbf{R}^+$ ;  $C_1 = \sqrt{ND} t_0^\kappa$ ,  $C' = 2\beta D^2 + 4D\beta C_1 + 4DC_2$ ,  $\theta > 1$ .

定理 1 的证明详见第 4 节. 由定理 1 可知,  $F_t(\bar{\mathbf{x}}(t))$  以速度  $O(1/(t+1))$  收敛到  $F_t(\mathbf{x}^*(t))$ , 而且收敛速度受网络智能体个数以及网络拓扑结构的影响. 由定理 1 可以得到算法 1 任意时刻的界, 同时得到算法 1 的悔界如下.

**定理 2.** 基于假设 1 和假设 2, 假设对于所有的  $i \in \{1, \dots, N\}$  和  $t \in \{1, \dots, T\}$ , 每个代价函数  $f_t^i$  是  $L$ -Lipchitz 和凸的, 并且  $f_t^i$  可能是非光滑的. 则对于

任意  $\mathbf{x}^* \in X$ , 可以得到

$$R(T) \leq 6\sqrt{\rho LD} \sqrt{T+1} \quad (18)$$

其中,

$$\rho = \max\{B_1, B_2\}$$

$$B_1 = \frac{9}{2}(LD + 2DC_2' + 2LC_1)$$

$$B_2 = \frac{81}{8} \times \left( \frac{2\eta^2 L}{D\theta^2} - \frac{4C''}{9} + \sqrt{\left( \frac{2\eta^2 L}{D\theta^2} \right)^2 - \frac{16}{9} \times \frac{\eta^2 L}{D\theta^2} C''} \right)$$

$$C_2' = \sqrt{N} \left\{ \left[ \delta + \frac{L(D+2C_1)}{D(1-\kappa)} \right] + 2 \frac{L(D+2C_1)}{D} \right\} \times t_0^\kappa$$

$$C'' = 2LD + 4LC_1 + 4DC_2$$

定理 2 的证明详见第 4 节. 根据定理 2, 可以得到强凸条件下的平方根悔界, 即算法 1 可实现悔界  $O(\sqrt{T})$ , 其中  $T$  是一个时间范围. 同时, 可以看到悔界由约束集  $X$  的直径  $D$  和局部代价函数梯度的上界  $L$  决定. 此外, 悔界也受通信网络连通性的影响. 与文献 [33] 和文献 [34] 相比, 本文使用了梯度追踪技术<sup>[13]</sup> 以提高算法的收敛速度. 在文献 [33] 和文献 [34] 中, 它们的悔界都为  $O(T^{3/4})$ , 而本文所提算法的悔界为  $O(\sqrt{T})$ , 明显优于以上两个工作. 具体比较见表 1.

表 1 不同算法的比较

Table 1 Comparison of different algorithms		
算法	凸代价函数	非凸代价函数
D-OCG <sup>[33]</sup>	$O(T^{3/4})$	—
D-BOCG <sup>[34]</sup>	$O(T^{3/4})$	—
本文算法	$O(\sqrt{T})$	$O(\sqrt{T})$

当代价函数可能是非凸时, 为了分析算法 1 的收敛性, 引入对偶间隙的定义为

$$\psi(t) = \max_{\mathbf{v} \in X} \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \mathbf{v} \rangle = \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle \quad (19)$$

其中,  $\hat{\mathbf{v}}(t) \in \arg \min_{\mathbf{v} \in X} \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v} \rangle$ . 如果  $\psi(t) = 0$ , 则估计值  $\bar{\mathbf{x}}(t)$  是  $\min_{\mathbf{x} \in X} F_t(\mathbf{x})$  的固定点. 因此, 本文使用类似文献 [25] 和文献 [30] 的方法分析对偶间隙, 非凸代价函数的收敛结果如下.

**定理 3.** 基于假设 1~3, 令步长  $\gamma(t) = 1/t^\kappa$ , 其中  $\kappa \in (0, 1)$ . 假设每个代价函数  $f_t^i$  是非凸的, 并且

时间范围  $T$  是均等的. 则  $\exists t_0 \in \mathbf{Z}^+$ , 对于所有的  $t \geq t_0$  和  $T \geq 13$ , 当  $\kappa \in [1/2, 1)$  时, 可以得到

$$\min_{t \in [\frac{T}{2}+1, T]} \psi(t) \leq \frac{1-\kappa}{T^{1-\kappa}} \times \left(1 - \left(\frac{2}{3}\right)^{1-\kappa}\right)^{-1} \times \alpha_\kappa \quad (20)$$

其中,

$$\alpha_\kappa = 4LD + \left(\frac{\beta D^2}{2} + 2DC_2\right) \log 2 + 2\beta DC_1 \left(\frac{2^\kappa - 1}{\kappa}\right)$$

当  $T \geq 6$  且  $\kappa \in (0, 1/2)$  时, 可以得到

$$\min_{t \in [\frac{T}{2}+1, T]} \psi(t) \leq \frac{1-\kappa}{T^\kappa} \times \left(1 - \left(\frac{2}{3}\right)^{1-\kappa}\right)^{-1} \times \hat{\alpha}_\kappa \quad (21)$$

其中,

$$\hat{\alpha}_\kappa = 4LD + \frac{2^\kappa - 1}{\kappa} (2DC_2 + 2\beta DC_1) + \frac{\beta D^2}{2} \times \frac{1 - (\frac{1}{2})^{1-2\kappa}}{1 - 2\kappa}$$

$$C_1 = \sqrt{N} D t_0^\kappa$$

$$C_2 = \sqrt{N} \left\{ \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right] + 2\beta(D + 2C_1) \right\} \times t_0^\kappa, \delta \in \mathbf{R}^+$$

定理 3 的证明详见第 4 节. 根据定理 3 可以得到, 如果  $\kappa \in [1/2, 1)$ , 算法 1 可以在每轮  $t \in [T/2 + 1, T]$  得到收敛速率  $O(1/T^{1-\kappa})$ ; 如果当  $\kappa \in (0, 1/2)$ , 算法 1 可以得到收敛速率  $O(1/T^\kappa)$ . 此外, 当设置  $\kappa = 1/2$  时, 算法可以得到最快的收敛速率  $O(1/\sqrt{T})$ . 可以看出, 收敛速率受梯度上界  $L$  和约束集直径  $D$  的影响.

### 4 性能分析

本节分析当代价函数为凸函数或潜在非凸函数时算法 1 的性能, 并给出主要结果的详细证明. 为了分析算法 1 的性能, 引入下面 3 个辅助向量, 并给出几个引理.

$$\bar{\mathbf{x}}(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i(t) \quad (22)$$

$$\bar{\mathbf{s}}_t(t) = \frac{1}{t} \times \frac{1}{N} \sum_{\tau=1}^t \sum_{i=1}^N \mathbf{s}_\tau^i(t) \quad (23)$$

$$\mathbf{g}_t(t) = \frac{1}{t} \times \frac{1}{N} \sum_{\tau=1}^t \sum_{i=1}^N \nabla f_\tau^i(\mathbf{z}_i(t)) \quad (24)$$

**引理 1.** 对于所有的  $t = 1, \dots, T$ , 有以下关系:

a)  $\bar{\mathbf{s}}_{t+1}(t+1) = \mathbf{g}_{t+1}(t+1);$

b)  $\bar{\mathbf{x}}(t+1) = (1 - \gamma(t)) \bar{\mathbf{x}}(t) + \gamma(t) \bar{\mathbf{v}}(t)$ , 其中  $\bar{\mathbf{v}}(t) = (1/N) \sum_{i=1}^N \mathbf{v}_i(t)$ .

引理 1 的证明见附录 A. 从引理 1 可以看出, 更新关系由平均序列  $\bar{\mathbf{s}}_t(t)$  和  $\bar{\mathbf{x}}(t)$  决定. 下面给出  $\|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2$  的有界性.

**引理 2.** 基于假设 1, 对于某些  $\kappa \in (0, 1]$  使  $\gamma(t) = 1/t^\kappa$ , 则对于所有的  $i = 1, \dots, N$  和  $t \geq t_0$ , 有

$$\max_{i \in \{1, \dots, N\}} \|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2 \leq C_1 t^{-\kappa} \quad (25)$$

其中,  $C_1 = \sqrt{N} D t_0^\kappa$ .

引理 2 的证明见附录 B. 从引理 2 可以看出, 当  $t$  趋于无穷大,  $\mathbf{z}_i(t)$  和  $\bar{\mathbf{x}}(t)$  的均方差趋近于零. 对于所有的  $i \in \{1, \dots, N\}$ , 下面给出  $\|\mathbf{S}_t^i(t) - \mathbf{g}_t(t)\|_2$  的有界性.

**引理 3.** 基于假设 1, 存在  $\kappa \in (0, 1)$  使得  $\gamma(t) = 1/t^\kappa$ , 则对于所有的  $i = 1, \dots, N$  和  $t \geq t_0$ , 有

$$\max_{i \in \{1, \dots, N\}} \|\mathbf{S}_t^i(t) - \mathbf{g}_t(t)\|_2 \leq C_2 t^{-\kappa} \quad (26)$$

其中,  $C_2 = \sqrt{N} \left\{ \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right] + 2\beta(D + 2C_1) \right\} t_0^\kappa$ , 且对于所有的  $i = 1, \dots, N$  和  $t \geq 1$ ,  $\delta$  满足  $\delta \geq \|\mathbf{s}_{t+1}^i(1) - \phi_{t+1}(1)\|_2$ .

引理 3 的证明见附录 C. 从引理 3 可以得出, 当  $t$  趋于无穷大时,  $\mathbf{S}_t^i(t)$  和  $\mathbf{g}_t(t)$  的均方误差趋近于零. 下面利用引理 1 ~ 3 证明定理 1.

