基于镜像下降的分布式弱凸优化算法研究

程松松 [1] 陈 茹 [1] 樊 渊 [1] 邱剑彬 [2]

摘 要 机器学习中的诸多非凸优化问题, 如鲁棒相位恢复、低秩矩阵补全以及稀疏字典学习等, 本质上可归结为弱凸优化问题. 然而, 弱凸优化问题固有的非凸特性使得此类问题的求解极具挑战. 此外, 由于系统复杂度和问题规模的增加以及相关参数的分布式存储需求, 传统的基于单个个体的集中式计算框架难以高效求解此类问题. 针对上述挑战, 设计一种分布式镜像下降算法, 并从 Bregman-Moreau 包络的角度分析其收敛性, 证明算法的收敛速度为 $\mathrm{O}(\ln K/\sqrt{K})$, 其中 K 为算法的迭代步数. 进一步地, 考虑目标函数梯度信息难以精确获取的情形, 采用正交随机方向矩阵法进行梯度估计. 相较于传统的基于随机向量的方法, 该方法利用多维方向信息进行估计, 从而显著提高梯度信息的估计精度和效率. 基于高效的梯度信息估计, 提出一种分布式零阶镜像下降算法, 并获得与已知精确梯度信息情形下相一致的收敛速度. 最后, 通过基于相位恢复问题的数值仿真和实验验证了所提出的两种算法的有效性.

关键词 分布式优化, 弱凸, 镜像下降, Bregman-Moreau 包络, 零阶梯度

引用格式 程松松, 陈茹, 樊渊, 邱剑彬. 基于镜像下降的分布式弱凸优化算法研究. 自动化学报, 2025, **51**(8): 1842−1856

DOI 10.16383/j.aas.c240784 **CSTR** 32138.14.j.aas.c240784

Distributed Mirror Descent Algorithm Design for Weakly Convex Optimization

CHENG Song-Song[1] CHEN Ru[1] FAN Yuan[1] QIU Jian-Bin[2]

Abstract Many non-convex optimization problems in machine learning, such as robust phase retrieval, low-rank matrix recovery, and sparse dictionary learning, can be fundamentally modeled as weakly convex optimization problems. However, the inherent non-convex property of weakly convex optimization makes solving such problems particularly challenging. Besides, with the increase of systems′ complexity and problem scale, as well as distributed storage of related parameters, traditional centralized computing frameworks based on a single node are difficult to solve such problems efficiently. To address these challenges, we design a distributed mirror descent algorithm and analyze its convergence from the perspective of the Bregman-Moreau envelope. It has been proven that the convergence rate of the algorithm is $\mathrm{O}(\ln K/\sqrt{K})$, where K is the number of iterations of the algorithm. Moreover, considering that exact gradient information of the objective function is difficult to obtain, we propose a gradient estimation scheme based on orthogonal random direction matrices. Compared to conventional methods based on random vectors, this method utilizes multi-dimensional directional information for estimation, significantly improving the accuracy and efficiency of gradient information estimation. Furthermore, we propose a distributed zeroth-order mirror descent algorithm based on the effective gradient information estimation result, and achieve a comparable convergence to the algorithm designed based on known exact gradient information. Finally, we validate the effectiveness of the two proposed algorithms by numerical simulations and experiments based on the phase retrieval problem.

Key words Distributed optimization, weakly convex, mirror descent, Bregman-Moreau envelope, zeroth-order gradient

Citation Cheng Song-Song, Chen Ru, Fan Yuan, Qiu Jian-Bin. Distributed mirror descent algorithm design for weakly convex optimization. *Acta Automatica Sinica*, 2025, **51**(8): 1842−1856

现实世界中, 很多问题都可以转化为优化问题来解决, 如基于传感器网络的定位问题、智能电网的资源分配问题以及推荐系统的设计问题等. 然而, 随着问题规模和复杂度的增加, 传统的集中式框架难以实现大规模复杂优化问题的求解. 基于网络化系统的分布式优化因其鲁棒性强、隐私性好以及不依赖中心节点等优点能够有效求解大规模优化问

收稿日期 2024-12-13 录用日期 2025-05-13
Manuscript received December 13, 2024; accepted May 13, 2025
国家自然科学基金 (62103003, U21B6001, 62273121), 安徽省科技创新攻坚计划 (202423i08050033) 资助
Supported by National Natural Science Foundation of China (62103003, U21B6001, 62273121) and Anhui Province Science and Technology Innovation Key Project (202423i08050033)
本文责任编委 王卓
Recommended by Associate Editor WANG Zhuo
1. 安徽大学电气工程与自动化学院, 计算智能与信号处理教育部重点实验室 合肥 230601 2. 哈尔滨工业大学智能控制与系统研究所 哈尔滨 150080
1. Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Electrical Engineering and Automation, Anhui University, Hefei 230601 2. Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150080

题[1−7]. 针对基于网络化系统的大规模优化问题, 涌现了诸多性能优良的分布式求解算法[8−11].

在实际场景中, 个体的决策变量往往受资源的限制, 相应的问题通常构建成带有集合约束的分布式优化问题. 针对有集合约束的凸优化问题, 文献 [12] 设计基于投影的分布式算法, 并分析算法的收敛性能. 文献 [13] 则研究集合约束下的分布式光滑凸优化问题, 从原始−对偶的角度设计基于投影的分布式比例−积分算法. 文献 [14] 考虑多个集合约束的分布式强凸优化问题, 指出多个集合约束情形下带有固定步长的投影梯度跟踪算法的非精确收敛性, 并通过设计衰减步长保证算法的收敛性. 但是上述算法在确定具体的投影点时基本都是基于欧氏距离框架进行计算. 这种基于欧氏距离的投影方法理论简单、便于分析, 但通常仅适用于处理简单的集合约束, 针对较复杂的约束集合 (如单纯形等), 难以给出投影点的显示表达. 镜像下降 (Mirror descent) 采用 Bregman 散度替代传统欧氏距离度量, 是一种基于对偶域优化的参数学习算法, 该类算法能够有效地处理带有复杂约束集合的优化问题. 针对凸优化问题, 文献 [15] 提出一种镜像下降算法, 指出该算法是投影次梯度算法的一般推广, 特别地, 当距离生成函数为欧氏二范数时, 则退化为投影次梯度算法. 针对有约束的随机凸优化问题, 文献 [16] 设计分布式镜像下降算法, 并结合 Epoch 技术进一步优化算法的收敛性能.

在现实工程应用中, 很多分布式优化问题在本质上是非凸的 (目标函数为非凸函数或可行域为非凸集合). 然而, 由于非凸优化问题的种类繁多, 目前尚缺乏统一的分布式求解算法, 并且现有算法的收敛性能往往不甚理想. 例如, 针对特定类型的非凸优化问题, 文献 [17] 设计依赖中心节点的近端交替方向乘子法 (Proximal alternating direction method of multiplier, P-ADMM), 但该方法仅能渐近收敛到算法稳定点的邻域. 值得注意的是, 现实中很多非凸优化问题, 例如鲁棒相位恢复、低秩矩阵补全和稀疏字典学习, 可以等效转化为弱凸优化模型, 以便更有效地求解. 针对可行集约束下的弱凸优化问题, 文献 [18] 提出一种基于动量的 Frank-Wolfe 算法, 但其收敛速度仅局限于 $\mathrm{O}(1/\ln K)$. 文献 [19] 首次设计投影次梯度方法, 并从 Moreau 包络的角度证明了算法能够以 $\mathrm{O}(\ln K/\sqrt{K})$ 的速度收敛到最优解. 进一步地, 文献 [20] 将次梯度方法推广到有通信延迟的情形, 并证明了在弱凸目标函数满足锐度特性的条件下, 该方法具有线性收敛速度.

目标函数的梯度信息在设计优化算法时直接影响算法的迭代方向. 然而在很多实际问题中, 目标函数的梯度信息往往难以甚至无法精确获取. 为此, 一种可行的策略是在决策变量邻域进行微小摄动, 并利用摄动点和原决策变量点的目标函数值进行差分, 从而估计梯度信息. 针对凸优化问题, 文献 [21] 从双向差分、左向差分以及右向差分三个维度提出梯度信息的估计策略, 并在此基础上设计分布式优化算法. 此外, 文献 [22−23] 研究分布式随机零阶梯度算法. 针对分布式非凸优化问题, 文献 [24] 从原始−对偶的角度设计分布式零阶优化算法, 并证明了其在 PŁ 条件下的线性收敛性. 而文献 [25] 则进一步将零阶梯度与梯度跟踪技术相结合, 提出一种新的分布式优化算法, 并系统地论证了其在满足 PŁ 条件下的线性收敛性.

基于以上讨论, 分布式有约束弱凸优化的求解尚存在如下问题有待进一步解决:

1) 针对带有简单约束的弱凸优化问题, 可采用传统的投影梯度方法进行求解. 然而针对带有复杂约束的弱凸优化问题, 投影梯度方法在保证解的可行性时计算负担较大. 如何设计更有效的分布式镜像下降算法求解该类问题?

2) 针对有约束弱凸优化问题, 现有的基于投影梯度方法构建的 Moreau 包络分析框架难以直接应用于分布式镜像下降算法的收敛性能分析. 如何拓展现有 Moreau 包络框架以分析分布式镜像下降算法的收敛性?

3) 在梯度信息不可精确获取的情形下, 如何突破现有梯度估计框架, 设计更高效的梯度估计策略, 以提升算法的求解性能? 此外, 基于零阶梯度设计的算法能否获得和基于精确梯度信息算法相匹配的收敛性能?

针对上述挑战, 本文聚焦于大规模有约束弱凸优化问题, 设计分布式镜像下降算法, 并进一步扩展至梯度信息不可精确获取的情形, 分析算法收敛性能, 并应用于相位恢复问题的求解. 本文的主要贡献简要总结如下:

1) 相较于文献 [12, 19] 中基于欧氏距离的投影算法, 本文设计更一般的基于 Bregman 散度的镜像下降算法, 能够有效处理更复杂的集合约束 (如单纯形约束等). 此外, 与文献 [15−16] 中针对强凸函数的镜像下降算法相比, 本文考虑更一般的弱凸优化问题, 从而拓宽了镜像下降算法的适用范围.

2) 针对分布式弱凸优化问题的非凸属性, 本文从 Bregman-Moreau 包络的角度建立算法收敛性能的分析框架, 并从邻近投影的角度分析算法的收

敛速度. 特别地, 文献 [19, 26] 中的相关方法和结果即为本文中距离生成函数 $\phi(\boldsymbol{x})$ 取 $\frac{1}{2}\|\boldsymbol{x}\|^2$ 时的特殊情况.

3) 相较于传统的基于随机向量的梯度估计策略[21−25], 本文从更高维度引入基于随机方向矩阵的梯度信息估计方法, 提升梯度信息的估计精度. 在此基础上, 本文设计分布式零阶优化算法, 并证明其收敛速度可达到 $\mathrm{O}(\ln K/\sqrt{K})$.

本文结构安排如下: 第 1 节介绍网络拓扑相关的基础知识并给出分布式弱凸优化问题以及相关的基础理论; 第 2 节考虑在梯度信息精确已知的情形下, 设计分布式镜像下降算法, 并分析算法的收敛性能; 第 3 节进一步考虑梯度信息未知情形, 设计基于零阶梯度的分布式镜像下降算法并给出算法收敛性能; 第 4 节通过相位恢复的数值仿真和实验验证了所提算法的有效性; 第 5 节总结全文.