**定理 1 证明.** 为了描述方便, 首先定义辅助变量  $f_t = (1/N) \sum_{i=1}^N f_t^i$  和  $F_t = (1/t) \sum_{\tau=1}^t f_\tau$ , 并将常数  $C$  定义为

$$C = \max \left\{ \frac{9}{2} (\beta D^2 + 2DC_2 + 2D\beta C_1), \frac{\frac{2\eta^2 \mu}{\theta^2} - \frac{4C'}{9} + \sqrt{\left(\frac{2\eta^2 \mu}{\theta^2}\right)^2 - \frac{16}{9} \times \frac{\eta^2 \mu}{\theta^2} C'}}{\frac{8}{81}} \right\} \quad (27)$$

其中,  $\theta > 1$ ,  $C' = 2\beta D^2 + 4D\beta C_1 + 4DC_2$ . 显然, 当  $t = t^*$  时, 定理 1 成立. 然后, 假设对于某些  $t \geq t^*$ ,  $\Delta(t) \leq C/(t+1)$  成立, 其中  $t^*$  的定义见式 (16). 从假设 2 可知函数  $F_t$  是  $\beta$ -光滑的, 因此根据引理 1 中的 b), 可以得出

$$F_t(\bar{\mathbf{x}}(t+1)) \leq F_t(\bar{\mathbf{x}}(t)) + \frac{\beta}{2} \|\bar{\mathbf{x}}(t+1) - \bar{\mathbf{x}}(t)\|_2^2 +$$

$$\begin{aligned}
& \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t+1) - \bar{\mathbf{x}}(t) \rangle \leq \\
& F_t(\bar{\mathbf{x}}(t)) + \frac{\beta}{2} \gamma^2(t) \|\bar{\mathbf{v}}(t) - \bar{\mathbf{x}}(t)\|_2^2 + \\
& \frac{\gamma(t)}{N} \sum_{i=1}^N \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v}_i(t) - \bar{\mathbf{x}}(t) \rangle \leq \\
& F_t(\bar{\mathbf{x}}(t)) + \frac{\beta}{2} \gamma^2(t) D^2 + \\
& \frac{\gamma(t)}{N} \sum_{i=1}^N \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v}_i(t) - \bar{\mathbf{x}}(t) \rangle \quad (28)
\end{aligned}$$

其中, 第 1 个不等式可由  $F_t$  是  $\beta$ -光滑函数得出, 第 2 个不等式依据引理 1 的 b) 得出, 第 3 个不等式根据  $X$  的有界性得到. 此外, 对于所有  $i=1, \dots, N$  和  $\mathbf{v} \in X$ , 可以得出

$$\begin{aligned}
& \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v}_i(t) - \bar{\mathbf{x}}(t) \rangle = \\
& \langle \mathbf{S}_t^i(t), \mathbf{v}_i(t) - \bar{\mathbf{x}}(t) \rangle + \\
& \langle \nabla F_t(\bar{\mathbf{x}}(t)) - \mathbf{S}_t^i(t), \mathbf{v}_i(t) - \bar{\mathbf{x}}(t) \rangle \leq \\
& \langle \mathbf{S}_t^i(t), \mathbf{v} - \bar{\mathbf{x}}(t) \rangle + \\
& D \times \|\mathbf{S}_t^i(t) - \nabla F_t(\bar{\mathbf{x}}(t))\|_2 \leq \\
& \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v} - \bar{\mathbf{x}}(t) \rangle + \\
& 2D \times \|\mathbf{S}_t^i(t) - \nabla F_t(\bar{\mathbf{x}}(t))\|_2 \quad (29)
\end{aligned}$$

其中, 第 1 个不等式通过在等式加上  $\mathbf{S}_t^i(t)$  再减去  $\mathbf{S}_t^i(t)$ , 由  $\mathbf{v}_i(t) \in \arg \min_{\mathbf{v} \in X} \langle \mathbf{v}, \mathbf{S}_t^i(t) \rangle$  确定; 最后一个不等式通过加上再减去  $\nabla F_t(\bar{\mathbf{x}}(t))$  得到. 为了得到式 (29) 的有界性, 需要得到  $\|\mathbf{S}_t^i(t) - \nabla F_t(\bar{\mathbf{x}}(t))\|_2$  的上界. 根据  $\mathbf{g}_t(t)$  的定义, 可以得出

$$\begin{aligned}
& \|\mathbf{S}_t^i(t) - \nabla F_t(\bar{\mathbf{x}}(t))\|_2 \leq \\
& \|\mathbf{S}_t^i(t) - \mathbf{g}_t(t)\|_2 + \|\mathbf{g}_t(t) - \nabla F_t(\bar{\mathbf{x}}(t))\|_2 \leq \\
& \frac{C_2}{t} + \frac{1}{Nt} \sum_{\tau=1}^t \sum_{i=1}^N \|\nabla f_\tau^i(\mathbf{z}_i(t)) - \nabla f_\tau^i(\bar{\mathbf{x}}(t))\|_2 \leq \\
& \frac{C_2}{t} + \beta \times \frac{1}{Nt} \sum_{\tau=1}^t \sum_{i=1}^N \|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2 \leq \\
& \frac{C_2}{t} + \frac{\beta}{t} \sum_{\tau=1}^t \frac{C_1}{t} \leq \\
& \frac{C_2}{t} + \frac{\beta C_1}{t} \quad (30)
\end{aligned}$$

其中, 第 1 个不等式可依据三角形不等式获得, 第 2 个不等式由范数的凸性和引理 3 得出, 第 3 个不等式利用函数  $\{f_\tau^1, \dots, f_\tau^N\}_{\tau=1}^T$  的  $\beta$ -光滑性得到, 第 4 个不等式根据引理 2 得出. 将式 (29) 和式 (30) 代入式 (28), 则对任意  $\mathbf{v} \in X$ , 有

$$\begin{aligned}
F_t(\bar{\mathbf{x}}(t+1)) & \leq F_t(\bar{\mathbf{x}}(t)) + \frac{\gamma(t)}{2} \beta D^2 + \\
& \gamma(t) \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v} - \bar{\mathbf{x}}(t) \rangle + \\
& 2D\gamma(t) \frac{C_2 + \beta C_1}{t} \quad (31)
\end{aligned}$$

同时, 还可以得到

$$\begin{aligned}
& F_t(\bar{\mathbf{x}}(t+1)) - F_t(\mathbf{x}^*(t+1)) \leq \\
& F_t(\bar{\mathbf{x}}(t+1)) - F_t(\mathbf{x}^*(t)) \leq \\
& F_t(\bar{\mathbf{x}}(t)) - F_t(\mathbf{x}^*(t)) + \frac{\gamma^2(t)}{2} \beta D^2 - \\
& \gamma(t) \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle + \\
& 2D\gamma(t) \frac{C_2 + \beta C_1}{t} \quad (32)
\end{aligned}$$

其中, 第 1 个不等式依据  $\mathbf{x}^*(t) = \arg \min_{\mathbf{x} \in X} F_t(\mathbf{x})$  获得, 最后一个不等式依据式 (31) 和  $\hat{\mathbf{v}}(t) \in \arg \min_{\mathbf{v} \in X} \langle \mathbf{v}, \nabla F_t(\bar{\mathbf{x}}(t)) \rangle$  得到. 此外, 根据  $\Delta(t) = F_t(\bar{\mathbf{x}}(t)) - F_t(\mathbf{x}^*(t))$ , 可以得出

$$\begin{aligned}
\Delta(t+1) & = \\
& \frac{t}{1+t} (F_t(\bar{\mathbf{x}}(t+1)) - F_t(\mathbf{x}^*(t+1))) + \\
& \frac{1}{1+t} (f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1))) \leq \\
& \frac{t}{1+t} (\Delta(t) - \gamma(t) \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle) + \\
& \frac{t}{1+t} \left( \frac{\gamma^2(t)}{2} \beta D^2 + 2D\gamma(t) \times \frac{C_2 + \beta C_1}{t} \right) + \\
& \frac{1}{1+t} (f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1))) \quad (33)
\end{aligned}$$

其中, 最后一个不等式根据式 (32) 得出.

另外, 假设  $F_t$  是  $\mu$ -强凸的且  $\mathbf{x}^*(t) \in \text{int}(X)$ , 其中,  $\text{int}(X)$  表示  $X$  的内点集. 则对于某些  $\xi \in [0, 1)$ ,  $\mathbf{x}^*(t)$  为

$$\mathbf{x}^*(t) = \bar{\mathbf{x}}(t) + \xi(\mathbf{w}(t) - \bar{\mathbf{x}}(t)) \quad (34)$$

其中,  $\mathbf{w}(t) \in \partial X$ . 由于  $F_t$  是  $\mu$ -强凸的, 可以得出

$$\begin{aligned}
& \frac{\mu}{2} \|\mathbf{x}^*(t) - \bar{\mathbf{x}}(t)\|_2^2 \leq \\
& F_t(\mathbf{x}^*(t)) - F_t(\bar{\mathbf{x}}(t)) - \\
& \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{x}^*(t) - \bar{\mathbf{x}}(t) \rangle = \\
& -\Delta(t) + \xi \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \mathbf{w}(t) \rangle \leq \\
& -\Delta(t) + \xi \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle \quad (35)
\end{aligned}$$

其中, 最后 1 个不等式根据  $\hat{\mathbf{v}}(t) \in \arg \min_{\mathbf{v} \in X} \langle \nabla F_t(\bar{\mathbf{x}}(t)), \mathbf{v} \rangle$  得出. 此外, 还可以得出

$$\begin{aligned} \frac{\mu}{2} \|\mathbf{x}^*(t) - \bar{\mathbf{x}}(t)\|_2^2 &= \\ \frac{\mu}{2} \xi^2 \|\mathbf{w}(t) - \bar{\mathbf{x}}(t)\|_2^2 &\geq \\ \frac{\mu}{2} \xi^2 \|\mathbf{w}(t) - \bar{\mathbf{x}}(t)\|_2^2 &\geq \frac{\mu}{2} \eta^2 \xi^2 \end{aligned} \quad (36)$$

其中, 第 1 个不等式依据下式得出

$$\|\mathbf{w}(t) - \bar{\mathbf{x}}(t)\|_2^2 \geq (1 - \xi)^2 \|\mathbf{w}(t) - \bar{\mathbf{x}}(t)\|_2^2 = \|\mathbf{w}(t) - \mathbf{x}^*(t)\|_2^2$$

最后一个不等式根据  $\eta$  的定义得出. 根据式 (35) 和式 (36), 可以得到

$$\begin{aligned} \Delta(t) &\leq \xi \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \bar{\mathbf{v}}(t) \rangle - \frac{\mu}{2} \eta^2 \xi^2 \leq \\ \frac{1}{2\mu\eta^2} (\langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle)^2 &\quad (37) \end{aligned}$$

其中, 最后一个不等式通过设  $\xi = \langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle / (\mu\eta^2)$  获得. 因此, 根据式 (37), 可以有

$$\langle \nabla F_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \hat{\mathbf{v}}(t) \rangle \geq \sqrt{2\mu\eta^2 \Delta(t)} \quad (38)$$

为了获得  $\Delta(t+1)$  的上界, 需要得出  $f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1))$  的上界. 由于对于所有的  $i \in \{1, \dots, N\}$  和  $t \in \{1, \dots, T\}$ , 每个代价函数  $f_t^i$  都是  $\mu$ -强凸的, 那么根据  $F_t$  的定义, 函数  $F_t$  也是  $\mu$ -强凸的. 类似地, 假设 3 表明  $F_t$  是  $\beta$ -光滑和  $L$ -Lipschitz 的. 由于  $\mathbf{x}^*(t) \in \arg \min_{\mathbf{x} \in X} F_t(\mathbf{x})$ , 则根据  $F_t$  的强凸性, 可以得出

$$\mu \|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t)\|_2^2 \leq F_t(\bar{\mathbf{x}}(t)) - F_t(\mathbf{x}^*(t)) \leq \frac{C}{t+1} \quad (39)$$