符号说明. 本文将正整数域、m 维和 $n\times m$ 维实数域分别记为 $\mathbf{N}_+$、$\mathbf{R}^m$ 和 $\mathbf{R}^{n\times m}$. $[l]$ 表示集合 $\{1, 2, \cdots, l\}$. 给定函数 $f(\boldsymbol{x})$, $\nabla f(\boldsymbol{x})$ 表示其梯度, $\langle\nabla f(\boldsymbol{x}), \boldsymbol{y}\rangle$ 为 $\nabla f(\boldsymbol{x})$ 和向量 $\boldsymbol{y}$ 的内积. $\mathbf{1}_N\in\mathbf{R}^N$ 表示一个全 1 的列向量. $\boldsymbol{I}_m\in\mathbf{R}^{m\times m}$ 为单位矩阵. $\|\boldsymbol{x}\|$ 为向量 $\boldsymbol{x}$ 的 2 范数. 此外, $\boldsymbol{P}^{(j)}$ 表示矩阵 $\boldsymbol{P}$ 的第 j 列, $[\cdot]_j$ 表示给定向量的第 j 个分量.

1 问题描述与基础知识

1.1 基础概念

如图 1 所示, 网络化系统在 k 时刻的网络拓扑图记为 $\mathcal{G}_k=(\mathcal{V}, \mathcal{E}_k)$, 其中, $\mathcal{V}=\{1, 2, \cdots, n\}$ 为图的节点集, $\mathcal{E}_k\in\mathcal{V}\times\mathcal{V}$ 为图在 k 时刻的边集. $(j, i)\in\mathcal{E}_k$ 表示在 k 时刻节点 i 能够从节点 j 获取信息, 同时称节点 j 是节点 i 的邻居. 节点 i 在 k 时刻的所有邻居构成的集合记为 $\mathcal{N}_{i,\,k}$. 图 $\mathcal{G}_k$ 的权重矩阵 $\boldsymbol{W}_k=[w_{ij,\,k}]\in\mathbf{R}^{n\times n}$, 如果 $(j, i)\in\mathcal{E}_k$, 则存在 $\eta\in(0, 1)$, 使得 $w_{ij,\,k}\geq\eta$, 否则 $w_{ij,\,k}=0$. 如果任意两个节点之间至少存在一条有向路径, 则该图称为强连通的. 基于上述定义, 对图 $\mathcal{G}_k$ 给出如下假设:

假设 1. 图 $\mathcal{G}_k=(\mathcal{V}, \mathcal{E}_k)$ 满足如下条件:

1) 权重矩阵 $[w_{ij,\,k}]$ 满足双随机特性, 即针对任意的 $i, j\in\mathcal{V}$ 和 $k\in\mathbf{N}_+$, 都有 $\sum_{\ell=1}^n w_{i\ell,\,k}=1$ 且 $\sum_{\ell=1}^n w_{\ell j,\,k}=1$ 成立;

2) 存在某个正整数 B, 使得针对任意 $s\in\mathbf{N}_+$, 图 $\left(\mathcal{V}, \cup_{k=sB+1}^{(s+1)B}\mathcal{E}_k\right)$ 均为强连通.

1.2 问题描述

考虑在一个包含有 n 个智能体的网络化系统

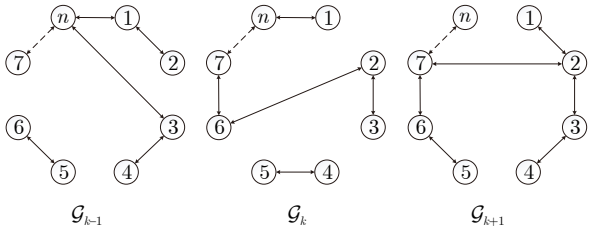


图 1 网络化系统示意图

Fig. 1 Networked system schematic

中, 需要求解如下优化问题:

$$\begin{aligned}&\min_{\boldsymbol{x}_o}\ f(\boldsymbol{x}_o)=\frac{1}{n}\sum_{i=1}^n f_i(\boldsymbol{x}_o)\\&\text{s.t.}\ \ \boldsymbol{x}_o\in\mathcal{X}\end{aligned}\tag{1}$$

其中, $f_i:\mathbf{R}^m\to\mathbf{R}$ 和 $\mathcal{X}\subset\mathbf{R}^m$ 分别为智能体 i 的局部目标函数和约束集合. 为实现分布式求解, 可将问题 (1) 等价地转化为如下形式:

$$\begin{aligned}&\min_{\boldsymbol{x}}\ f(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^n f_i(\boldsymbol{x}_i)\\&\text{s.t.}\ \ \boldsymbol{x}_i=\boldsymbol{x}_j\in\mathcal{X},\quad\forall\, i,\, j\in\mathcal{V}\end{aligned}\tag{2}$$

其中, $\boldsymbol{x}=[\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}, \cdots, \boldsymbol{x}_n^{\mathrm{T}}]^{\mathrm{T}}\in\mathbf{R}^{nm}$ 为全局决策变量. 不失一般性, 假设智能体 i 仅能获取其局部的相关信息 (如决策变量和目标函数的函数值或梯度值) 且仅能与其邻居进行决策变量的交互.

针对问题 (2), 给出如下合理的假设.

假设 2. 局部目标函数 $f_i(\cdot)$ 满足弱凸条件, 即 $\exists\sigma>0$, 使得 $f_i(\boldsymbol{x}_o)+\frac{\sigma}{2}\|\boldsymbol{x}_o\|^2$ 为关于 $\boldsymbol{x}_o$ 的凸函数.

假设 3. 局部目标函数 $f_i(\cdot)$ 及其梯度函数 $\nabla f_i(\cdot)$ 分别满足 L -Lipschitz 连续和 λ -Lipschitz 连续, 即存在正常数 L 和 λ, 使得对任意 $\boldsymbol{x}_1$, $\boldsymbol{x}_2\in\mathcal{X}$, 均有 $\|f_i(\boldsymbol{x}_2)-f_i(\boldsymbol{x}_1)\|\leq L\|\boldsymbol{x}_2-\boldsymbol{x}_1\|$ 和 $\|\nabla f_i(\boldsymbol{x}_2)-\nabla f_i(\boldsymbol{x}_1)\|\leq\lambda\|\boldsymbol{x}_2-\boldsymbol{x}_1\|$ 成立.

注 1. 根据假设 2, 本文所考虑的问题是一类非凸的分布式有约束优化问题, 其在鲁棒相位恢复、稀疏字典学习以及低秩矩阵补全等领域具有广泛的应用. 然而, 由于问题固有的非凸特性和集合约束的存在 (特别是当集合约束具有复杂结构时, 如单纯形约束等), 如何实现该类问题的有效求解, 是一个充满挑战且具有学术价值的问题. 根据假设 3, 局部目标函数 $f_i(\cdot)$ 的梯度有界, 即针对任意 $\boldsymbol{x}\in\mathcal{X}$, 均有 $\|\nabla f_i(\boldsymbol{x})\|\leq L$ 成立.

针对问题 (2) 的目标函数, 有如下基础性结论.

引理 1[19]**.** 若假设 2 成立, 则对任意 $\boldsymbol{x}_i\in\mathbf{R}^m$, $i\in\mathcal{V}$, 恒有

$$f\left(\sum_{i=1}^{n}\theta_i\boldsymbol{x}_i\right)\leq\sum_{i=1}^{n}\theta_i f(\boldsymbol{x}_i)+\frac{\sigma}{2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\theta_i\theta_j\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2$$

其中, $\sum_{i=1}^{n}\theta_i=1$ 且 $\theta_i\in[0,\,1]$, $\forall i\in\mathcal{V}$.

1.3　基础知识

为便于后续设计分布式镜像下降算法, 在此给出镜像映射的一个基础概念.

定义 1[27]**.** 考虑函数 $\phi(\cdot):\mathbf{R}^m\to\mathbf{R}\cup+\infty$, 并记其定义域内部构成的集合为 $\mathcal{D}_\phi$. 若如下条件成立, 则 $\phi(\cdot)$ 是 $\mathcal{X}$ 相容镜像映射.

1) $\phi(\cdot)$ 是下半连续且严格凸;

2) $\phi(\cdot)$ 在 $\mathcal{D}_\phi$ 上可微, 且 $\phi(\cdot)$ 的梯度可取 $\mathbf{R}^m$ 中的任意值, 即 $\nabla\phi(\mathcal{D}_\phi)=\mathbf{R}^m$;

3) $\mathcal{X}\cap\mathcal{D}_\phi$ 非空且 $\mathcal{X}\subset\mathrm{cl}\mathcal{D}_\phi$, $\mathrm{cl}\mathcal{D}_\phi$ 为 $\mathcal{D}_\phi$ 的闭包.

考虑函数 $\phi(\cdot)$ 是 $\mathcal{X}$ 相容镜像映射, 并满足如下条件.

假设 4. 函数 $\phi(\cdot)$ 是 σ_ϕ 强凸的, 其中 $\sigma_\phi\geq1$, 即对于任意 $\boldsymbol{u},\,\boldsymbol{v}\in\mathbf{R}^m$, 有

$$\phi(\boldsymbol{v})\geq\phi(\boldsymbol{u})+\langle\nabla\phi(\boldsymbol{u}),\,\boldsymbol{v}-\boldsymbol{u}\rangle+\frac{\sigma_\phi}{2}\|\boldsymbol{v}-\boldsymbol{u}\|^2$$

定义 2. 基于强凸函数 $\phi(\boldsymbol{x})$ 的 Bregman 散度为

$$V_\phi(\boldsymbol{v},\,\boldsymbol{u})=\phi(\boldsymbol{v})-\phi(\boldsymbol{u})-\langle\nabla\phi(\boldsymbol{u}),\,\boldsymbol{v}-\boldsymbol{u}\rangle$$

注 2. 定义 2 中给出的 Bregman 散度是传统欧氏距离函数的一般性推广. 特别地, 当 $\phi(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{x}\|^2$ 时, Bregman 散度则退化为传统的欧氏距离. 这种更一般的距离函数有助于后续算法的设计与分析.

针对 Bregman 散度 $V_\phi(\cdot,\,\cdot)$, 给出如下假设.

假设 5. 函数 $V_\phi(\boldsymbol{v},\,\boldsymbol{u})$ 具有 G -Lipschitz 连续梯度, 同时满足

$$V_\phi\left(\sum_{j=1}^{n}\theta_j\boldsymbol{v}_j,\,\sum_{j=1}^{n}\theta_j\boldsymbol{u}_j\right)\leq\sum_{j=1}^{n}\theta_j V_\phi(\boldsymbol{v}_j,\,\boldsymbol{u}_j)$$

注 3. Bregman 散度 $V_\phi(\cdot,\,\cdot)$ 是相较于传统欧氏距离的更一般形式. 当 $\phi(\cdot)=\frac{1}{2}\|\cdot\|^2$ 时, 相应的 Bregman 散度 $V_\phi(\boldsymbol{v},\,\boldsymbol{u})=\frac{1}{2}\|\boldsymbol{v}-\boldsymbol{u}\|^2$ 为关于 $(\boldsymbol{v},\,\boldsymbol{u})$ 的凸函数且具有 Lipschitz 连续梯度, 故假设 5 成立; 当 $\phi(\boldsymbol{v})=\sum_{i=1}^{m}[\boldsymbol{v}]_i\ln[\boldsymbol{v}]_i$ 时, $V_\phi(\boldsymbol{v},\,\boldsymbol{u})=\sum_{i=1}^{n}[\boldsymbol{v}]_i(\ln[\boldsymbol{v}]_i-\ln[\boldsymbol{u}]_i)$ 为关于 $(\boldsymbol{v},\,\boldsymbol{u})$ 的凸函数, 且在集合 $\mathcal{X}=\{\boldsymbol{v}=\sum_{i=1}^{m}[\boldsymbol{v}]_i=1,\,[\boldsymbol{v}]_i\in[\eta,\,1],\,\eta>0\}$ 上具有 Lipschitz 连续梯度, 故假设 5 成立.

针对目标函数 $f(\boldsymbol{x})$, 其 Bregman-Moreau 包络和相应的邻近投影可定义如下:

定义 3. 函数 $f(\boldsymbol{x})$ 的 Bregman-Moreau 包络为

$$f_\epsilon(\boldsymbol{x})=\min_{\boldsymbol{v}\in\mathbf{R}^m}f(\boldsymbol{v})+\frac{1}{\epsilon}V_\phi(\boldsymbol{v},\,\boldsymbol{x}),\ \epsilon\in\left(0,\,\frac{1}{4n(\sigma+1)}\right)$$

对应地, 使函数 $f(\boldsymbol{v})+\frac{1}{\epsilon}V_\phi(\boldsymbol{v},\,\boldsymbol{x})$ 关于 $\boldsymbol{v}$ 取得最小值的决策变量为 $\check{\boldsymbol{x}}$, 记为

$$\mathrm{prox}_{\epsilon f}(\boldsymbol{x})=\check{\boldsymbol{x}}=\arg\min_{\boldsymbol{v}\in\mathbf{R}^m}f(\boldsymbol{v})+\frac{1}{\epsilon}V_\phi(\boldsymbol{v},\,\boldsymbol{x})$$

其中, $\mathrm{prox}_{\epsilon f}(\boldsymbol{x})=\check{\boldsymbol{x}}$ 称为邻近投影. 值得指出的是 $\mathrm{prox}_{\epsilon f}(\boldsymbol{x})$ 仅在分析中使用, 并不参与算法的实际计算.

注 4. Bregman-Moreau 包络实质上是在 Bregman 散度意义下, 寻求距离 $\boldsymbol{x}$ 较近且使得函数 $f(\boldsymbol{v})$ 较小的点 $\check{\boldsymbol{x}}$, 其中, $\check{\boldsymbol{x}}$ 与 $\boldsymbol{x}$ 的距离和函数值 $f(\boldsymbol{v})$ 之间的权重通过参数 ϵ 进行调节. 特别地, 当 Bregman 散度退化为传统欧氏距离时, 本文采用的 Bregman-Moreau 包络与传统的 Moreau 包络等价.

2　分布式算法设计与分析

2.1　分布式镜像下降算法设计

针对式 (2) 中的分布式弱凸优化问题, 本文设计如下分布式镜像下降算法.

算法 1. 分布式镜像下降算法

输入. 初始化迭代次数 K, 步长 α_k;

初始化. $\boldsymbol{x}_{i,\,1}\in\mathbf{R}^m$, $\forall\,i\in\mathcal{V}$;

1: **for** $k=1,\,2,\,\cdots,\,K-1$ **do**

2:　　计算 $f_i(\cdot)$ 在 $\boldsymbol{x}_{i,\,k}$ 处的真实梯度 $\nabla f_i(\boldsymbol{x}_{i,\,k})$;

3:　　在距离生成函数 $\phi(\cdot)$ 的梯度域进行梯度下降 $\nabla\phi(\boldsymbol{y}_{i,\,k})=\nabla\phi(\boldsymbol{x}_{i,\,k})-\alpha_k\nabla f_i(\boldsymbol{x}_{i,\,k})$;

4:　　关于距离函数 $V_\phi(\cdot,\,\cdot)$ 求解邻近投影 $\boldsymbol{z}_{i,\,k+1}=\arg\min_{\boldsymbol{x}_o\in\mathcal{X}}V_\phi(\boldsymbol{x}_o,\,\boldsymbol{y}_{i,\,k})$;

5:　　智能体与其邻居交互信息并加权平均 $\boldsymbol{x}_{i,\,k+1}=\sum_{j\in\mathcal{N}_{i,\,k+1}}w_{ij,\,k+1}\boldsymbol{z}_{j,\,k+1}$;

6: **end for**

返回. $\boldsymbol{x}_{i,\,K}$.

注 5. 在算法 1 中, $\boldsymbol{y}_{i,\,k}$ 的迭代机制为 $\nabla\phi(\boldsymbol{y}_{i,\,k})=\nabla\phi(\boldsymbol{x}_{i,\,k})-\alpha_k\nabla f_i(\boldsymbol{x}_{i,\,k})$. 这种迭代机制解释如下: 首先将变量 $\boldsymbol{x}_{i,\,k}$ 映射到函数 $\phi(\cdot)$ 的梯度空间, 并在该空间内执行目标函数的梯度下降, 得到 $\nabla\phi(\boldsymbol{y}_{i,\,k})$. 随后, 根据函数 $\phi(\cdot)$ 的具体形式, 通过逆映射将 $\nabla\phi(\boldsymbol{y}_{i,\,k})$ 转换回原始决策变量空间, 从而获得 $\boldsymbol{y}_{i,\,k}$. 因此, $\boldsymbol{y}_{i,\,k}$ 的具体迭代方式依赖于函数

$\phi(\cdot)$ 的设计. 特别地, 当设计 $\phi(\cdot)=\frac{1}{2}\|\cdot\|^2$ 时, 其梯度映射为恒等映射, 从而直接得到 $\boldsymbol{y}_{i,\,k}=\boldsymbol{x}_{i,\,k}-\alpha_k\nabla f_i(\boldsymbol{x}_{i,\,k})$.

注 6. 针对强凸的优化问题, 文献 [16] 设计分布式镜像下降算法. 受文献 [16] 中算法的启发, 本文设计算法 1. 与文献 [16] 相比, 本文具有如下区别和挑战: 首先, 本文将文献 [16] 中的目标函数由强凸弱化为更一般的弱凸; 其次, 针对弱凸优化问题, 文献 [16] 中采用的算法收敛性分析方法已不再适用, 如何有效分析本文分布式镜像下降算法的收敛性能, 需要寻求新的技术和方法.

2.2 算法收敛性分析

为便于后续分析算法的收敛性, 首先给出如下邻近投影:

$$\begin{cases}\check{\boldsymbol{x}}_k=\arg\min\limits_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})+\dfrac{1}{\epsilon}V_\phi(\boldsymbol{x},\,\bar{\boldsymbol{x}}_k)\\ \hat{\boldsymbol{x}}_{i,\,k}=\arg\min\limits_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})+\dfrac{1}{\epsilon}V_\phi(\boldsymbol{x},\,\boldsymbol{x}_{i,\,k})\\ \hat{\boldsymbol{z}}_{i,\,k}=\arg\min\limits_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})+\dfrac{1}{\epsilon}V_\phi(\boldsymbol{x},\,\boldsymbol{z}_{i,\,k})\end{cases}\tag{3}$$

根据 Bregman-Moreau 包络的定义, $\check{\boldsymbol{x}}_k$, $\hat{\boldsymbol{x}}_{i,\,k}$ 和 $\hat{\boldsymbol{z}}_{i,\,k}$ 分别是距离 $\bar{\boldsymbol{x}}_k$, $\boldsymbol{x}_{i,\,k}$ 和 $\boldsymbol{z}_{i,\,k}$ 较近且使得 $f(\boldsymbol{x})$ 函数值较小的决策变量. 根据式 (3) 的最优性条件, 有

$$0\in\nabla f(\check{\boldsymbol{x}}_k)+\partial\boldsymbol{I}_{\mathcal{X}}(\check{\boldsymbol{x}}_k)+\frac{1}{\epsilon}(\nabla\phi(\check{\boldsymbol{x}}_k)-\nabla\phi(\bar{\boldsymbol{x}}_k))\tag{4}$$

其中, $\boldsymbol{I}_{\mathcal{X}}(\boldsymbol{x})$ 为集合 $\mathcal{X}$ 的示性函数: 当 $\boldsymbol{x}\in\mathcal{X}$ 时, $\boldsymbol{I}_{\mathcal{X}}(\boldsymbol{x})=0$; 否则 $\boldsymbol{I}_{\mathcal{X}}(\boldsymbol{x})=\infty$. $\partial\boldsymbol{I}_{\mathcal{X}}(\check{\boldsymbol{x}}_k)$ 为函数在 $\check{\boldsymbol{x}}_k$ 处的次梯度集合[1]. 基于式 (4), 有

$$\begin{aligned}&\operatorname{dist}(0,\,\nabla f(\check{\boldsymbol{x}}_k)+\partial\boldsymbol{I}_{\mathcal{X}}(\check{\boldsymbol{x}}_k))\;=\\ &\qquad\frac{1}{\epsilon}\|\nabla\phi(\check{\boldsymbol{x}}_k)-\nabla\phi(\bar{\boldsymbol{x}}_k)\|\end{aligned}\tag{5}$$

其中, $\operatorname{dist}(0,\,\nabla f(\check{\boldsymbol{x}}_k)+\partial\boldsymbol{I}_{\mathcal{X}}(\check{\boldsymbol{x}}_k))$ 表示 0 到集合 $\nabla f(\check{\boldsymbol{x}}_k)+\partial\boldsymbol{I}_{\mathcal{X}}(\check{\boldsymbol{x}}_k)$ 的距离. 针对式 (1) 中的问题, 其最优性条件为 $0\in\nabla f(\boldsymbol{x}^*)+\partial\boldsymbol{I}_{\mathcal{X}}(\boldsymbol{x}^*)$, 即 $\operatorname{dist}(0,\,\nabla f(\boldsymbol{x}^*)+\partial\boldsymbol{I}_{\mathcal{X}}(\boldsymbol{x}^*))=0$. 因此, 结合式 (5) 中的结论, 可以通过 $\frac{1}{\epsilon}\|\nabla\phi(\check{\boldsymbol{x}}_k)-\nabla\phi(\bar{\boldsymbol{x}}_k)\|$ 衡量 $\check{\boldsymbol{x}}_k$ 与最优解间的距离.

此外, 针对算法 1, 给出如下基础性结果.

引理 2. 若假设 1 ~ 4 成立, 局部变量 $\boldsymbol{z}_{i,\,k}$ 和 $\boldsymbol{x}_{i,\,k}$ 的一致性误差具有如下上界:

$$\|\boldsymbol{z}_{i,\,k}-\bar{\boldsymbol{z}}_k\|\le\gamma\beta^{k-1}\sum_{i=1}^n\|\boldsymbol{z}_{i,\,1}\|+\frac{nL\gamma}{\sigma_\phi}\sum_{\kappa=1}^{k-1}\beta^{k-1-\kappa}\alpha_\kappa$$

$$\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|\le\gamma\beta^{k-1}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|+\frac{nL\gamma}{\sigma_\phi}\sum_{\kappa=1}^{k-1}\beta^{k-\kappa}\alpha_\kappa$$

其中, $\bar{\boldsymbol{z}}_k=\frac{1}{n}\sum_{i=1}^n\boldsymbol{z}_{i,\,k}$, $\bar{\boldsymbol{x}}_k=\frac{1}{n}\sum_{i=1}^n\boldsymbol{x}_{i,\,k}$, $\beta=(1-\eta/4n^2)^{1/B}$, $\gamma=(1-\eta/4n^2)^{-2}$.

引理 2 的证明见附录 A.

引理 3[19]**.** 针对 $\beta=(1-\eta/4n^2)^{1/B}\in(0,\,1)$ 和步长序列 $\{\alpha_k\}$, 当 $\alpha_k=1/k^q$, $q\in(0,\,1]$ 时, 有 $\sum_{k=0}^{K-1}\beta^k\alpha_{K-k-1}=\mathrm{O}\big(\frac{\alpha_{K-1}}{1-\beta}\big)$.

引理 4[28]**.** 若假设 2 成立, 则针对式 (3) 中的 $\check{\boldsymbol{x}}_k$ 与 $\hat{\boldsymbol{x}}_{i,\,k}$, 有 $\|\check{\boldsymbol{x}}_k-\hat{\boldsymbol{x}}_{i,\,k}\|\le\frac{1}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|$.

基于以上讨论, 算法 1 的收敛性能总结如下.