其中, 最后一个不等式根据归纳假设得出. 这样, 依据式 (39), 还可得出

$$\|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t)\|_2 \leq \sqrt{\frac{C}{\mu(t+1)}} \quad (40)$$

根据  $F_t$  的定义, 有

$$\begin{aligned} F_{t+1}(\mathbf{x}^*(t)) - F_{t+1}(\mathbf{x}^*(t+1)) &= \\ \frac{t}{t+1} (F_t(\mathbf{x}^*(t)) - F_t(\mathbf{x}^*(t+1))) &+ \\ \frac{1}{t+1} (f_{t+1}(\mathbf{x}^*(t)) - f_{t+1}(\mathbf{x}^*(t+1))) &\leq \\ \frac{1}{t+1} (f_{t+1}(\mathbf{x}^*(t)) - f_{t+1}(\mathbf{x}^*(t+1))) &\leq \\ \frac{L}{t+1} \|\mathbf{x}^*(t) - \mathbf{x}^*(t+1)\|_2 &\leq \frac{LD}{t+1} \end{aligned} \quad (41)$$

其中, 第 1 个不等式由  $F_t(\mathbf{x}^*(t)) \leq F_t(\mathbf{x}^*(t+1))$  得出, 第 2 个不等式依据  $F_t$  是  $L$ -Lipschitz 的获得, 最后一个不等式根据  $X$  的有界性得到. 依据式 (41), 得出

$$\|\mathbf{x}^*(t) - \mathbf{x}^*(t+1)\|_2 \leq \sqrt{\frac{LD}{\mu(t+1)}} \leq \sqrt{\frac{C}{\mu(t+1)}} \quad (42)$$

其中, 最后一个不等式根据  $C \geq LD$  得到. 由于  $\|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t+1)\|_2 \leq D$ , 可以得出

$$\begin{aligned} \|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t+1)\|_2 &\leq \min \left\{ \sqrt{\frac{4C}{\mu(t+1)}}, D \right\} \leq \\ \frac{C}{9L} \times \sqrt{\frac{1}{t+1}} &\quad (43) \end{aligned}$$

其中,  $C \geq \max\{324L^2/\mu, 9LD\}$ . 此外, 还可以得到

$$\begin{aligned} \|\bar{\mathbf{x}}(t+1) - \bar{\mathbf{x}}(t)\|_2 &= \gamma(t) \|\bar{\mathbf{v}}(t) - \bar{\mathbf{x}}(t)\|_2 \leq \\ D \times \gamma(t) &\leq \frac{C}{9L} \times \sqrt{\frac{1}{t+1}} \end{aligned} \quad (44)$$

其中, 最后一个不等式根据不等式  $2/(t+2) \leq \sqrt{1/(t+1)}$  对所有的  $t \geq 2$  都成立得出. 因此, 根据式 (43) 和式 (44), 得出

$$\|\bar{\mathbf{x}}(t+1) - \mathbf{x}^*(t+1)\|_2 \leq \frac{2C}{9L} \times \sqrt{\frac{1}{t+1}} \quad (45)$$

其中, 最后一个不等式根据三角不等式得到. 此外, 由于  $f_{t+1}$  是  $L$ -Lipschitz 的, 可以得出

$$f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1)) \leq \frac{2C}{9} \times \sqrt{\frac{1}{t+1}} \quad (46)$$

将式 (38) 和式 (46) 代入式 (33), 可以得到

$$\begin{aligned} \Delta(t+1) &\leq \sqrt{\Delta(t)} \left( \sqrt{\Delta(t)} - \gamma(t) \sqrt{2\mu\eta^2} \right) + \\ \frac{2C}{9(t+1)^{3/2}} + \frac{\gamma^2(t)}{2} \beta D^2 + & \\ 2D\gamma(t) \times \frac{\beta C_1 + C_2}{t+1} &\quad (47) \end{aligned}$$

当  $\Delta(t) - \gamma(t) \sqrt{2\mu\eta^2 \Delta(t)} \leq 0$ , 由式 (47) 得

$$\begin{aligned} \Delta(t+1) &\leq \\ \frac{2C}{9(t+1)^{3/2}} + \frac{2\beta D^2}{(t+1)^2} + 4D \frac{\beta C_1 + C_2}{(t+1)^2} &\leq \\ \left( 2\beta D^2 + 4DC_2 + \frac{2}{9}C + 4D\beta C_1 \right) \times \frac{1}{t+1} &\leq \\ \left( \frac{1}{3}C + 3\beta D^2 + 6DC_2 + 6D\beta C_1 \right) \times \frac{1}{t+2} &\quad (48) \end{aligned}$$

其中, 最后一个不等式依据  $n \geq 1$  时成立的不等式

$$\begin{aligned} \frac{1}{n+t-1} &\leq \frac{n+1}{n} \times \frac{1}{n+t} \text{ 得出. 根据 } C \text{ 的定义, 可得} \\ C &\geq \frac{9}{2} (\beta D^2 + 2DC_2 + 2D\beta C_1) \end{aligned} \quad (49)$$

因此, 得出

$$\Delta(t+1) \leq C \times \frac{1}{t+2}$$

当  $\Delta(t) - \gamma(t) \sqrt{2\mu\eta^2 \Delta(t)} > 0$ , 根据式 (47), 有

$$\begin{aligned} \Delta(t+1) - \frac{C}{t+2} &\leq \\ C \left( \frac{1}{t+1} - \frac{1}{t+2} \right) - \frac{\eta\sqrt{2\mu C}}{(t+1)^{\frac{3}{2}}} + \\ \frac{2C}{9} \frac{1}{(t+1)^{\frac{3}{2}}} + \frac{4DC_2 + 2\beta D^2 + 4D\beta C_1}{(t+1)^2} \\ \frac{C}{(t+1)^2} - \frac{\eta\sqrt{2\mu C}}{(t+1)^{\frac{3}{2}}} + \frac{2C}{9} \frac{1}{(t+1)^{\frac{3}{2}}} + \\ \frac{4DC_2 + 2\beta D^2 + 4D\beta C_1}{(t+1)^2} &\leq \\ \frac{1}{(t+1)^{\frac{3}{2}}} \left( \frac{C}{\sqrt{t+1}} - \eta\sqrt{2\mu C} + \frac{2C}{9} + \right. \\ (4DC_2 + 2\beta D^2 + 4D\beta C_1) \times \frac{1}{\sqrt{t+1}} &\left. \right) \leq \\ \frac{1}{(t+1)^{\frac{3}{2}}} \left( \frac{C}{\sqrt{t+1}} + \frac{2C}{9} (1-\theta) + \right. \\ (4DC_2 + 2\beta D^2 + 4D\beta C_1)(1-\theta) &\left. \right) \quad (50) \end{aligned}$$

其中, 第 2 个不等式依据不等式  $1/(t+1) - 1/(t+2) \leq 1/(t+1)^2$  得出, 最后一个不等式根据  $C$  的定义得到. 由于函数  $1/\sqrt{t+1}$  是一个单调递减函数, 因此  $t^*$  定义可由式 (16) 给出.

由于若  $\theta > 1$ , 则  $t^*$  值存在. 因此, 如果  $t > t^*$ , 则式 (50) 的右侧是非正的, 即

$$\Delta(t+1) \leq C \times \frac{1}{t+2}$$

则得到算法 1 任意时刻的界.  $\square$

定理 1 给出了算法 1 任意时刻的界. 利用定理 1, 下面给出定理 2 的证明.

**定理 2 证明.** 引入一个辅助函数:

$$\tilde{f}_t^i(\mathbf{x}) = \langle \nabla f_t^i(\bar{\mathbf{x}}(t)), \mathbf{x} \rangle + \frac{L}{D} \|\mathbf{x} - \mathbf{x}^*\|_2^2 \quad (51)$$

其中,  $\nabla f_t^i(\bar{\mathbf{x}}(t))$  表示  $f_t^i$  在  $\bar{\mathbf{x}}(t)$  处的梯度或次梯度满足  $\|\nabla f_t^i(\bar{\mathbf{x}}(t))\|_2 \leq L$ . 因此, 可得

$$\nabla \tilde{f}_t^i(\mathbf{x}) = \nabla f_t^i(\bar{\mathbf{x}}(t)) + \frac{2L}{D}(\mathbf{x} - \mathbf{x}^*) \quad (52)$$

由于  $\|\nabla f_t^i(\bar{\mathbf{x}}(t))\|_2 \leq L$  和  $\|\mathbf{x} - \mathbf{x}^*\|_2 \leq D$ , 则  $\|\nabla \tilde{f}_t^i(\mathbf{x})\|_2 \leq 3L$ . 因此, 对于所有的  $i \in \{1, \dots, N\}$

和  $t \in \{1, \dots, T\}$ , 函数  $\tilde{f}_t^i$  是  $3L$ -Lipschitz. 根据  $\tilde{f}_t^i$  的定义, 可得

$$\begin{aligned} \tilde{f}_t^i(\mathbf{x} + \mathbf{y}) - \tilde{f}_t^i(\mathbf{x}) &= \langle \nabla f_t^i(\bar{\mathbf{x}}(t)), \mathbf{y} \rangle + \\ \frac{L}{D} \|\mathbf{y}\|_2^2 + \frac{2L}{D} \langle \mathbf{x} - \mathbf{x}^*, \mathbf{y} \rangle &= \\ \langle \nabla \tilde{f}_t^i(\bar{\mathbf{x}}(t)), \mathbf{y} \rangle + \frac{L}{D} \|\mathbf{y}\|_2^2 &\quad (53) \end{aligned}$$

因此,  $\tilde{f}_t^i$  是  $(L/D)$ -强凸和  $(L/D)$ -光滑的. 根据定理 1, 可得

$$\tilde{\Delta}(t) = \tilde{F}_t(\bar{\mathbf{x}}(t)) - \tilde{F}_t(\mathbf{x}^*(t)) \leq \rho \times \frac{1}{t+1} \quad (54)$$

其中,

$$\rho = \max\{B_1, B_2\}$$

$$B_1 = \frac{9}{2}(LD + 2DC_2' + 2LC_1)$$

$$B_2 = \left( \frac{2\eta^2 L}{D\theta^2} - \frac{4C''}{9} + \sqrt{\left( \frac{2\eta^2 L}{D\theta^2} \right)^2 - \frac{16}{9} \times \frac{\eta^2 L}{D\theta^2} C''} \right) / \left( \frac{8}{81} \right)$$

且

$$\begin{aligned} C_2' &= \sqrt{N} \left\{ \left[ \delta + \frac{L(D+2C_1)}{D(1-\kappa)} \right] + 2 \frac{L(D+2C_1)}{D} \right\} \times t_0^\kappa \\ C'' &= 2LD + 4LC_1 + 4DC_2 \end{aligned}$$

$$\tilde{F}_t(\mathbf{x}) = \frac{1}{Nt} \sum_{\tau=1}^t \sum_{i=1}^N \tilde{f}_\tau^i(\mathbf{x})$$