定理 1. 假设 1 ~ 3 成立, 令 $\alpha_k=1/\sqrt{k}$, $\forall\,k\in[K]$, 若所有智能体 $i\in\mathcal{V}$ 依据算法 1 生成 $\boldsymbol{x}_{i,\,k}$, 则有

$$\inf_{1\le k\le K}\frac{1}{\epsilon^2}\|\nabla\phi(\bar{\boldsymbol{x}}_k)-\nabla\phi(\check{\boldsymbol{x}}_k)\|^2\le\frac{c_1}{\sqrt{K}}+\frac{c_2\ln K}{\sqrt{K}}$$

其中,

$$\begin{aligned}c_1=\;&\frac{2G^2}{1-4n\sigma\epsilon}\Bigg[\frac{2\sigma n^2\gamma^2[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2(1-\beta)}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|^2+\\ &\frac{nL\gamma(2-\epsilon\sigma)}{\epsilon(1-\epsilon\sigma)(1-\beta)}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|+\\ &\frac{n\sigma\gamma^2}{(1-\beta)(1-\epsilon\sigma)^2}\sum_{i=1}^n\|\boldsymbol{z}_{i,\,1}\|^2+\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,1})-\\ &\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,K+1})\Bigg]\\ c_2=\;&\frac{2G^2}{1-4n\sigma\epsilon}\Bigg[\frac{nL^2}{2\epsilon}+\frac{(2-\epsilon\sigma)\gamma\beta n^2L^2}{\epsilon\sigma_\phi(1-\epsilon\sigma)(1-\beta)}+\\ &\frac{2[2(1-\epsilon\sigma)^2+1]\sigma\gamma^2\beta^2L^2n^3}{\epsilon\sigma_\phi^2(1-\epsilon\sigma)^2(1-\beta)^2}+\\ &\frac{\sigma n^2L^2\gamma^2}{\sigma_\phi^2(1-\beta)^2(1-\epsilon\sigma)^2}\Bigg]\end{aligned}$$

证明. 根据算法 1, 有

$$\begin{aligned}&\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\,\boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\;=\\ &\quad\frac{1}{\alpha_k}\langle\nabla\phi(\boldsymbol{x}_{i,\,k})-\nabla\phi(\boldsymbol{y}_{i,\,k}),\,\boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\overset{(a)}{=}\\ &\quad\frac{1}{\alpha_k}[V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})-V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{y}_{i,\,k})+\\ &\quad V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})]\end{aligned}\tag{6}$$

[1] 考虑非光滑凸函数 $h(\cdot):\mathbf{R}^m\to\mathbf{R}$, $\boldsymbol{u}$ 为其定义域 $\mathrm{dom}h$ 中一点. 若向量 $\boldsymbol{g}\in\mathbf{R}^m$ 满足 $h(\boldsymbol{v})\ge h(\boldsymbol{u})+\langle\boldsymbol{g},\boldsymbol{v}-\boldsymbol{u}\rangle$, $\forall\boldsymbol{v}\in\mathrm{dom}h$, 则称 $\boldsymbol{g}$ 为函数 $h(\cdot)$ 在点 $\boldsymbol{u}$ 处的一个次梯度; 此外, 记 $\partial h(\boldsymbol{u})=\{\boldsymbol{g}|\boldsymbol{g}\in\mathbf{R}^m,\,h(\boldsymbol{v})\ge h(\boldsymbol{u})+\langle\boldsymbol{g},\,\boldsymbol{v}-\boldsymbol{u}\rangle,\forall\boldsymbol{v}\in\mathrm{dom}h\}$ 为函数 $h(\cdot)$ 在点 $\boldsymbol{u}$ 处的次梯度集合[29].

其中, $\overset{(a)}{=}$ 是根据 $\langle\nabla\phi(\boldsymbol{b})-\nabla\phi(\boldsymbol{c}),\ \boldsymbol{b}-\boldsymbol{a}\rangle = V_\phi(\boldsymbol{a},\ \boldsymbol{b}) - V_\phi(\boldsymbol{a},\ \boldsymbol{c}) + V_\phi(\boldsymbol{b},\ \boldsymbol{c})$ 所得. 由于 $\hat{\boldsymbol{x}}_{i,\,k}\in\mathcal{X}$ 且 $\boldsymbol{z}_{i,\,k+1}$ 是将 $\boldsymbol{y}_{i,\,k}$ 关于距离函数 $V_\phi(\cdot,\ \cdot)$ 向集合 $\mathcal{X}$ 的投影, 故 $V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{z}_{i,\,k+1})\leq V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{y}_{i,\,k})$. 因此, 有

$$\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\ \boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\leq\frac{1}{\alpha_k}[V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{x}_{i,\,k})- V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{z}_{i,\,k+1})+V_\phi(\boldsymbol{x}_{i,\,k},\ \boldsymbol{y}_{i,\,k})] \tag{7}$$

将式 (7) 关于 i 从 1 到 n 求和, 即

$$\sum_{i=1}^{n}\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\ \boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\ \leq \frac{1}{\alpha_k}\left[\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{x}_{i,\,k})-\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{z}_{i,\,k+1})+\sum_{i=1}^{n}V_\phi(\boldsymbol{x}_{i,\,k},\ \boldsymbol{y}_{i,\,k})\right] \tag{8}$$

针对 $V_\phi(\boldsymbol{x}_{i,\,k},\ \boldsymbol{y}_{i,\,k})$, 有

$$V_\phi(\boldsymbol{x}_{i,\,k},\ \boldsymbol{y}_{i,\,k})+V_\phi(\boldsymbol{y}_{i,\,k},\ \boldsymbol{x}_{i,\,k})\overset{(a)}{=}\langle\nabla\phi(\boldsymbol{x}_{i,\,k})-\nabla\phi(\boldsymbol{y}_{i,\,k}),\ \boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\rangle\overset{(b)}{=}\langle\alpha_k\nabla f_i(\boldsymbol{x}_{i,\,k}),\ \boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\rangle\overset{(c)}{\leq}\frac{L^2\alpha_k^2}{2}+\frac{1}{2}\|\boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\|^2 \tag{9}$$

其中, $\overset{(a)}{=}$ 是根据 $V_\phi(\boldsymbol{x}_{i,\,k},\ \boldsymbol{y}_{i,\,k})$ 的定义所得, $\overset{(b)}{=}$ 是根据算法 1 所得, $\overset{(c)}{\leq}$ 则源于 Young 不等式和假设 3. 因此, 有

$$V_\phi(\boldsymbol{x}_{i,\,k},\ \boldsymbol{y}_{i,\,k})\leq\frac{L^2\alpha_k^2}{2}+\frac{1}{2}\|\boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\|^2- V_\phi(\boldsymbol{y}_{i,\,k},\ \boldsymbol{x}_{i,\,k})\leq\frac{L^2\alpha_k^2}{2} \tag{10}$$

将式 (10) 代入式 (8), 有

$$\sum_{i=1}^{n}\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\ \boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\leq\frac{nL^2\alpha_k}{2}+\frac{1}{\alpha_k}\left[\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{x}_{i,\,k})-\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{z}_{i,\,k+1})\right] \tag{11}$$

根据目标函数 $f_i(\cdot)$ 的弱凸性, 有

$$\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\ \boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\geq f_i(\boldsymbol{x}_{i,\,k})-f_i(\hat{\boldsymbol{x}}_{i,\,k})-\frac{\sigma}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2 \tag{12}$$

结合式 (11) 和式 (12), 可得

$$\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{z}_{i,\,k+1})\ \leq \sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{x}_{i,\,k})+\frac{nL^2\alpha_k^2}{2}+\alpha_k\sum_{i=1}^{n}\left[\frac{\sigma}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2+f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\boldsymbol{x}_{i,\,k})\right] \tag{13}$$

针对 $f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\boldsymbol{x}_{i,\,k})$, 有

$$f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\boldsymbol{x}_{i,\,k})=f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\check{\boldsymbol{x}}_k)+f_i(\check{\boldsymbol{x}}_k)-f_i(\bar{\boldsymbol{x}}_k)+f_i(\bar{\boldsymbol{x}}_k)-f_i(\boldsymbol{x}_{i,\,k})\overset{(a)}{\leq}f_i(\check{\boldsymbol{x}}_k)-f_i(\bar{\boldsymbol{x}}_k)+L(\|\hat{\boldsymbol{x}}_{i,\,k}-\check{\boldsymbol{x}}_k\|+\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|)\overset{(b)}{\leq}f_i(\check{\boldsymbol{x}}_k)-f_i(\bar{\boldsymbol{x}}_k)+\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\| \tag{14}$$

其中, $\overset{(a)}{\leq}$ 是根据函数 $f_i(\cdot)$ 的 L-Lipschitz 连续性所得, $\overset{(b)}{\leq}$ 是根据引理 4 所得. 类似地, 针对 $\frac{1}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2$, 有

$$\frac{1}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2=\frac{1}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\check{\boldsymbol{x}}_k+\check{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\ \leq \|\hat{\boldsymbol{x}}_{i,\,k}-\check{\boldsymbol{x}}_k\|^2+\|\check{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\ \leq \frac{1}{(1-\epsilon\sigma)^2}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|^2+\|\check{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2 \tag{15}$$

针对 $\frac{1}{2}\|\check{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2$, 有

$$\frac{1}{2}\|\check{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\leq\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2+\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2 \tag{16}$$

将式 (14) ~ (16) 代入式 (13), 有

$$\begin{aligned}\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{z}_{i,\,k+1})\ \leq\ &\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{x}_{i,\,k})\ +\alpha_k\sum_{i=1}^{n}\left[f_i(\check{\boldsymbol{x}}_k)-f_i(\bar{\boldsymbol{x}}_k)+\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|\ +2\sigma\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2+\sigma(2+\frac{1}{(1-\epsilon\sigma)^2})\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\right]\ +\frac{nL^2\alpha_k^2}{2}\overset{(a)}{\leq}\\ &\sum_{i=1}^{n}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\ \boldsymbol{x}_{i,\,k})+\alpha_k(2n\sigma-\frac{1}{2\epsilon})\times\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2+\alpha_k\sum_{i=1}^{n}\left[\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\sigma\frac{2(1-\epsilon\sigma)^2+1}{(1-\epsilon\sigma)^2}\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\right]+\frac{nL^2\alpha_k^2}{2}\end{aligned} \tag{17}$$

其中, $\overset{(a)}{\leq}$ 是基于不等式 $f(\check{\boldsymbol{x}}_k)+\frac{1}{2\epsilon}\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2\ \leq$

$f(\check{\boldsymbol{x}}_k)+\frac{1}{\epsilon}V_\phi(\check{\boldsymbol{x}}_k,\bar{\boldsymbol{x}}_k)\le f(\bar{\boldsymbol{x}}_k)$ 所得. 根据 Bregman-Moreau 包络的定义, 有

$$f_\epsilon(\boldsymbol{z}_{i,\,k+1})=\min_{\boldsymbol{v}\in\mathcal{X}} f(\boldsymbol{v})+\frac{1}{\epsilon}V_\phi(\boldsymbol{v},\,\boldsymbol{z}_{i,\,k+1}) \tag{18}$$

特别地, 令式 (18) 中 $\boldsymbol{v}$ 的取值为 $\hat{\boldsymbol{x}}_{i,\,k}$, 有

$$f_\epsilon(\boldsymbol{z}_{i,\,k+1})\le f(\hat{\boldsymbol{x}}_{i,\,k})+\frac{1}{\epsilon}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{z}_{i,\,k+1}) \tag{19}$$