和

$$\mathbf{x}^*(t) = \arg \min_{\mathbf{x} \in X} \tilde{F}_t(\mathbf{x})$$

因此, 对任何  $\mathbf{x}^* \in X$ , 可得

$$\sum_{t=1}^T \tilde{f}_t(\mathbf{x}^*(t)) \leq \sum_{t=1}^T \tilde{f}_t(\mathbf{x}^*) \quad (55)$$

其中,  $\tilde{f}_t = (1/N) \sum_{i=1}^N \tilde{f}_t^i$ . 由于  $\tilde{F}_t$  是  $(L/D)$ -强凸且  $\mathbf{x}^*(t) = \arg \min_{\mathbf{x} \in X} \tilde{F}_t(\mathbf{x})$ , 可得

$$\tilde{F}_t(\bar{\mathbf{x}}(t)) - \tilde{F}_t(\mathbf{x}^*(t)) \geq \frac{L}{D} \|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t)\|_2^2 \quad (56)$$

根据式 (54) 和式 (56), 可得

$$\|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t)\|_2 \leq \sqrt{\frac{D\tilde{\Delta}(t)}{L}} \leq \sqrt{\frac{\rho D}{L}} \times \sqrt{\frac{1}{t+1}} \quad (57)$$

因此, 对于所有的  $t \geq t^*$ , 可以得出

$$\tilde{f}_t(\bar{\mathbf{x}}(t)) \leq \tilde{f}_t(\mathbf{x}^*(t)) + 3\sqrt{\rho LD} \times \sqrt{\frac{1}{t+1}} \quad (58)$$



其中, 利用了  $\tilde{f}_t^i$  是  $3L$ -Lipschitze 的特性. 从  $t = 1$  到  $t = T$  对式 (58) 求和, 可得

$$\sum_{t=1}^T \left( \tilde{f}_t(\bar{\mathbf{x}}(t)) - \tilde{f}_t(\mathbf{x}^*) \right) \leq 6\sqrt{\rho LD} \times \sqrt{T+1} \quad (59)$$

其中, 不等式根据式 (55) 和以下不等式得出

$$\sum_{t=1}^T \frac{1}{\sqrt{t+1}} \leq \int_0^{T+1} \frac{1}{\sqrt{\tau}} d\tau = 2\sqrt{T+1}$$

因此, 根据式 (59), 可以得到

$$\sum_{t=1}^T \langle \nabla f_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \mathbf{x}^* \rangle \leq 6\sqrt{\rho LD} \times \sqrt{T+1} \quad (60)$$

其中, 使用了不等式

$$-\sum_{t=1}^T \left( \frac{L}{D} \right) \|\bar{\mathbf{x}}(t) - \mathbf{x}^*\|_2 \leq 0$$

依据  $f_t$  是凸函数, 可以得出

$$f_t(\bar{\mathbf{x}}(t)) - f_t(\mathbf{x}^*) \leq \langle \nabla f_t(\bar{\mathbf{x}}(t)), \bar{\mathbf{x}}(t) - \mathbf{x}^* \rangle \quad (61)$$

根据式 (60) 和式 (61), 可以得到

$$\sum_{t=1}^T f_t(\bar{\mathbf{x}}(t)) - f_t(\mathbf{x}^*) \leq 6\sqrt{\rho LD} \times \sqrt{T+1} \quad (62)$$

□

最后, 本文分析非凸函数时算法 1 的收敛性能, 即分析对偶间隙的上界, 具体结果如定理 3 所述. 下面给出定理 3 的证明.

**定理 3 证明.** 根据  $\psi(t)$  的定义, 即式 (19), 利用式 (33) 和  $\gamma(t) = 1/t^\kappa$ , 可以得到

$$\begin{aligned} \Delta(t+1) &\leq \frac{t}{t+1} (\Delta(t) - \gamma(t)\psi(t)) + \\ &\frac{t}{1+t} \left( \frac{\gamma^2(t)}{2} \beta D^2 + 2D\gamma(t) \times \frac{C_2 + \beta C_1}{t^\kappa} \right) + \\ &\frac{1}{1+t} (f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1))) \end{aligned} \quad (63)$$

其中, 不等式根据引理 2 和引理 3 得出. 因此, 根据式 (63), 可得

$$\begin{aligned} \frac{t}{t+1} \gamma(t)\psi(t) &\leq \frac{t}{t+1} \left( \Delta(t) + \frac{\gamma^2(t)}{2} \beta D^2 \right) - \\ &\Delta(t+1) + \frac{t}{t+1} \times \\ &2D\gamma(t) \times \frac{C_2 + \beta C_1}{t^\kappa} + \\ &\frac{f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1))}{1+t} \end{aligned} \quad (64)$$

另外, 根据  $\Delta(t)$  的定义, 可得

$$\begin{aligned} &\frac{(f_{t+1}(\bar{\mathbf{x}}(t+1)) - f_{t+1}(\mathbf{x}^*(t+1))) - \Delta(t+1)}{1+t} = \\ &-\frac{t}{t+1} (F_t(\bar{\mathbf{x}}(t)) - F_t(\mathbf{x}^*(t+1))) \end{aligned} \quad (65)$$

结合式 (64) 和式 (65), 可以得出

$$\begin{aligned} \gamma(t)\psi(t) &\leq \Delta(t) + \frac{\gamma^2(t)}{2} \beta D^2 - \\ &(F_t(\bar{\mathbf{x}}(t+1)) - F_t(\mathbf{x}^*(t+1))) + \\ &2D\gamma(t) \times \frac{C_2 + \beta C_1}{t^\kappa} \end{aligned} \quad (66)$$

根据  $F_t$  的定义, 还可以得到

$$\begin{aligned} F_t(\bar{\mathbf{x}}(t)) - F_{t-1}(\bar{\mathbf{x}}(t)) &= \\ \frac{1}{t} \sum_{s=1}^{t-1} f_s(\bar{\mathbf{x}}(t)) + \frac{1}{t} f_t(\bar{\mathbf{x}}(t)) - F_{t-1}(\bar{\mathbf{x}}(t)) &= \\ \frac{t-1}{t} F_{t-1}(\bar{\mathbf{x}}(t)) + \frac{1}{t} f_t(\bar{\mathbf{x}}(t)) - F_{t-1}(\bar{\mathbf{x}}(t)) &= \\ \frac{1}{t} (f_t(\bar{\mathbf{x}}(t)) - F_{t-1}(\bar{\mathbf{x}}(t))) \end{aligned} \quad (67)$$

这样, 根据式 (67), 可以得到

$$\begin{aligned} &\sum_{t=T/2+1}^T (\Delta(t) - (F_t(\bar{\mathbf{x}}(t+1)) - F_t(\mathbf{x}^*(t+1)))) = \\ &\sum_{t=T/2+1}^T (F_t(\bar{\mathbf{x}}(t)) - F_t(\bar{\mathbf{x}}(t+1))) - \\ &\sum_{t=T/2+1}^T (F_t(\mathbf{x}^*(t)) - F_t(\mathbf{x}^*(t+1))) = \\ &F_{T/2+1}(\bar{\mathbf{x}}(T/2+1)) - F_T(\bar{\mathbf{x}}(T+1)) + \\ &F_T(\mathbf{x}^*(T+1)) - F_{T/2+1}(\mathbf{x}^*(T/2+1)) + \\ &\sum_{t=T/2+2}^T \frac{1}{t} \times (f_t(\bar{\mathbf{x}}(t)) - F_{t-1}(\bar{\mathbf{x}}(t))) - \\ &\sum_{t=T/2+2}^T \frac{1}{t} \times (f_t(\mathbf{x}^*(t)) - F_{t-1}(\mathbf{x}^*(t))) + \\ &L \times \|\bar{\mathbf{x}}(T/2+1) - \mathbf{x}^*(T/2+1)\|_2 + \\ &L \times \sum_{t=T/2+2}^T \frac{2}{t} \|\bar{\mathbf{x}}(t) - \mathbf{x}^*(t)\|_2 \leq \\ &2LD \times \left( 1 + \sum_{t=T/2+2}^T \frac{1}{t} \right) \leq \\ &2LD \times (1 + \log 2) \leq 4LD \end{aligned} \quad (68)$$

其中, 第 1 个不等式根据  $F_t$  和  $f_t$  是  $L$ -Lipschitz 得出, 第 2 个不等式由  $D$  的定义得到, 第 3 个和第 4 个不等式依据  $\sum_{t=T/2+1}^T t^{-1} \leq \log 2T/(T+4) \leq \log 2 \leq 1$  获得.

从  $t = T/2 + 1$  到  $t = T$  对式 (66) 求和, 可得

$$\begin{aligned} \sum_{t=T/2+1}^T \gamma(t) \psi(t) &\leq 4LD + \\ &\sum_{t=T/2+1}^T \left( \frac{\gamma^2(t)}{2} \beta D^2 + 2D\gamma(t) \times \frac{C_2 + \beta C_1}{t^\kappa} \right) \end{aligned} \quad (69)$$

由于  $\gamma(t) = t^{-\kappa}$ , 则当  $\kappa \in [1/2, 1)$ , 可以得到

$$\sum_{t=T/2+1}^T \gamma^2(t) = \sum_{t=T/2+1}^T t^{-2\kappa} \leq \sum_{t=T/2+1}^T t^{-1} \leq \log 2 \quad (70)$$

由  $\psi(t) \geq 0$  和  $\gamma(t) \geq 0$ , 还可以得到

$$\begin{aligned} \left[ \min_{t \in [T/2+1, T]} \psi(t) \right] \times \left[ \sum_{t=T/2}^T \gamma(t) \right] &\leq \\ \sum_{t=T/2}^T \gamma(t) \psi(t) &\leq \\ 4LD + \left( \beta \frac{D^2}{2} + 2\beta DC_1 + 2DC_2 \right) \log 2 \end{aligned} \quad (71)$$

因此, 对于所有的  $T \geq 6$ , 可得

$$\begin{aligned} \sum_{t=T/2+1}^T \gamma(t) &= \sum_{t=T/2+1}^T \frac{1}{t^\kappa} \geq \int_{T/2+1}^T t^{-\kappa} dt = \\ \frac{T^{1-\kappa}}{1-\kappa} \left( 1 - \left( \frac{1}{2} + \frac{1}{T} \right)^{1-\kappa} \right) &\geq \\ \frac{T^{1-\kappa}}{1-\kappa} \left( 1 - \left( \frac{2}{3} \right)^{1-\kappa} \right) \end{aligned} \quad (72)$$

将式 (72) 代入式 (71), 则对所有的  $T \geq 6$ , 有

$$\min_{t \in [T/2+1, T]} \psi(t) \leq C_4 \left( 1 - \left( \frac{2}{3} \right)^{1-\kappa} \right)^{-1} \times \frac{1-\kappa}{T^{1-\kappa}} \quad (73)$$

其中,

$$C_4 = 4LD + \left( \frac{\beta D^2}{2} + 2\beta DC_1 + 2DC_2 \right) \times \log 2$$

当  $\kappa \in (0, 1/2)$  时, 有  $t^{-2\kappa} > t^{-1}$ . 因此, 可以得出

$$\sum_{t=T/2+1}^T \frac{1}{t^{2\kappa}} \leq \int_{t=T/2}^T \frac{1}{t^{2\kappa}} dt = \frac{1 - \left( \frac{1}{2} \right)^{1-2\kappa}}{1-2\kappa} \times T^{1-2\kappa} \quad (74)$$

利用式 (74), 可得

$$\begin{aligned} 4LD + \sum_{t=T/2+1}^T \left( \frac{\gamma^2(t)}{2} \beta D^2 + 2D\gamma(t) \times \right. \\ \left. \frac{C_2 + \beta C_1}{t^\kappa} \right) &\leq \left( \frac{\beta D^2}{2} + 2DC_2 + 2\beta DC_1 \right) \times \\ \frac{1 - \left( \frac{1}{2} \right)^{1-2\kappa}}{1-2\kappa} \times T^{1-2\kappa} &+ 4LD \times T^{1-2\kappa} \end{aligned}$$

将式 (75) 代入式 (69), 并除以  $(1-\kappa)^{-1}(1-(2/3)^{1-\kappa})T^{1-\kappa}$ , 可得

$$\min_{t \in [T/2+1, T]} \psi(t) \leq \frac{1-\kappa}{1 - \left( \frac{2}{3} \right)^{1-\kappa}} \times \frac{C_5}{T^\kappa} \quad (75)$$

其中,

$$\begin{aligned} C_5 = 4LD + \left( \frac{\beta D^2}{2} + 2DC_2 + 2\beta DC_1 \right) \times \\ \frac{1 - \left( \frac{1}{2} \right)^{1-2\kappa}}{1-2\kappa} \end{aligned}$$

□

## 5 仿真实验

仿真实验考虑一个数据平均分布的多智能体网络, 每个节点通过局部通信和局部计算, 与邻居节点协作完成网络中的全局任务, 使用算法 1 解决网络系统的多分类问题.

在多分类问题中, 每个局部代价函数为

$$\begin{aligned} f_i^i(X_i(t)) = \\ \log \left( 1 + \sum_{\ell \neq y_i(t)} \exp \left( \mathbf{x}_\ell^T \mathbf{e}_i(t) - \mathbf{x}_{y_i(t)}^T \mathbf{e}_i(t) \right) \right) \end{aligned}$$

其中,  $\mathbf{e}_i(t) \in \mathbf{R}^d$  是类型  $C = \{1, \dots, c\}$  的一个元素, 表示节点  $i$  的一个数据实例;  $\mathbf{X}_i(t) = [\mathbf{x}_1^T, \dots, \mathbf{x}_c^T] \in \mathbf{R}^{c \times d}$  表示节点  $i$  的一个决策矩阵; 定义  $X = \{X \in \mathbf{R}^{c \times d} \mid \|X\|_* \leq \zeta\}$ , 其中  $\|X\|_*$  是  $X$  的核范数; 学习率设为  $2/(t+2)$ .

本实验在 aloi 和 news20 数据集上验证了算法 1 的性能. 实验 1 考察节点数对算法 1 的影响, 当节点分别为 4、64、128 和 256 时算法 1 的性能如图 1 所示. 从图 1 可以看出, 悔界随着节点数量的增加而增加, 且在不同节点数时算法 1 都是收敛的. 实验 2 采用 64 个节点比较算法 1 与 D-OCG<sup>[33]</sup> 算法的性能, 结果如图 2 所示. 从图 2 可以看出, 算法 1 均比 D-OCG 算法具有更快的收敛速率. 实验 3 考

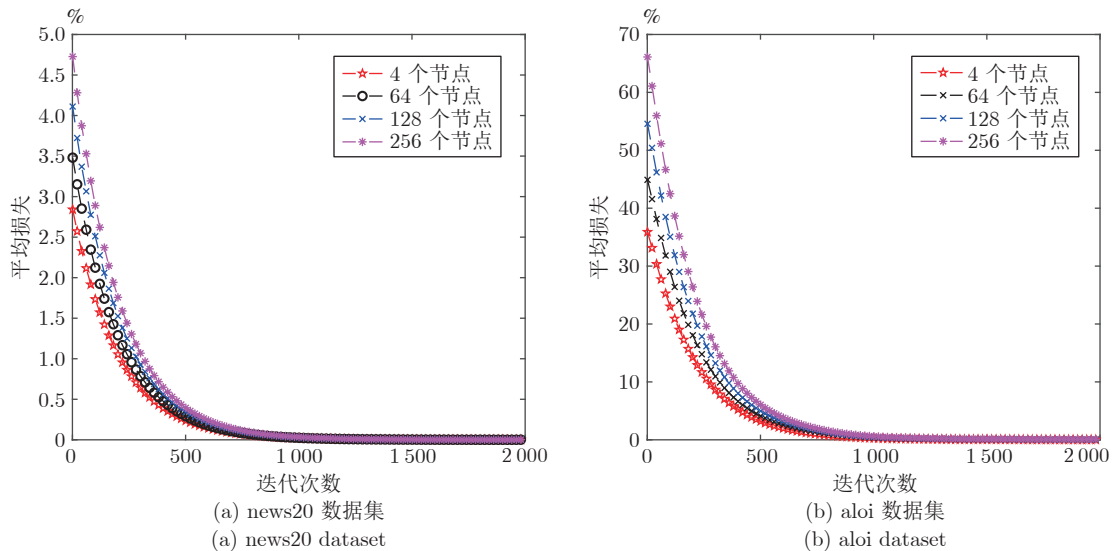


图 1 在 news20 和 aloi 数据集上不同节点数下本文算法的性能

Fig. 1 Performance of the proposed algorithm at different nodes on news20 and aloi datasets

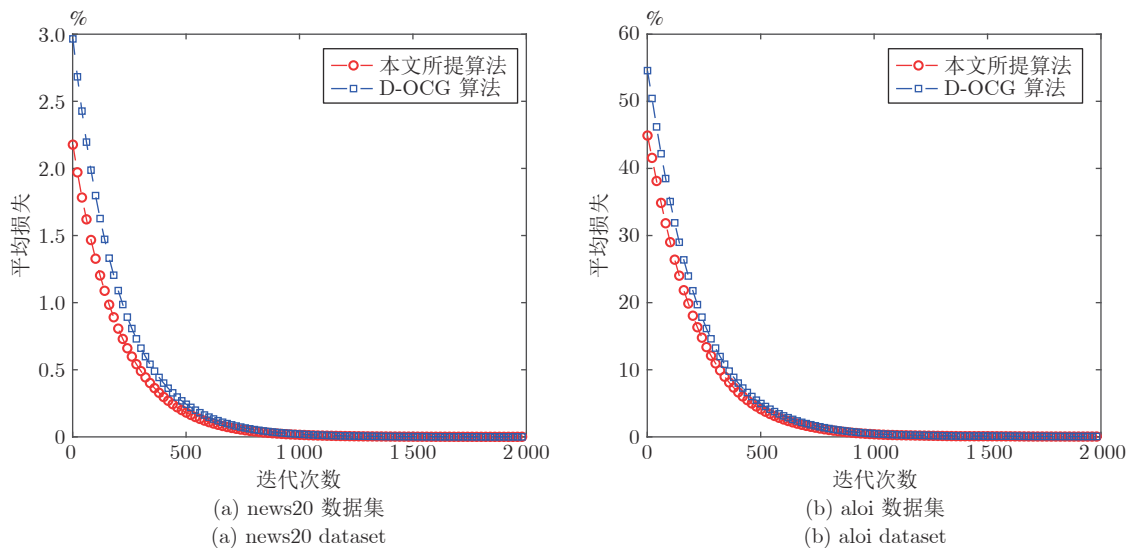


图 2 在 news20 和 aloi 数据集上 64 节点下本文算法和 D-OCG 算法的性能比较

Fig. 2 The performance comparison between the proposed algorithm and the D-OCG algorithm at 64 nodes on news20 and aloi datasets

察不同网络拓扑对算法 1 性能的影响, 采用 64 个节点组成完全图、随机图和循环图三种网络拓扑, 算法 1 性能如图 3 所示. 从图 3 可以看出, 采用完全图拓扑时算法的收敛速率略快于采用随机图<sup>[37]</sup>和循环图拓扑时算法的收敛速率.

## 6 结束语

本文研究了多智能体网络系统中的分布式在线约束优化问题, 其中全局代价函数是局部代价函数之和, 且局部代价函数是时变的. 为了解决这个问题, 本文提出了一种分布式条件梯度在线学习算法,

以避免代价高昂的投影算子. 理论分析表明, 对于凸代价函数, 所提算法可以达到悔界  $O(\sqrt{T})$ , 其中  $T$  表示时间范围; 对于非凸代价函数, 所提算法以  $O(\sqrt{T})$  的速率收敛到平稳点. 仿真实验验证了所提算法的性能和理论分析的结果. 该算法可为多智能体网络系统的资源分配、数据管理、调度控制等问题提供参考.