将式 (17) 代入式 (19), 可得

$$\begin{aligned}&\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k+1})\le\sum_{i=1}^n f(\hat{\boldsymbol{x}}_{i,\,k})+\frac{nL^2\alpha_k^2}{2\epsilon}+\\&\frac{1}{\epsilon}\Bigg[\sum_{i=1}^n V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})+\alpha_k\sum_{i=1}^n\Bigg(\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\times\\&\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\sigma\frac{2(1-\epsilon\sigma)^2+1}{(1-\epsilon\sigma)^2}\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\Bigg)+\\&\alpha_k\left(2n\sigma-\frac{1}{2\epsilon}\right)\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2\Bigg]=\sum_{i=1}^n\Bigg[f(\hat{\boldsymbol{x}}_{i,\,k})+\\&\frac{1}{\epsilon}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})\Bigg]+\frac{1}{\epsilon}\Bigg[\alpha_k\left(2n\sigma-\frac{1}{2\epsilon}\right)\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2+\\&\alpha_k\sum_{i=1}^n\Bigg(\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\\&\sigma\frac{2(1-\epsilon\sigma)^2+1}{(1-\epsilon\sigma)^2}\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\Bigg)\Bigg]+\frac{nL^2\alpha_k^2}{2\epsilon}\end{aligned} \tag{20}$$

针对 $f(\hat{\boldsymbol{x}}_{i,\,k})+\frac{1}{\epsilon}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})$, 有

$$\begin{aligned}&f(\hat{\boldsymbol{x}}_{i,\,k})+\frac{1}{\epsilon}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})=f\left(\sum_{j=1}^n w_{ij,\,k}\hat{\boldsymbol{x}}_{i,\,k}\right)+\\&\frac{1}{\epsilon}V_\phi\left(\sum_{j=1}^n w_{ij,\,k}\hat{\boldsymbol{x}}_{i,\,k},\,\sum_{j=1}^n w_{ij,\,k}\boldsymbol{z}_{j,\,k}\right)\overset{(a)}{\le}\\&f\left(\sum_{j=1}^n w_{ij,\,k}\hat{\boldsymbol{z}}_{j,\,k}\right)+\\&\frac{1}{\epsilon}V_\phi\left(\sum_{j=1}^n w_{ij,\,k}\hat{\boldsymbol{z}}_{j,\,k},\,\sum_{j=1}^n w_{ij,\,k}\boldsymbol{z}_{j,\,k}\right)\overset{(b)}{\le}\\&\sum_{j=1}^n w_{ij,\,k}\left[f(\hat{\boldsymbol{z}}_{j,\,k})+\frac{1}{\epsilon}V_\phi(\hat{\boldsymbol{z}}_{j,\,k},\,\boldsymbol{z}_{j,\,k})\right]+\\&\frac{\sigma}{2}\sum_{j=1}^{n-1}\sum_{l=j+1}^n w_{ij,\,k}w_{il,\,k}\|\hat{\boldsymbol{z}}_{j,\,k}-\hat{\boldsymbol{z}}_{l,\,k}\|^2\overset{(c)}{\le}\\&\frac{\sigma}{2(1-\epsilon\sigma)^2}\sum_{j=1}^{n-1}\sum_{l=j+1}^n w_{ij,\,k}w_{il,\,k}\|\boldsymbol{z}_{j,\,k}-\\&\boldsymbol{z}_{l,\,k}\|^2+\sum_{j=1}^n w_{ij,\,k}f_\epsilon(\boldsymbol{z}_{j,\,k})\end{aligned} \tag{21}$$

其中, $\overset{(a)}{\le}$ 是根据 $\hat{\boldsymbol{x}}_{i,\,k}$ 的定义和 $\sum_{j=1}^n w_{i,\,j}\hat{\boldsymbol{z}}_{j,\,k}\in\mathcal{X}$ 所得, $\overset{(b)}{\le}$ 和 $\overset{(c)}{\le}$ 分别是基于引理 1 和引理 4 所得. 将式 (21) 代入式 (20), 有

$$\begin{aligned}&\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k+1})\le\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k})+\frac{nL^2\alpha_k^2}{2\epsilon}+\\&\frac{1}{\epsilon}\Bigg[\alpha_k\left(2n\sigma-\frac{1}{2\epsilon}\right)\|\check{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2+\\&\alpha_k\sum_{i=1}^n\Bigg(\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\\&\sigma\frac{2(1-\epsilon\sigma)^2+1}{(1-\epsilon\sigma)^2}\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\Bigg)\Bigg]+\\&\frac{\sigma}{2(1-\epsilon\sigma)^2}\sum_{i=1}^n\sum_{j=1}^{n-1}\sum_{l=j+1}^n w_{ij,\,k}w_{il,\,k}\|\boldsymbol{z}_{j,\,k}-\\&\boldsymbol{z}_{l,\,k}\|^2\le\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k})+\frac{nL^2\alpha_k^2}{2\epsilon}+b_k\end{aligned} \tag{22}$$

其中,

$$\begin{aligned}b_k=&\frac{1}{\epsilon}\Bigg[\alpha_k\sum_{i=1}^n\Bigg(\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\\&\sigma\frac{2(1-\epsilon\sigma)^2+1}{(1-\epsilon\sigma)^2}\|\bar{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2\Bigg)\Bigg]+\frac{\sigma}{2(1-\epsilon\sigma)^2}\times\\&\sum_{i=1}^n\sum_{j=1}^{n-1}\sum_{l=j+1}^n w_{ij,\,k}w_{il,\,k}\|\boldsymbol{z}_{j,\,k}-\boldsymbol{z}_{l,\,k}\|^2\end{aligned} \tag{23}$$

针对 $\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|$, 由引理 2, 可得

$$\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|\le\gamma\beta^{k-1}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|+\frac{nL\gamma\beta\alpha_k}{\sigma_\phi(1-\beta)} \tag{24}$$

基于式 (24), 有

$$\begin{aligned}&\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|^2\le 2n\gamma^2\beta^{2k-2}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|^2+\\&\frac{2n^2L^2\gamma^2\beta^2\alpha_k^2}{\sigma_\phi^2(1-\beta)^2}\le 2n\gamma^2\beta^{k-1}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|^2+\\&\frac{2n^2L^2\gamma^2\beta^2\alpha_k^2}{\sigma_\phi^2(1-\beta)^2}\end{aligned} \tag{25}$$

此外, 根据引理 2, 有

$$\begin{aligned}&\sum_{i=1}^n\sum_{j=1}^{n-1}\sum_{l=j+1}^n w_{ij,\,k}w_{il,\,k}\|\boldsymbol{z}_{j,\,k}-\boldsymbol{z}_{l,\,k}\|^2\le\\&2n\gamma^2\beta^{k-1}\sum_{i=1}^n\|\boldsymbol{z}_{i,\,1}\|^2+\frac{2n^2L^2\gamma^2\alpha_k^2}{\sigma_\phi^2(1-\beta)^2}\end{aligned} \tag{26}$$

将式 (24) ~ (26) 代入式 (23), 可得

$$b_k \le r_1\alpha_k^2 + r_2\beta^{k-1}$$

其中,

$$r_1 = \frac{(2-\epsilon\sigma)\gamma\beta n^2L^2}{\epsilon\sigma_\phi(1-\epsilon\sigma)(1-\beta)} + \frac{2[2(1-\epsilon\sigma)^2+1]\sigma\alpha_1\gamma^2\beta^2L^2n^3}{\epsilon\sigma_\phi^2(1-\epsilon\sigma)^2(1-\beta)^2} + \frac{\sigma n^2L^2\gamma^2}{\sigma_\phi^2(1-\beta)^2(1-\epsilon\sigma)^2}$$

$$r_2 = \frac{nL\gamma\alpha_1(2-\epsilon\sigma)}{\epsilon(1-\epsilon\sigma)}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\| + \frac{2\sigma\alpha_1n^2\gamma^2[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|^2 + \frac{n\sigma\gamma^2}{(1-\epsilon\sigma)^2}\sum_{i=1}^n\|\boldsymbol{z}_{i,\,1}\|^2$$

进一步, 有

$$\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k+1}) - \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k}) \le \frac{nL^2\alpha_k^2}{2\epsilon} + b_k$$

不等式 (22) 可写为如下形式

$$\frac{1}{\epsilon}\left(\frac{1}{2\epsilon}-2n\sigma\right)\alpha_k\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2 \le \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k}) - \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,k+1}) + \frac{nL^2\alpha_k^2}{2\epsilon} + b_k \tag{27}$$

针对式 (27), 关于 k 从 1 到 K 求和, 有

$$\sum_{k=1}^K\frac{1}{\epsilon}\left(\frac{1}{2\epsilon}-2n\sigma\right)\alpha_k\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2 \le \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,1}) - \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,K+1}) + \sum_{k=1}^K\frac{nL^2\alpha_k^2}{2\epsilon} + \sum_{k=1}^K b_k$$

基于 $\epsilon \in (0,\ \frac{1}{4n(\sigma+1)})$, 进一步化简得到

$$\inf_{1\le k\le K}\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2 \le \frac{2\epsilon^2}{1-4n\sigma\epsilon}\times \frac{\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,1}) - \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,K+1}) + \sum_{k=1}^K\frac{nL^2\alpha_k^2}{2\epsilon} + \sum_{k=1}^K b_k}{\sum_{k=1}^K\alpha_k}$$

若 $\alpha_k = 1/\sqrt{k}$, 对于足够大的 K, 有

$$\inf_{1\le k\le K}\frac{1}{\epsilon^2}\|\nabla\phi(\bar{\boldsymbol{x}}_k)-\nabla\phi(\check{\boldsymbol{x}}_k)\|^2 \le \sum_{k=1}^K\frac{G^2}{\epsilon^2}\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2 \le \frac{c_1}{\sqrt{K}} + \frac{c_2\ln K}{\sqrt{K}}$$

其中,

$$c_1 = \frac{2G^2}{1-4n\sigma\epsilon}\left[\frac{2\sigma n^2\gamma^2[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2(1-\beta)}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|^2 + \frac{nL\gamma(2-\epsilon\sigma)}{\epsilon(1-\epsilon\sigma)(1-\beta)}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\| + \frac{n\sigma\gamma^2}{(1-\beta)(1-\epsilon\sigma)^2}\sum_{i=1}^n\|\boldsymbol{z}_{i,\,1}\|^2 + \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,1}) - \sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,K+1})\right]$$

$$c_2 = \frac{2G^2}{1-4n\sigma\epsilon}\left[\frac{nL^2}{2\epsilon} + \frac{(2-\epsilon\sigma)\gamma\beta n^2L^2}{\epsilon\sigma_\phi(1-\epsilon\sigma)(1-\beta)} + \frac{2[2(1-\epsilon\sigma)^2+1]\sigma\gamma^2\beta^2L^2n^3}{\epsilon\sigma_\phi^2(1-\epsilon\sigma)^2(1-\beta)^2} + \frac{\sigma n^2L^2\gamma^2}{\sigma_\phi^2(1-\beta)^2(1-\epsilon\sigma)^2}\right]$$

□

3 分布式零阶镜像下降算法

3.1 分布式算法设计

针对式 (2) 中的分布式弱凸优化问题, 考虑梯度信息不可精确获取的情形 (如目标函数的具体解析形式难以精确获取、目标函数梯度受到噪声干扰以及分布式系统中梯度信息涉及隐私数据而无法明确给出等). 在此背景下, 本文基于正交随机方向矩阵, 设计如下分布式零阶镜像下降算法.