## 附录 A 引理 1 的证明

**证明.** 1) 关系 a) 的证明. 对式 (9) 从  $i = 1, \dots, N$  求和, 其中  $t \geq 1$ , 除以  $N(t+1)$  得到

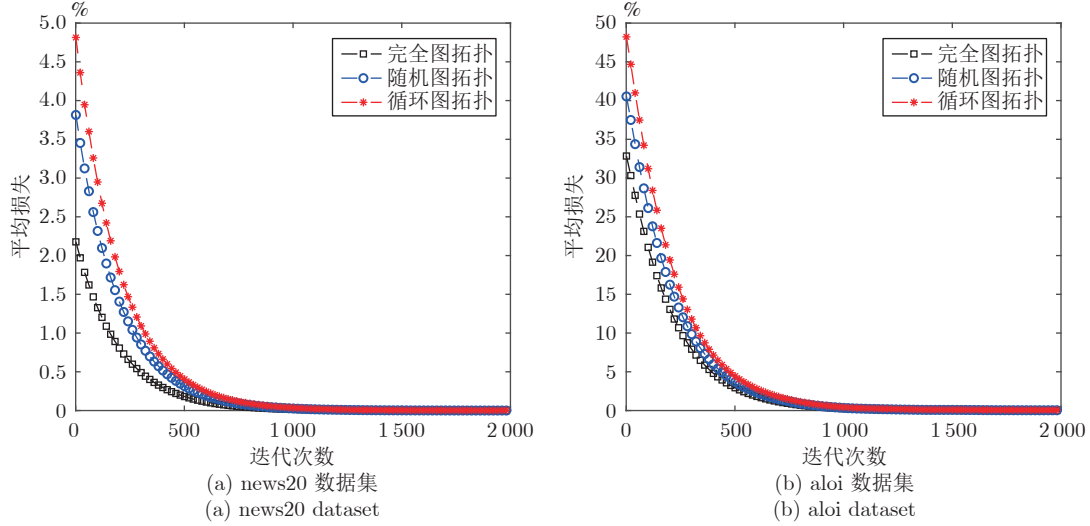


图 3 本文算法在具有固定 64 个节点和不同拓扑结构的 news20 和 aloi 数据集上的性能

Fig. 3 Performance of the proposed algorithm on news20 and aloi datasets with fixed 64 nodes and different topologies

$$\begin{aligned}
 \bar{\mathbf{s}}_{t+1}(t+1) - \frac{1}{t+1} \times \frac{1}{N} \sum_{\tau=1}^{t+1} \sum_{i=1}^N \mathbf{s}_{\tau}^i(t+1) = & \\
 \frac{1}{1+t} \times \frac{1}{N} \sum_{\tau=1}^{t+1} \sum_{i=1}^N \sum_{j \in N_i} a_{ij} \mathbf{s}_{\tau}^j(t) + & \\
 \frac{1}{t+1} \times \frac{1}{N} \sum_{\tau=1}^{t+1} \sum_{i=1}^N \nabla f_{\tau}^i(\mathbf{z}_i(t+1)) - & \\
 \frac{1}{t+1} \times \frac{1}{N} \sum_{\tau=1}^{t+1} \sum_{i=1}^N \nabla f_{\tau}^i(\mathbf{z}_i(t)) = & \\
 \bar{\mathbf{s}}_{t+1}(t) + \mathbf{g}_{t+1}(t+1) - \mathbf{g}_{t+1}(t) & \quad (\text{A1})
 \end{aligned}$$

由于邻接矩阵  $\mathbf{A}$  是一个双随机矩阵, 即  $\mathbf{1}^T \mathbf{A} = \mathbf{1}^T$ . 由式 (A1) 可得

$$\bar{\mathbf{s}}_{t+1}(t+1) = \bar{\mathbf{s}}_{t+1}(0) + \mathbf{g}_{t+1}(t+1) - \mathbf{g}_{t+1}(0)$$

其中,  $\mathbf{g}_{t+1}(0) = \frac{1}{N(t+1)} \sum_{\tau=1}^{t+1} \sum_{i=1}^N \nabla f_{\tau}^i(\mathbf{z}_i(0))$ . 由于对所有的  $i=1, \dots, N$  和  $\tau \geq 1$ , 有  $\mathbf{s}_{\tau}^i(0) = \nabla f_{\tau}^i(\mathbf{z}_i(0))$ , 则  $\bar{\mathbf{s}}_{t+1}(0) = \mathbf{g}_{t+1}(0)$ .

2) 关系 b) 的证明. 根据  $\bar{\mathbf{x}}(t)$  的定义以及式 (8) 和式 (12), 可以得出

$$\begin{aligned}
 \bar{\mathbf{x}}(t+1) &= \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i(t+1) = \\
 \frac{1}{N} \sum_{i=1}^N [(1-\gamma(t)) \mathbf{z}_i(t) + \gamma(t) \mathbf{v}_i(t)] &= \\
 \frac{(1-\gamma(t))}{N} \sum_{i=1}^N \sum_{j=1}^N a_{ij} \mathbf{x}_j(t) + \frac{\gamma(t)}{N} \sum_{i=1}^N \mathbf{v}_i(t) &=
 \end{aligned}$$

$$(1-\gamma(t)) \bar{\mathbf{x}}(t) + \gamma(t) \bar{\mathbf{v}}(t) \quad (\text{A2})$$

其中, 最后一个等式根据矩阵  $\mathbf{A}$  由双随机矩阵得出, 即  $\mathbf{1}^T \mathbf{A} = \mathbf{1}^T$ , 且  $\bar{\mathbf{v}}(t) = (1/N) \sum_{i=1}^N \mathbf{v}_i(t)$ .  $\square$

## 附录 B 引理 2 的证明

证明. 根据范数的性质, 有

$$\max_{i \in \{1, \dots, N\}} \|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2 \leq \sqrt{\sum_{i=1}^N \|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2^2} \quad (\text{B1})$$

为了证明引理 2, 只需要证明不等式

$$\sqrt{\sum_{i=1}^N \|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2^2} \leq \frac{C_1}{t^{\kappa}} \quad (\text{B2})$$

下面对  $t$  使用归纳法证明式 (B2). 可以看出, 从  $t=1$  到  $t=t_0$ , 式 (B2) 成立. 假设对于某些  $t \geq t_0$ , 式 (B2) 也成立. 由于当  $\kappa \in (0, 1]$ , 有  $\gamma(t) = t^{-\kappa}$ , 则有  $\mathbf{x}_i(t+1) = (1-t^{-\kappa}) \mathbf{z}_i(t) + t^{-\kappa} \mathbf{v}_i(t)$ . 因此, 根据引理 1 的 b), 可以得到

$$\begin{aligned}
 \sum_{i=1}^N \|\mathbf{z}_i(t+1) - \bar{\mathbf{x}}(t+1)\|_2^2 &= \\
 \sum_{i=1}^N \left\| \sum_{j=1}^N a_{ij} (1-t^{-\kappa}) \mathbf{z}_j(t) + \right. & \\
 \left. \sum_{j=1}^N a_{ij} t^{-\kappa} \mathbf{v}_j(t) - (1-t^{-\kappa}) \bar{\mathbf{x}}(t) - t^{-\kappa} \bar{\mathbf{v}}(t) \right\|_2^2 &\leq
 \end{aligned}$$

$$\lambda^2 \sum_{j=1}^N \left\| (1-t^{-\kappa})(z_j(t) - \bar{x}(t)) + t^{-\kappa}(v_j(t) - \bar{v}(t)) \right\|_2^2 \tag{B3}$$

其中, 最后一个不等式依据式 (13) 得到. 为了获得  $\sum_{i=1}^N \|z_i(t+1) - \bar{x}(t+1)\|_2^2$  的上界, 需要得到

$$\sum_{j=1}^N \left\| (1-t^{-\kappa})(z_j(t) - \bar{x}(t)) + t^{-\kappa}(v_j(t) - \bar{v}(t)) \right\|_2^2$$

的上界. 为此, 首先有

$$\sum_{j=1}^N \left\| (1-t^{-\kappa})(z_j(t) - \bar{x}(t)) + t^{-\kappa}(v_j(t) - \bar{v}(t)) \right\|_2^2 \leq$$

$$\begin{aligned} & \sum_{j=1}^N \|z_j(t) - \bar{x}(t)\|_2^2 + Nt^{-2\kappa}D^2 + \\ & 2Dt^{-\kappa}\sqrt{N} \sqrt{\sum_{j=1}^N \|z_j(t) - \bar{x}(t)\|_2^2} \leq \\ & t^{-2\kappa}(C_1^2 + ND^2) + 2Dt^{-2\kappa}C_1\sqrt{N} \leq \\ & \left( \frac{t_0^\kappa + 1}{t_0^\kappa} \times \frac{C_1}{t^\kappa} \right)^2 \end{aligned} \tag{B4}$$

其中, 第 1 个不等式依据柯西-施瓦兹不等式、约束集  $X$  的有界性和不等式  $\sum_{i=1}^N |w_i| \leq \sqrt{N} \sqrt{\sum_{i=1}^N w_i^2}$  得到. 第 2 个和第 3 个不等式由归纳假设得到. 因此, 根据式 (14), 对于所有的  $t \geq t_0$ , 有

$$\begin{aligned} \lambda \times \frac{1+t_0^\kappa}{t_0^\kappa \times t^\kappa} & \leq \frac{t_0^\kappa}{1+t_0^\kappa} \times \left( \frac{t_0}{1+t_0} \right)^\kappa \times \\ & \frac{1+t_0^\kappa}{t_0^\kappa \times t^\kappa} = \left( \frac{t_0}{1+t_0} \right)^\kappa \times \frac{1}{t^\kappa} \leq \\ & \left( \frac{t}{1+t} \right)^\kappa \times \frac{1}{t^\kappa} = \frac{1}{(1+t)^\kappa} \end{aligned} \tag{B5}$$

其中, 第 2 个不等式根据函数  $\varphi(\nu) = (\nu/(1+\nu))^\kappa$  是关于  $\nu$  的单调递增函数得到. 结合式 (14) 和式 (B5), 可以得到

$$\sqrt{\sum_{i=1}^N \|z_i(t+1) - \bar{x}(t+1)\|_2^2} \leq \frac{C_1}{(1+t)^\kappa}$$

因此, 式 (B2) 成立.  $\square$

### 附录 C 引理 3 的证明

证明. 首先, 利用范数的性质, 有

$$\max_{i \in \{1, \dots, N\}} \|S_t^i(t) - g_t(t)\|_2 \leq \sqrt{\sum_{i=1}^N \|S_t^i(t) - g_t(t)\|_2^2} \tag{C1}$$

为了证明式 (25), 只需证明式 (C2)

$$\sqrt{\sum_{i=1}^N \|S_t^i(t) - g_t(t)\|_2^2} \leq \frac{C_2}{t^\kappa} \tag{C2}$$

其中,  $C_2 = \sqrt{N}[\delta + \frac{\beta(D+2C_1)}{1-\kappa}] + 2\beta(D+2C_1) \times t_0^\kappa$ , 且对于所有的  $i = 1, \dots, N$  和  $t \geq 1$ ,  $\delta$  是一个满足  $\delta \geq \|s_{t+1}^i(1) - \phi_{t+1}(1)\|_2$  的正常数.

下面采用归纳法证明式 (C2). 可以看出, 从  $t = 1$  到  $t = t_0$ , 式 (C2) 成立. 假设对于某个  $t$ , 式 (C2) 成立. 引入一个辅助变量  $\sigma f_\tau^i(t+1) = \nabla f_\tau^i(z_i(t+1)) - \nabla f_\tau^i(z_i(t))$ , 将其代入式 (9), 可以得到

$$s_\tau^i(t+1) = \sum_{j \in N_i} a_{ij} s_\tau^j(t) + \sigma f_\tau^i(t+1) \tag{C3}$$

根据  $S_t^i(t)$  的定义, 有

$$\begin{aligned} & \sum_{i=1}^N \|S_{t+1}^i(t+1) - g_{t+1}(t+1)\|_2^2 \leq \\ & \lambda^2 \sum_{i=1}^N \left\| \frac{1}{t+1} \sum_{\tau=1}^{t+1} \sum_{j=1}^N a_{ij} s_\tau^j(t) + \right. \\ & \left. \frac{1}{t+1} \sum_{\tau=1}^{t+1} \sigma f_\tau^i(t+1) - g_{t+1}(t+1) \right\|_2^2 = \\ & \lambda^2 \sum_{i=1}^N \left\| \frac{t}{t+1} (S_t^i(t) - g_t(t)) + \right. \\ & \left. \frac{1}{t+1} s_{t+1}^i(t+1) + \frac{1}{t+1} \sum_{\tau=1}^t \sigma f_\tau^i(t+1) + \right. \\ & \left. \frac{t}{t+1} g_t(t) - g_{t+1}(t+1) \right\|_2^2 \end{aligned} \tag{C4}$$

其中, 不等式依据式 (9) 和式 (13) 得出. 根据  $g_t(t)$  的定义, 可以得到

$$\begin{aligned} & g_{t+1}(t+1) - \frac{t}{t+1} g_t(t) = \\ & \frac{1}{t+1} \frac{1}{N} \sum_{\tau=1}^{t+1} \sum_{i=1}^N \nabla f_\tau^t(z_i(t+1)) - \\ & \frac{1}{t+1} \frac{1}{N} \sum_{\tau=1}^t \sum_{i=1}^N \nabla f_\tau^t(z_i(t)) = \end{aligned}$$

$$\begin{aligned} & \frac{1}{N(t+1)} \sum_{\tau=1}^t \sum_{i=1}^N (\nabla f_{\tau}^i(\mathbf{z}_i(t+1)) - \nabla f_{\tau}^i(\mathbf{z}_i(t))) + \\ & \frac{1}{t+1} \frac{1}{N} \sum_{i=1}^N \nabla f_{t+1}^i(\mathbf{z}_i(t+1)) = \\ & \frac{1}{t+1} \sum_{\tau=1}^t \left( \frac{1}{N} \sum_{i=1}^N \sigma f_{\tau}^i(t+1) \right) + \\ & \frac{1}{t+1} \phi_{t+1}(t+1) \end{aligned} \quad (C5)$$

其中,  $\phi_{t+1}(t+1) = (1/N) \sum_{i=1}^N \nabla f_{t+1}^i(\mathbf{z}_i(t+1))$ .  
并且可以进一步得到

$$\begin{aligned} & \|\mathbf{s}_{t+1}^i(t+1) - \phi_{t+1}(t+1)\|_2 = \\ & \left\| \sum_{j=1}^N a_{ij} \mathbf{s}_{t+1}^j(t) - \phi_{t+1}(t) + \right. \\ & \left. \sigma f_{t+1}^i(t+1) - [\phi_{t+1}(t+1) - \phi_{t+1}(t)] \right\|_2 \leq \\ & \left\| \sum_{j=1}^N a_{ij} \mathbf{s}_{t+1}^j(t) - \phi_{t+1}(t) \right\|_2 + \\ & \|\sigma f_{t+1}^i(t+1) - [\phi_{t+1}(t+1) - \phi_{t+1}(t)]\|_2 \leq \\ & \lambda \|\mathbf{s}_{t+1}^i(t) - \phi_{t+1}(t)\|_2 + \\ & \|\sigma f_{t+1}^i(t+1) - [\phi_{t+1}(t+1) - \phi_{t+1}(t)]\|_2 \end{aligned} \quad (C6)$$

其中, 第 1 个不等式依据三角不等式得出, 最后一个不等式根据式 (13) 获得. 另外,  $\|\sigma f_{t+1}^i(t+1) - [\phi_{t+1}(t+1) - \phi_{t+1}(t)]\|_2$  由式 (C7) 进行计算

$$\begin{aligned} & \|\sigma f_{t+1}^i(t+1) - [\phi_{t+1}(t+1) - \phi_{t+1}(t)]\|_2^2 = \\ & \|\sigma f_{t+1}^i(t+1)\|_2^2 + \|\phi_{t+1}(t+1) - \phi_{t+1}(t)\|_2^2 - \\ & 2 \langle \sigma f_{t+1}^i(t+1), \phi_{t+1}(t+1) - \phi_{t+1}(t) \rangle \leq \\ & \|\sigma f_{t+1}^i(t+1)\|_2^2 \end{aligned} \quad (C7)$$

其中, 最后一个不等式依据  $\phi_{t+1}(t+1)$  的定义得出. 将式 (C7) 代入式 (C6), 得到

$$\begin{aligned} & \|\mathbf{s}_{t+1}^i(t+1) - \phi_{t+1}(t+1)\|_2 \leq \\ & \lambda \|\mathbf{s}_{t+1}^i(t) - \phi_{t+1}(t)\|_2 + \\ & \|\sigma f_{t+1}^i(t+1)\|_2 \end{aligned} \quad (C8)$$

根据  $\sigma f_{t+1}^i(t+1)$  的定义, 可以得到

$$\begin{aligned} & \|\sigma f_{t+1}^i(t+1)\|_2 = \\ & \|\nabla f_{t+1}^i(\mathbf{z}_i(t+1)) - \nabla f_{t+1}^i(\mathbf{z}_i(t))\|_2 \leq \\ & \beta \|\mathbf{z}_i(t+1) - \mathbf{z}_i(t)\|_2 = \\ & \beta \left\| \sum_{j=1}^N a_{ij} (\mathbf{x}_j(t+1) - \mathbf{z}_j(t) + \right. \\ & \left. (\mathbf{z}_j(t) - \mathbf{z}_i(t))) \right\|_2 \leq \\ & \beta \sum_{j=1}^N a_{ij} (\|\mathbf{x}_j(t+1) - \mathbf{z}_j(t)\|_2 + \\ & \|\mathbf{z}_j(t) - \mathbf{z}_i(t)\|_2) \leq \\ & \beta \sum_{j=1}^N a_{ij} \|t^{-\kappa} (\mathbf{v}_j(t) - \mathbf{z}_j(t))\|_2 + \\ & \beta \sum_{j=1}^N a_{ij} (\|\mathbf{z}_j(t) - \bar{\mathbf{x}}(t)\|_2 + \|\mathbf{z}_i(t) - \bar{\mathbf{x}}(t)\|_2) \leq \\ & \beta \sum_{j=1}^N a_{ij} (Dt^{-\kappa} + 2C_1 t^{-\kappa}) = \\ & (D + 2C_1) \beta t^{-\kappa} \end{aligned} \quad (C9)$$

其中, 第 1 个不等式根据式 (4) 得到, 第 2 个不等式可由欧几里得范数的凸性和三角不等式推出, 第 3 个不等式依据式 (12) 和三角不等式获得, 第 4 个不等式从引理 1 推出. 将式 (C9) 与 (C8) 合并, 可以得到

$$\begin{aligned} & \|\mathbf{s}_{t+1}^i(t+1) - \phi_{t+1}(t+1)\|_2 \leq \\ & \lambda^{t+1} \|\mathbf{s}_{t+1}^i(1) - \phi_{t+1}(1)\|_2 + \\ & \beta (D + 2C_1) \sum_{\tau=0}^{t-1} \lambda^{\tau} (t - \tau)^{-\kappa} \leq \\ & \|\mathbf{s}_{t+1}^i(1) - \phi_{t+1}(1)\|_2 + \\ & \frac{\beta (D + 2C_1)}{1 - \kappa} t^{1-\kappa} \end{aligned} \quad (C10)$$

其中, 最后一个不等式根据下面不等式得出

$$\begin{aligned} & \sum_{\tau=1}^{t-1} \lambda^{\tau} (t - \tau)^{-\kappa} \leq \sum_{\theta=1}^t \theta^{-\kappa} \leq \\ & \int_0^t \theta^{-\kappa} d\theta = \frac{1}{1 - \kappa} t^{1-\kappa} \end{aligned}$$

其中,  $\lambda \in (0, 1)$ . 此外, 还可以得到

$$\begin{aligned}
& \left\| \frac{1}{t+1} \sum_{\tau=1}^t \sigma f_{\tau}^i(t+1) - \right. \\
& \quad \left. \frac{1}{t+1} \sum_{\tau=1}^t \left( \frac{1}{N} \sum_{i=1}^N \sigma f_{\tau}^i(t+1) \right) \right\|_2 \leq \\
& \frac{1}{t+1} \sum_{\tau=1}^t \left\| \sigma f_{\tau}^i(t+1) - \frac{1}{N} \sum_{i=1}^N \sigma f_{\tau}^i(t+1) \right\|_2 = \\
& \frac{1}{t+1} \sum_{\tau=1}^t \left\| \left( 1 - \frac{1}{N} \right) \sigma f_{\tau}^i(t+1) - \right. \\
& \quad \left. \frac{1}{N} \sum_{j \neq i} \sigma f_{\tau}^j(t+1) \right\|_2 \leq \\
& \frac{1}{t+1} \sum_{\tau=1}^t \left( \left( 1 - \frac{1}{N} \right) \left\| \sigma f_{\tau}^i(t+1) \right\|_2 + \right. \\
& \quad \left. \frac{1}{N} \sum_{j \neq i} \left\| \sigma f_{\tau}^j \right\|_2 \right) \leq \\
& 2 \frac{1}{t+1} \sum_{\tau=1}^t \left( 1 - \frac{1}{N} \right) (D + 2C_1) \beta t^{-\kappa} \leq \\
& 2(D + 2C_1) \beta t^{-\kappa} \times \frac{t}{t+1} \tag{C11}
\end{aligned}$$

其中, 第 1 个不等式根据欧几里得范数的凸性推出, 第 2 个不等式依据三角不等式获得, 第 3 个不等式从式 (C9) 推出, 第 4 个不等式依据  $N$  是一个正整数得到. 将式 (C10) 和式 (C11) 代入式 (C4), 并使  $\|s_{t+1}^i(1) - \phi_{t+1}(1)\|_2 \leq \delta$ ,  $\delta$  是一个正常数. 则对于所有的  $i \in \{1, \dots, N\}$ , 可以得到

$$\begin{aligned}
& \sum_{i=1}^N \left\| \frac{t}{t+1} (S_t^i(t) - g_t(t)) + \frac{1}{t+1} s_{t+1}^i(t+1) + \right. \\
& \quad \left. \frac{1}{t+1} \sum_{\tau=1}^t \sigma f_{\tau}^i(t+1) + \frac{t}{t+1} g_t(t) - \right. \\
& \quad \left. g_{t+1}(t+1) \right\|_2^2 \leq \\
& \sum_{i=1}^N \left\| S_t^i(t) - g_t(t) \right\|_2^2 + 4N(D + 2C_1)^2 \beta^2 t^{-2\kappa} + \\
& 4 \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right] (D + 2C_1) \beta t^{-2\kappa} + \\
& \frac{N}{t^{2\kappa}} \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right]^2 +
\end{aligned}$$

$$\begin{aligned}
& 2\sqrt{N} \sqrt{\sum_{i=1}^N \|S_t^i(t) - g_t(t)\|_2^2} \times \\
& \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} + 2(D + 2C_1)\beta \right] \leq \\
& C_2^2 t^{-2\kappa} + N \left( \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right)^2 t^{-2\kappa} + \\
& 4N(D + 2C_1)^2 \beta^2 t^{-2\kappa} + 4\sqrt{N}(D + 2C_1)C_2 \beta t^{-2\kappa} + \\
& 2\sqrt{N} \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right] [2(D + 2C_1)\beta + C_2] t^{-2\kappa} \leq \\
& t^{-2\kappa} \left[ C_2 + \sqrt{N} \left[ \delta + \frac{\beta(D + 2C_1)}{1 - \kappa} \right] + \right. \\
& \quad \left. 2\beta(D + 2C_1) \right]^2 \leq \left( \frac{t_0^{\kappa} + 1}{t_0^{\kappa}} \times \frac{C_2}{t^{\kappa}} \right)^2 \tag{C12}
\end{aligned}$$