算法 2. 分布式零阶镜像下降算法

输入. 初始化迭代次数 K, 步长 α_k;

初始化. $\boldsymbol{x}_{i,\,1}\in\mathbf{R}^m,\ \forall\, i\in\mathcal{V}$;

1: **for** $k=1,\ 2,\ \cdots,\ K-1$ **do**

2: 计算在 $\boldsymbol{x}_{i,\,k}$ 处的差分 $[\nabla_k f_i(\boldsymbol{x}_{i,\,k})]_j = \frac{1}{h_k}[f_i(\boldsymbol{x}_{i,\,k}+h_k\boldsymbol{P}_{i,\,k}^{(j)}) - f_i(\boldsymbol{x}_{i,\,k})]$;

3: 在距离生成函数 $\phi(\cdot)$ 的梯度域进行梯度下降 $\nabla\phi(\boldsymbol{y}_{i,\,k}) = \nabla\phi(\boldsymbol{x}_{i,\,k}) - \alpha_k\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})$;

4: 关于距离函数 $V_\phi(\cdot,\ \cdot)$ 求解邻近投影 $\boldsymbol{z}_{i,\,k+1} = \arg\min_{\boldsymbol{x}_o\in\mathcal{X}} V_\phi(\boldsymbol{x}_o,\ \boldsymbol{y}_{i,\,k})$;

5: 智能体与其邻居交互信息并加权平均

$\boldsymbol{x}_{i,\,k+1} = \sum_{j\in\mathcal{N}_{i,\,k+1}} w_{ij,\,k+1}\boldsymbol{z}_{j,\,k+1}$;

6: **end for**

返回. $\boldsymbol{x}_{i,\,K}$.

注 7. 在目标函数的梯度信息难以精确获取的情形下, 有多种处理方式, 如随机梯度法、有限差分法等. 算法 2 以有限差分法为例给出零阶镜像下降法的设计框架. 相较于算法 1, 算法 2 弱化了对目标函数梯度信息的依赖. 具体地, 算法 2 采用正交随机方向矩阵序列 $\{\boldsymbol{P}_{i,\,k}\}\in\mathbf{R}^{m\times l}$ 中的列向量 $\boldsymbol{P}_{i,\,k}^{(j)}$ 作为局部决策变量的摄动向量, 并求取目标函数在 $\boldsymbol{x}_{i,\,k}+h_k\boldsymbol{P}_{i,\,k}^{(j)}$ 和 $\boldsymbol{x}_{i,\,k}$ 两点处的差分, 差分因子为 h_k. 将差分结果作为估计向量 $\nabla_k f_i(\boldsymbol{x}_{i,\,k})\in\mathbf{R}^l$ 的第 j 个分量. 最终, 取 $\boldsymbol{P}_{i,\,k}\nabla f_i(\boldsymbol{x}_{i,\,k})\in\mathbf{R}^m$ 为对真实梯度信息 $\nabla f_i(\boldsymbol{x}_{i,\,k})$ 的估计.

针对算法 2 中的正交随机方向矩阵 $\boldsymbol{P}_{i,\,k}$, 给出如下假设.

假设 6. 每个随机矩阵 $\boldsymbol{P}_{i,\,k}\in\mathbf{R}^{m\times l}$, $\forall\, i\in\mathcal{V}$, $k\in\mathbf{N}_+$ 满足 $\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\boldsymbol{P}_{i,\,k}\overset{\text{a.s.}}{=}\frac{m}{l}\boldsymbol{I}_l$, $\mathrm{E}\{\boldsymbol{P}_{i,\,k}\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\}=\boldsymbol{I}_m$, 其中, a.s. 表示几乎必然 (Almost surely).

注 8. 满足假设 6 的正交随机方向矩阵 $\boldsymbol{P}_{i,\,k}$ 可通过如下两种方案生成: 1) 从 m 维单位矩阵中随机选取 l 列构成矩阵 $\hat{\boldsymbol{P}}_{i,\,k}\in\mathbf{R}^{m\times l}$, 随机生成对角线元素为 ±1 的对角矩阵 $\boldsymbol{D}_{i,\,k}\in\mathbf{R}^{l\times l}$, 则可构造 $\boldsymbol{P}_{i,\,k}=\sqrt{m/l}\hat{\boldsymbol{P}}_{i,\,k}\boldsymbol{D}_{i,\,k}$; 2) 由标准正态分布 N(0, 1) 随机生成矩阵 $\boldsymbol{Z}_{i,\,k}\in\mathbf{R}^{m\times m}$, 对矩阵 $\boldsymbol{Z}_{i,\,k}$ 进行 QR 分解获得 $\boldsymbol{Z}_{i,\,k}=\boldsymbol{Q}_{i,\,k}\boldsymbol{R}_{i,\,k}$, 截取单位矩阵 $\boldsymbol{I}$ 的前 l 列可得矩阵 $\hat{\boldsymbol{D}}\in\mathbf{R}^{m\times l}$, 则可构造 $\boldsymbol{P}_{i,\,k}=\sqrt{m/l}\boldsymbol{Q}_{i,\,k}\hat{\boldsymbol{D}}$.

算法 2 中的估计方法对目标函数梯度信息的估计效果可概括如下.

引理 5. 若假设 3 和假设 6 成立, 则有以下结论:

1) $\mathrm{E}\{\|\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})\|\}\le\frac{\lambda m h_k}{2\sqrt{l}}$;

2) $\mathrm{E}\left\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})\|^2\right\}\le 2B_k^2$, 其中, $B_k^2=\frac{\lambda^2h_k^2m^3}{4l^2}+L^2$.

引理 5 的证明见附录 B.

注 9. 相较于传统基于随机向量的梯度信息估计方法, 算法 2 采用基于正交随机矩阵 $\boldsymbol{P}\in\mathbf{R}^{l\times m}$ 的估计策略. 该算法的核心思想是通过 l 个正交方向激发梯度信息模态, 从而实现对梯度信息的更精确估计. 依据引理 5, 正交随机方向矩阵 $\boldsymbol{P}_{i,\,k}$ 的列数 l 越大, 梯度信息估计误差期望值 $\mathrm{E}\{\|\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})\|\}$和梯度信息期望值 $\mathrm{E}\{\|\nabla_k f_i(\boldsymbol{x}_{i,\,k})\|\}$ 的上界越小. 特别地, 现有基于随机向量的估计方法正是本文提出的梯度估计方法在 $l=1$ 时的特殊情况[21].

3.2 算法收敛性分析

针对算法 2, 其收敛性能总结如下.

定理 2. 假设 1 ~ 6 成立, 令 $\alpha_k=1/\sqrt{k}$, $\forall\, k\in[K]$, 若所有智能体 $i\in\mathcal{V}$ 依据算法 2 生成 $\boldsymbol{x}_{i,\,k}$, 则有

$$\inf_{1\le k\le K}\frac{1}{\epsilon^2}\|\nabla\phi(\bar{\boldsymbol{x}}_k)-\nabla\phi(\check{\boldsymbol{x}}_k)\|^2\le\frac{c_3}{\sqrt{K}}+\frac{c_4\ln K}{\sqrt{K}}$$

其中,

$$\begin{aligned}c_3=&\frac{2G^2}{1-4n\epsilon(\sigma+1)}\Bigg[\frac{2(\sigma+1)n^2\gamma^2[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2(1-\beta)}\times\\&\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|^2+\\&\frac{n\sigma\gamma^2}{(1-\epsilon\sigma)^2(1-\beta)}\sum_{i=1}^n\|\boldsymbol{z}_{i,\,1}\|^2+\\&\frac{nL\gamma(2-\epsilon\sigma)}{\epsilon(1-\epsilon\sigma)(1-\beta)}\sum_{i=1}^n\|\boldsymbol{x}_{i,\,1}\|+\\&\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,1})-\sum_{i=1}^n f_\epsilon(\boldsymbol{z}_{i,\,K+1})\Bigg]\\c_4=&\frac{2G^2}{1-4n\epsilon(\sigma+1)}\Bigg[\frac{n^2B_1^2}{\epsilon}+\frac{n\lambda^2m^3}{8\epsilon l^2}+\\&\frac{2[2(1-\epsilon\sigma)^2+1](\sigma+1)\gamma^2\beta^2B_1^2n^3}{\epsilon\sigma_\phi^2(1-\epsilon\sigma)^2(1-\beta)^2}+\\&\frac{(2-\epsilon\sigma)\gamma B_1\beta n^2L}{\epsilon\sigma_\phi(1-\epsilon\sigma)(1-\beta)}+\frac{2\sigma n^2B_1^2\gamma^2}{\sigma_\phi^2(1-\beta)^2(1-\epsilon\sigma)^2}\Bigg]\end{aligned}$$

证明. 根据算法 2 以及参考定理 1 证明过程中的式 (6) ~ (8), 有

$$\begin{aligned}&\sum_{i=1}^n\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\boldsymbol{x}_{i,\,k}-\hat{\boldsymbol{x}}_{i,\,k}\rangle\le\\&\quad\frac{1}{\alpha_k}\Bigg[\sum_{i=1}^n V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})-\sum_{i=1}^n V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{z}_{i,\,k+1})+\\&\quad\sum_{i=1}^n V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})\Bigg]\end{aligned}\tag{28}$$

针对 $V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})$, 有

$$\begin{aligned}&V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})+V_\phi(\boldsymbol{y}_{i,\,k},\,\boldsymbol{x}_{i,\,k})\overset{(a)}{=}\\&\quad\langle\nabla\phi(\boldsymbol{x}_{i,\,k})-\nabla\phi(\boldsymbol{y}_{i,\,k}),\,\boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\rangle\overset{(b)}{=}\\&\quad\langle\alpha_k\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\rangle\end{aligned}\tag{29}$$

其中, $\overset{(a)}{=}$ 是根据 $V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})$ 的定义所得, $\overset{(b)}{=}$ 是根据算法 2 所得. 对式 (29) 取期望, 得

$$\mathrm{E}\{V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})\}+\mathrm{E}\{V_\phi(\boldsymbol{y}_{i,\,k},\,\boldsymbol{x}_{i,\,k})\}\;=\;\mathrm{E}\left\{\langle\alpha_k\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\rangle\right\}\overset{(a)}{\leq}B_k^2\alpha_k^2+\frac{1}{2}\mathrm{E}\{\|\boldsymbol{x}_{i,\,k}-\boldsymbol{y}_{i,\,k}\|^2\}\tag{30}$$

其中, $\overset{(a)}{\leq}$ 源于 Young 不等式和引理 5. 因此, 参考定理 1 证明过程中的式 (10), 可得

$$\mathrm{E}\{V_\phi(\boldsymbol{x}_{i,\,k},\,\boldsymbol{y}_{i,\,k})\}\leq B_k^2\alpha_k^2\tag{31}$$

进一步, 对式 (28) 取期望并结合式 (31), 有

$$\sum_{i=1}^{n}\mathrm{E}\{\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle\}\leq\frac{1}{\alpha_k}\sum_{i=1}^{n}\mathrm{E}\left\{V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})-V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{z}_{i,\,k+1})\right\}+nB_k^2\alpha_k\tag{32}$$

针对 $\left\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\right\rangle$, 有

$$\begin{aligned}&\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle\;=\\&\langle\boldsymbol{P}_{i,\,k}\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})-\nabla f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle+\\&\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle+\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\\&\boldsymbol{P}_{i,\,k}\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle\end{aligned}\tag{33}$$

对式 (33) 两边同时乘以 α_k 并取期望, 有

$$\begin{aligned}&\alpha_k\mathrm{E}\{\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle\}\overset{(a)}{=}\\&\alpha_k\mathrm{E}\{\langle\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k}),\\&\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle\}+\\&\alpha_k\mathrm{E}\{\langle\nabla f_i(\boldsymbol{x}_{i,\,k}),\,\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\rangle\}\overset{(b)}{\leq}\\&\alpha_k\mathrm{E}\left\{f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\boldsymbol{x}_{i,\,k})+\frac{\sigma}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2\right\}+\\&\frac{\lambda m\sqrt{m}\alpha_k h_k}{2l}\mathrm{E}\left\{\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|\right\}\overset{(c)}{\leq}\\&\alpha_k\mathrm{E}\left\{f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\boldsymbol{x}_{i,\,k})+\frac{\sigma}{2}\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2\right\}+\\&\frac{h_k^2}{2}\mathrm{E}\left\{\|\hat{\boldsymbol{x}}_{i,\,k}-\boldsymbol{x}_{i,\,k}\|^2\right\}+\frac{\lambda^2m^3\alpha_k^2}{8l^2}\end{aligned}\tag{34}$$