其中, 第 1 个不等式依据三角不等式、柯西-施瓦兹不等式和不等式  $\sum_{i=1}^N |w_i| \leq \sqrt{N} \sqrt{\sum_{i=1}^N w_i^2}$  推出, 第 2 个不等式根据归纳假设获得, 最后一个不等式从  $C_2$  的定义得到. 依据式 (14), 对于所有的  $t \geq t_0$ , 可以得出

$$\lambda \times \frac{t_0^{\kappa} + 1}{t_0^{\kappa}} \leq \left( \frac{t}{t+1} \right)^{\kappa} \tag{C13}$$

由此, 对式 (C12) 两边取平方根完成归纳. 然后, 将式 (C1) 和式 (C2) 结合, 即完成引理 3 的证明.  $\square$

## References

- Zhu J L, Xu C Q, Guan J F, Wu D P. Differentially private distributed online algorithms over time-varying directed networks. *IEEE Transactions on Signal and Information Processing over Networks*, 2018, 4(1): 4-17
- Rabbat M, Nowak R. Distributed optimization in sensor networks. In: Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks Proc. Berkeley, USA: IEEE, 2004. 20-27
- Zhang M C, Hao B W, Ge Q B, Zhu J L, Zheng R J, Wu Q T. Distributed adaptive subgradient algorithms for online learning over time-varying networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, 52(7): 4518-4529
- Abu-Elkheir M, Hayajneh M, Abu A N. Data management for the internet of things: Design primitives and solution. *Sensors*, 2013, 13(11): 15582-15612
- Sun Lu-Ming, Zhang Shao-Min, Ji Tao, Li Cui-Ping, Chen Hong. Survey of data management techniques powered by artificial intelligence. *Journal of Software*, 2020, 31(3): 600-619 (孙路明, 张少敏, 姬涛, 李翠平, 陈红. 人工智能赋能的数据管理技术研究. *软件学报*, 2020, 31(3): 600-619)

- 6 Huang Gang, Li Jun-Hua. Multi-UAV cooperative target allocation based on AC-DSDE evolutionary algorithm. *Acta Automatica Sinica*, 2021, **47**(1): 173–184  
(黄刚, 李军华. 基于 AC-DSDE 进化算法多 UAVs 协同目标分配. *自动化学报*, 2021, **47**(1): 173–184)
- 7 Beck A, Nedić A, Ozdaglar A, Teboulle M. An  $O(1/k)$  gradient method for network resource allocation problems. *IEEE Transactions on Control of Network Systems*, 2014, **1**(1): 64–73
- 8 Liu Mei-Qin, Han Xue-Yan, Zhang Sen-Lin, Zheng Rong-Hao, Lan Jian. Research status and prospect of target tracking technologies via underwater sensor networks. *Acta Automatica Sinica*, 2021, **47**(2): 235–251  
(刘妹琴, 韩学艳, 张森林, 郑荣濠, 兰剑. 基于水下传感器网络的目标跟踪技术研究现状与展望. *自动化学报*, 2021, **47**(2): 235–251)
- 9 Jiang Hong-Yi, Wang Yong-Juan, Kang Jin-Yu. A survey of object detection models and its optimization methods. *Acta Automatica Sinica*, 2021, **47**(6): 1232–1255  
(蒋弘毅, 王永娟, 康锦煜. 目标检测模型及其优化方法综述. *自动化学报*, 2021, **47**(6): 1232–1255)
- 10 Wen G H, Yu X H, Liu Z W, Yu W W. Adaptive consensus-based robust strategy for economic dispatch of smart grids subject to communication uncertainties. *IEEE Transactions on Industrial Informatics*, 2018, **14**(6): 2484–2496
- 11 Li Xiao-Ling, Wang Huai-Min, Guo Chang-Guo, Ding Bo, Li Xiao-Yong. Research and development of distributed constraint optimization problems. *Chinese Journal of Computers*, 2015, **38**(8): 1656–1671  
(李小玲, 王怀民, 郭长国, 丁博, 李小勇. 分布式约束优化问题研究及其进展. *计算机学报*, 2015, **38**(8): 1656–1671)
- 12 Nedić A, Olshevsky A. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 2015, **60**(3): 601–615
- 13 Qu G N, Li N. Harnessing smoothness to accelerate distributed optimization. *IEEE Transactions on Control of Network Systems*, 2018, **5**(3): 1245–1260
- 14 Shalev-Shwartz S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 2011, **4**(2): 107–194
- 15 Hazan E, Kalai A, Kale S, Agarwal A. Logarithmic regret algorithms for online convex optimization. In: Proceedings of the International Conference on Computational Learning Theory. Berlin, Germany: Springer, 2006. 499–513
- 16 Mateos-Núñez D, Cortés J. Distributed online convex optimization over jointly connected digraphs. *IEEE Transactions on Network Science and Engineering*, 2014, **1**(1): 23–37
- 17 Xu C Q, Zhu J L, Wu D O. Decentralized online learning methods based on weight-balancing over time-varying digraphs. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021, **5**(3): 394–406
- 18 Akbari M, Gharesifard B, Linder T. Distributed online convex optimization on time-varying directed graphs. *IEEE Transactions on Control of Network Systems*, 2017, **4**(3): 417–428
- 19 Raginsky M, Kiarashi N, Willett R. Decentralized online convex programming with local information. In: Proceedings of the American Control Conference. San Francisco, USA: IEEE, 2011. 5363–5369
- 20 Hosseini S, Chapman A, Mesbahi M. Online distributed convex optimization on dynamic networks. *IEEE Transactions on Automatic Control*, 2016, **61**(11): 3545–3550
- 21 Yan F, Sundaram S, Vishwanathan S V N, Qi Y. Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties. *IEEE Transactions on Knowledge and Data Engineering*, 2013, **25**(11): 2483–2493
- 22 Liu S, Xie L H, Quevedo D E. Event-triggered quantized communication-based distributed convex optimization. *IEEE Transactions on Control of Network Systems*, 2018, **5**(1): 167–178
- 23 Liang Q K, Modiano E. Network utility maximization in adversarial environments. In: Proceedings of the IEEE Conference on Computer Communications. Honolulu, USA: IEEE, 2018. 594–602
- 24 Frank M, Wolfe P. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 1956, **3**(1–2): 95–110
- 25 Jaggi M. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In: Proceedings of the 30th International Conference Machine Learning. Atlanta, USA: JMLR, 2013. 427–435
- 26 Clarkson K L. Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. *ACM Transactions on Algorithms*, 2010, **6**(4): 1–34
- 27 Zhu J L, Wu Q T, Zhang M C, Zheng R J, Li K Q. Projection-free decentralized online learning for submodular maximization over time-varying networks. *Journal of Machine Learning Research*, 2021, **22**: 1–42
- 28 Hazan E, Kale S. Projection-free online learning. In: Proceedings of the 29th International Conference Machine Learning. Edinburgh, UK: JMLR, 2012. 521–528
- 29 Garber D, Hazan E. A linearly convergent variant of the conditional gradient algorithm under strong convexity, with applications to online and stochastic optimization. *SIAM Journal on Optimization*, 2016, **26**(3): 1493–1528
- 30 Lafond J, Wai H T, Moulines E. On the online Frank-Wolfe algorithms for convex and non-convex optimizations. arXiv preprint arXiv: 1510.01171, 2016.
- 31 Hazan E, Luo H P. Variance-reduced and projection-free stochastic optimization. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York, USA: JMLR, 2016. 1263–1271
- 32 Wai H T, Lafond J, Scaglione A, Moulines E. Decentralized Frank-Wolfe algorithm for convex and non-convex problems. *IEEE Transactions on Automatic Control*, 2017, **62**(11): 5522–5537
- 33 Zhang W P, Zhao P L, Zhu W W, Hoi S, Zhang T. Projection-free distributed online learning in networks. In: Proceedings of the 34th International Conference of Machine Learning. Sydney, Australia: PMLR, 2017. 4054–4062
- 34 Wan Y Y, Tu W W, Zhang L J. Projection-free distributed online convex optimization with  $O(\sqrt{T})$  communication complexity. In: Proceedings of the 37th International Conference on Machine Learning. Virtual Event: PMLR, 2020. 9818–9828
- 35 Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge: Cambridge University, 2004.
- 36 Liu S, Qiu Z R, Xie L H. Convergence rate analysis of distributed optimization with projected subgradient algorithm. *Automatica*, 2017, **83**: 162–169
- 37 Watts D J, Strogatz S H. Collective dynamics of “small-world” networks. *Nature*, 1998, **393**: 440–442





**吴庆涛** 河南科技大学信息工程学院教授. 主要研究方向为工业互联网, 智能系统, 模式识别和机器学习.

E-mail: wqt8921@haust.edu.cn

**(WU Qing-Tao** Professor at the School of Information Engineering, Henan University of Science and

Technology. His research interest covers industrial internet, intelligent system, pattern recognition, and machine learning.)



**朱军龙** 河南科技大学信息工程学院副教授. 主要研究方向为大规模优化, 分布式多智能体优化, 随机优化及其在机器学习中的应用.

E-mail: jlzhu@haust.edu.cn

**(ZHU Jun-Long** Associate professor at the School of Information Engineering, Henan University of Science and Technology.

His research interest covers large-scale optimization, distributed multi-agent optimization, and stochastic optimization and their applications in machine learning.)



**葛泉波** 南京信息工程大学自动化学院教授. 主要研究方向为信息融合, 非线性滤波, 无人系统和机器学习. 本文通信作者.

E-mail: quanboge@163.com

**(GE Quan-Bo** Professor at the School of Automation, Nanjing University of Information Science and Technology.

His research interest covers information fusion, nonlinear filtering, unmanned system, and machine learning. Corresponding author of this paper.)



**张明川** 河南科技大学信息工程学院教授. 主要研究方向为新型生成网络, 智能信息处理, 医疗辅助诊断和机器学习.

E-mail: zhang\_mch@haust.edu.cn

**(ZHANG Ming-Chuan** Professor at the School of Information Engineering, Henan University of Science and Technology.

His research interest covers new generation network, intelligent information processing, medical aided diagnosis, and machine learning.)