其中, $\overset{(a)}{=}$ 是根据假设 6 所得, $\overset{(b)}{\leq}$ 是基于引理 5 和目标函数 $f_i(\cdot)$ 的弱凸性所得, $\overset{(c)}{\leq}$ 是源于 Young 不等式. 结合式 (32) 和式 (34), 可得

$$\begin{aligned}&\sum_{i=1}^{n}\mathrm{E}\left\{V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{z}_{i,\,k+1})\right\}\;\leq\\&\frac{n\lambda^2m^3\alpha_k^2}{8l^2}+nB_k^2\alpha_k^2+\\&\alpha_k\sum_{i=1}^{n}\mathrm{E}\left\{f_i(\hat{\boldsymbol{x}}_{i,\,k})-f_i(\boldsymbol{x}_{i,\,k})\right\}+\\&\frac{\sigma\alpha_k+h_k^2}{2}\|\breve{\boldsymbol{x}}_k-\boldsymbol{x}_{i,\,k}\|^2+\sum_{i=1}^{n}\mathrm{E}\left\{V_\phi(\breve{\boldsymbol{x}}_k,\,\boldsymbol{x}_{i,\,k})\right\}\end{aligned}$$

参考定理 1 证明过程中的式 (14) ~ (17), 有

$$\begin{aligned}&\sum_{i=1}^{n}\mathrm{E}\left\{V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{z}_{i,\,k+1})\right\}\;\leq\frac{n\lambda^2m^3\alpha_k^2}{8l^2}+nB_k^2\alpha_k^2+\\&(2n\sigma\alpha_k+2nh_k^2-\frac{\alpha_k}{2\epsilon})\mathrm{E}\{\|\breve{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2\}+\\&\frac{\alpha_k(\sigma+h_k^2)[2(1-\epsilon\sigma)^2+1]}{(1-\epsilon\sigma)^2}\sum_{i=1}^{n}\mathrm{E}\{\|\boldsymbol{x}_{i,k}-\bar{\boldsymbol{x}}_k\|^2\}+\\&\frac{L(2-\epsilon\sigma)\alpha_k}{1-\epsilon\sigma}\sum_{i=1}^{n}\mathrm{E}\{\|\boldsymbol{x}_{i,k}-\bar{\boldsymbol{x}}_k\|\}+\\&\sum_{i=1}^{n}\mathrm{E}\left\{V_\phi(\hat{\boldsymbol{x}}_{i,k},\boldsymbol{x}_{i,k})\right\}\end{aligned}$$

进一步, 根据式 (18), 有

$$\begin{aligned}&\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k+1})\leq\sum_{i=1}^{n}\left[f(\hat{\boldsymbol{x}}_{i,\,k})+\frac{1}{\epsilon}V_\phi(\hat{\boldsymbol{x}}_{i,\,k},\,\boldsymbol{x}_{i,\,k})\right]+\\&\frac{1}{\epsilon}\left[nB_k^2\alpha_k^2+\left(2n\sigma\alpha_k+2nh_k^2-\frac{\alpha_k}{2\epsilon}\right)\|\breve{\mathbf{x}}_k-\bar{\mathbf{x}}_k\|^2+\right.\\&\alpha_k\sum_{i=1}^{n}\left(\frac{(\sigma+h_k^2)[2(1-\epsilon\sigma)^2+1]}{(1-\epsilon\sigma)^2}\|\mathbf{x}_{i,k}-\bar{\mathbf{x}}_k\|^2+\right.\\&\left.\left.\frac{L(2-\epsilon\sigma)}{1-\epsilon\sigma}\|\bar{\mathbf{x}}_k-\mathbf{x}_{i,k}\|\right)+\frac{n\lambda^2m^3\alpha_k^2}{8l^2}\right]\end{aligned}$$

结合定理 1 证明过程中的式 (21), 可得

$$\begin{aligned}&\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k+1})\leq\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k})+\frac{nB_k^2\alpha_k^2}{\epsilon}+\\&\frac{n\lambda^2m^3\alpha_k^2}{8\epsilon l^2}+\sum_{i=1}^{n}\left[\frac{L(2-\epsilon\sigma)\alpha_k}{\epsilon(1-\epsilon\sigma)}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\right.\\&\left.\frac{\alpha_k(\sigma+h_k^2)[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|^2\right]+\\&\frac{\alpha_k}{\epsilon}\left(2n\sigma+2n\frac{h_k^2}{\alpha_k}-\frac{1}{2\epsilon}\right)\|\breve{\boldsymbol{x}}_k-\bar{\boldsymbol{x}}_k\|^2+\\&\frac{\sigma}{2(1-\epsilon\sigma)^2}\sum_{i=1}^{n}\sum_{j=1}^{n-1}\sum_{l=j+1}^{n}w_{ij,\,k}w_{il,\,k}\|\boldsymbol{z}_{j,\,k}-\\&\boldsymbol{z}_{l,\,k}\|^2\leq\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k})+\frac{n\lambda^2m^3\alpha_k^2}{8\epsilon l^2}+\frac{nB_k^2\alpha_k^2}{\epsilon}+\tilde{b}_k\end{aligned}\tag{35}$$

其中,

$$\tilde{b}_k = \sum_{i=1}^{n}\left[\frac{L(2-\epsilon\sigma)\alpha_k}{\epsilon(1-\epsilon\sigma)}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|+\frac{\alpha_k(\sigma+h_k^2)[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2}\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|^2\right]+\frac{\sigma}{2(1-\epsilon\sigma)^2}\sum_{i=1}^{n}\sum_{j=1}^{n-1}\sum_{l=j+1}^{n}w_{ij,\,k}w_{il,\,k}\|\boldsymbol{z}_{j,\,k}-\boldsymbol{z}_{l,\,k}\|^2$$

根据引理 2 和引理 5, 局部决策变量 $\boldsymbol{z}_{i,\,k}$ 和 $\boldsymbol{x}_{i,\,k}$ 具有如下一致性结果:

$$\|\boldsymbol{z}_{i,\,k}-\bar{\boldsymbol{z}}_k\|\le\gamma\beta^{k-1}\sum_{i=1}^{n}\|\boldsymbol{z}_{i,\,1}\|+\frac{\sqrt{2}n\gamma B_k}{\sigma_\phi}\sum_{\kappa=1}^{k-1}\beta^{k-1-\kappa}\alpha_\kappa$$

$$\|\boldsymbol{x}_{i,\,k}-\bar{\boldsymbol{x}}_k\|\le\gamma\beta^{k-1}\sum_{i=1}^{n}\|\boldsymbol{x}_{i,\,1}\|+\frac{n\gamma B_k}{\sigma_\phi}\sum_{\kappa=1}^{k-1}\beta^{k-\kappa}\alpha_\kappa$$

参考定理 1 证明过程中的式 (25) 和式 (26), 有

$$\tilde{b}_k\le r_3\alpha_k^2+r_4\beta^{k-1}$$

其中,

$$r_3=\frac{2[2(1-\epsilon\sigma)^2+1](\sigma+h_1^2)\alpha_1\gamma^2\beta^2B_1^2n^3}{\epsilon\sigma_\phi^2(1-\epsilon\sigma)^2(1-\beta)^2}+\frac{(2-\epsilon\sigma)\gamma B_1\beta n^2L}{\epsilon\sigma_\phi(1-\epsilon\sigma)(1-\beta)}+\frac{2\sigma n^2B_1^2\gamma^2}{\sigma_\phi^2(1-\beta)^2(1-\epsilon\sigma)^2}$$

$$r_4=\frac{n\sigma\gamma^2}{(1-\epsilon\sigma)^2}\sum_{i=1}^{n}\|\boldsymbol{z}_{i,\,1}\|^2+\frac{nL\gamma\alpha_1(2-\epsilon\sigma)}{\epsilon(1-\epsilon\sigma)}\sum_{i=1}^{n}\|\boldsymbol{x}_{i,\,1}\|+\frac{2(\sigma+h_1^2)\alpha_1n^2\gamma^2[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2}\sum_{i=1}^{n}\|\boldsymbol{x}_{i,\,1}\|^2$$

进一步, 整理不等式 (35), 可得

$$\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k+1})-\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k})\le\frac{n\lambda^2m^3\alpha_k^2}{8\epsilon l^2}+\frac{nB_k^2\alpha_k^2}{\epsilon}+\tilde{b}_k$$

此外, 不等式 (35) 可以写为

$$\frac{1}{\epsilon}\left(\frac{1}{2\epsilon}-2n\sigma-2n\frac{h_k^2}{\alpha_k}\right)\alpha_k\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2\le\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k})-\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,k+1})+\frac{n\lambda^2m^3\alpha_k^2}{8\epsilon l^2}+\tilde{b}_k+\frac{nB_k^2\alpha_k^2}{\epsilon}\tag{36}$$

针对式 (36), 关于 k 从 1 到 K 求和, 有

$$\sum_{k=1}^{K}\frac{1}{\epsilon}\left(\frac{1}{2\epsilon}-2n\sigma-2n\frac{h_k^2}{\alpha_k}\right)\alpha_k\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2\le\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,1})-\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,K+1})+\sum_{k=1}^{K}\tilde{b}_k+\sum_{k=1}^{K}\left(\frac{n\lambda^2m^3\alpha_k^2}{8\epsilon l^2}+\frac{nB_k^2\alpha_k^2}{\epsilon}\right)\tag{37}$$

在选取步长 $\alpha_k=1/\sqrt{k}$ 的基础上, 令差分因子满足 $h_k^2=\alpha_k$. 结合条件 $4n\epsilon(\sigma+1)\in(0,\,1)$, 式 (37) 可进一步化简为

$$\inf_{1\le k\le K}\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2\le\frac{2\epsilon^2}{1-4n\epsilon(\sigma+1)}\frac{1}{\sum\limits_{k=1}^{K}\alpha_k}\times\left[\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,1})-\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,K+1})+\sum_{k=1}^{K}\tilde{b}_k+\sum_{k=1}^{K}\left(\frac{n\lambda^2m^3}{8\epsilon l^2}+\frac{nB_k^2}{\epsilon}\right)\alpha_k^2\right]$$

当 K 足够大时, 有

$$\inf_{1\le k\le K}\frac{1}{\epsilon^2}\|\nabla\phi(\bar{\boldsymbol{x}}_k)-\nabla\phi(\check{\boldsymbol{x}}_k)\|^2\le\sum_{k=1}^{K}\frac{G^2}{\epsilon^2}\|\bar{\boldsymbol{x}}_k-\check{\boldsymbol{x}}_k\|^2\le\frac{c_3}{\sqrt{K}}+\frac{c_4\ln K}{\sqrt{K}}$$

其中,

$$c_3=\frac{2G^2}{1-4n\epsilon(\sigma+1)}\left[\frac{2(\sigma+1)n^2\gamma^2[2(1-\epsilon\sigma)^2+1]}{\epsilon(1-\epsilon\sigma)^2(1-\beta)}\times\sum_{i=1}^{n}\|\boldsymbol{x}_{i,\,1}\|^2+\frac{n\sigma\gamma^2}{(1-\epsilon\sigma)^2(1-\beta)}\sum_{i=1}^{n}\|\boldsymbol{z}_{i,\,1}\|^2+\frac{nL\gamma(2-\epsilon\sigma)}{\epsilon(1-\epsilon\sigma)(1-\beta)}\sum_{i=1}^{n}\|\boldsymbol{x}_{i,\,1}\|+\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,1})-\sum_{i=1}^{n}f_\epsilon(\boldsymbol{z}_{i,\,K+1})\right]$$

$$c_4=\frac{2G^2}{1-4n\epsilon(\sigma+1)}\left[\frac{n^2B_1^2}{\epsilon}+\frac{n\lambda^2m^3}{8\epsilon l^2}+\right.$$

$$\frac{2[2(1-\epsilon\sigma)^2+1](\sigma+1)\gamma^2\beta^2B_1^2n^3}{\epsilon\sigma_\phi^2(1-\epsilon\sigma)^2(1-\beta)^2}+$$

$$\frac{(2-\epsilon\sigma)\gamma B_1\beta n^2L}{\epsilon\sigma_\phi(1-\epsilon\sigma)(1-\beta)}+\frac{2\sigma n^2B_1^2\gamma^2}{\sigma_\phi^2(1-\beta)^2(1-\epsilon\sigma)^2}\Bigg]$$

□

4　仿真实验

为验证本文所提出算法的有效性和相关理论结果的正确性, 我们考虑如下基于相位恢复问题的数值仿真与实验:

$$\min_{\boldsymbol{x}\in\mathbf{R}^m} f(\boldsymbol{x})=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{q}\sum_{j=1}^{q}\left|\langle\boldsymbol{\omega}_{i,\,j},\,\boldsymbol{x}\rangle^2-\boldsymbol{s}_{i,\,j}\right|\right)$$

其中, $\boldsymbol{s}_{i,\,j}=\langle\boldsymbol{\omega}_{i,\,j},\,\boldsymbol{x}^*\rangle^2$. 在以下 3 个数值仿真和实验中, 网络化系统中的各节点均通过环形网络拓扑连接, 测量向量 $\boldsymbol{\omega}_{i,\,j}\in\mathbf{R}^m$ 的元素依据标准正态分布 $\mathrm{N}\,(0,\,1)$ 随机生成. 且距离生成函数 $\phi(\boldsymbol{u})$ 设定为 $\frac{1}{2}\|\boldsymbol{u}\|^2$, 即 Bregman 散度满足 $V_\phi(\boldsymbol{v},\,\boldsymbol{u})=\frac{1}{2}\|\boldsymbol{v}-\boldsymbol{u}\|^2$. 此外, $\phi(\boldsymbol{u})$ 的梯度为 $\nabla\phi(\boldsymbol{u})=\boldsymbol{u}$, 算法 1 和算法 2 中的 $\boldsymbol{y}_{i,\,k}$ 的计算式分别为 $\boldsymbol{y}_{i,\,k}=\boldsymbol{x}_{i,\,k}-\alpha_k\nabla f_i(\boldsymbol{x}_{i,\,k})$ 和 $\boldsymbol{y}_{i,\,k}=\boldsymbol{x}_{i,\,k}-\alpha_k\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})$.

例 1. 考虑如下数值仿真算例, 智能体个数 $n=5$, 变量维数 $m=5$, 单个智能体处理的测量数组数量 $q=10$. 真实信号 $\boldsymbol{x}^*$ 中的元素依据正态分布 $\mathrm{N}\,(5,\,1)$ 随机生成.

首先, 算法 1 的步长设计为 $\alpha_k=0.01/\sqrt{k}$; 算法 2 的步长和差分因子分别选取 $\alpha_k=0.01/\sqrt{k}$ 和 $h_k^2=0.01/\sqrt{k}$, 且正交随机方向数 $l=5$. 仿真结果如图 2 所示, 其中, log-dist$(f(\bar{\boldsymbol{x}}_k),\,f^*)$ 表示 $f(\bar{\boldsymbol{x}}_k)$ 和最优值 f^* 之间的距离的对数值. 仿真结果表明, 本文所提出的两种算法均具有目标函数值随迭代次数增加而渐近收敛至最优值的趋势.

此外, 为显示算法 2 中采用的梯度信息估计策略的有效性, 分别考虑 $l=4$、6、8 三种情况并与文献 [21] 中的传统基于随机向量的零阶优化算法进行比较. 算法相关参数选取为 $\alpha_k=0.01l/\sqrt{k}$ 且 $h_k^2=0.01/\sqrt{k}$. 相关仿真结果如图 3 所示. 根据仿真结果, 随着 l 的增大 (即梯度估计中正交方向数量的增加), 梯度估计的精度显著提高, 从而使得算法的收敛性能得到提升. 此外, 通过比较 $l=4$、6、8 与文献 [21] 中算法的仿真结果可知, 本文所提出的算法具有更好的收敛性能.

例 2. 为验证本文算法在图像恢复任务中的有效性, 本实验考虑基于 MNIST 数据集的数字图像实验, 并与文献 [21] 中的基于随机向量差分的分布式零阶算法进行对比. 具体而言, 考虑图像维度 $d=28\times28$, 智能体个数 $n=28$, 变量维度 $m=784$. 算法步长和差分因子的选取与例 1 相同, 算法 2 中的正交方向数 l 选为 20 和 784. 两个算法的总迭代步数均设为 500.

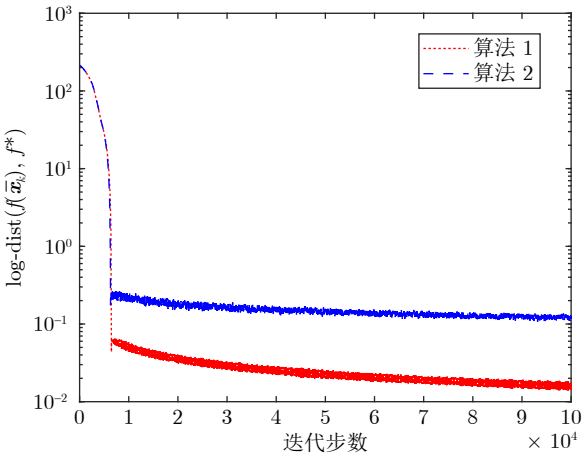


图 2　例 1 中算法 1 和算法 2 的仿真结果

Fig. 2　Simulation results of Algorithms 1 and 2 in Example 1

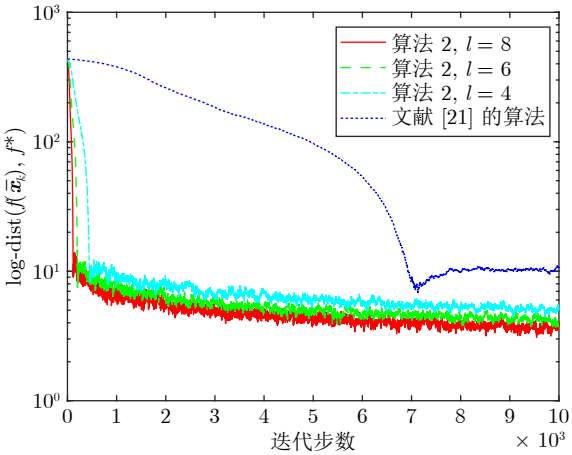


图 3　例 1 中算法 2 与文献 [21] 中算法的仿真结果

Fig. 3　Simulation results of Algorithm 2 and the algorithm from [21] in Example 1

实验结果如图 4 所示, 其中第 1 行从左到右分别是原始图像、算法的迭代初始值以及算法 1 的图像恢复效果; 第 2 行从左到右分别为文献 [21] 中算法以及算法 2 在 $l=20$ 和 $l=784$ 两种情形下的图像恢复效果. 实验结果显示, 文献 [21] 中的算法由于梯度信息激励不够导致恢复效果不理想; 算法 1 在获取真实梯度信息的情形下, 能够有效恢复图像; 当梯度信息不可精确获取时, 参数 l 的取值越大, 算法 2 的图像恢复效果越好.

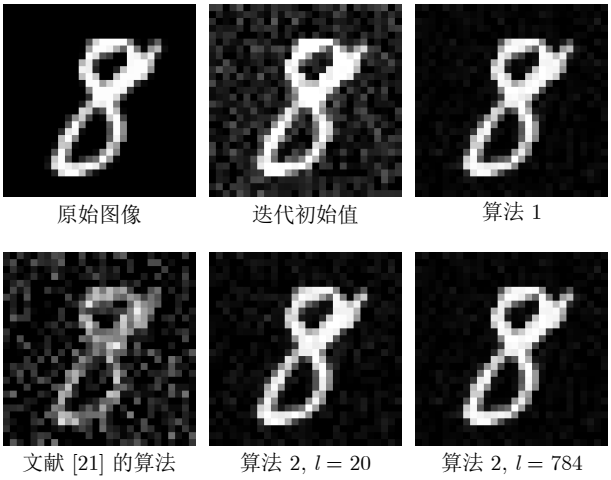


图 4 例 2 中算法 1 和算法 2 以及文献 [21] 中算法的图像恢复效果

Fig. 4 Image recovery results of Algorithms 1 and 2, and the algorithm from [21] in Example 2

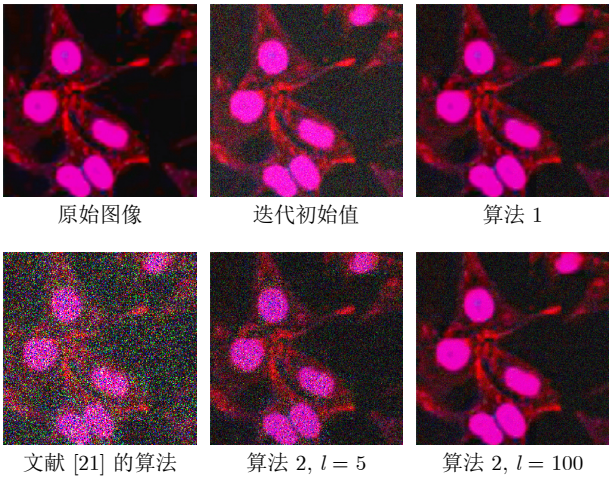


图 5 例 3 中算法 1 和算法 2 以及文献 [21] 中算法的图像恢复效果

Fig. 5 Image recovery results of Algorithms 1 and 2, and the algorithm from [21] in Example 3

例 3. 为进一步验证所提出的算法在处理更复杂相位恢复问题时的有效性, 本实验设计分辨率要求更高的彩色图像恢复实验. 所采用的图像为激光扫描显微镜下小鼠胰腺癌细胞中的核蛋白 HMGB1[30]. 为便于分布式存储与计算, 我们将原图像分为 64 个子块, 每个子块图像的像素数组 $\boldsymbol{X}^*$ 的维数为 $d = 32 \times 32 \times 3$. 将图像信息 $\boldsymbol{X}^*$ 转化为向量形式 $\boldsymbol{x}^* = \mathrm{Vec}(\boldsymbol{X}^*) \in \mathbf{R}^m$, $m = 3\,072$. 考虑采用包含有 $n = 5$ 个计算节点的网络化系统实现该图像的相位恢复, 单个计算节点处理测量向量的个数为 $q = \lceil 3m/n \rceil = 1\,844$, 其中, $\lceil \cdot \rceil$ 表示向上取整. 算法 1 和算法 2 以及文献 [21] 中算法均选取 $\alpha_k = 0.01/\sqrt{k}$ 且 $h_k^2 = 10^{-7}/\sqrt{k}$. 算法 2 考虑正交随机方向数 l 为 5 和 100 两种情形.

图像恢复效果如图 5 所示, 其中第 1 行从左往右分别为原始图像、算法迭代的初始值以及算法 1 的恢复效果; 第 2 行从左向右分别为文献 [21] 中算法和算法 2 在 $l = 5$ 和 $l = 100$ 两种情形下的图像恢复效果. 实验结果表明, 本文提出的两种算法都能处理复杂的相位恢复问题. 特别地, 在梯度信息不可精确获取的情形下, 与文献 [21] 中算法相比, 本文提出的梯度估计方法随着正交方向数 l 的增加, 图像恢复效果呈现提升的趋势.

5 结束语

针对广泛存在的大规模弱凸优化问题, 本文设计分布式镜像下降算法, 从 Bregman-Moreau 包络的角度分析算法的收敛性. 此外, 考虑梯度信息无法精确获取, 提出基于正交随机方向矩阵的分布式零阶镜像下降算法. 特别地, 正交随机方向矩阵列数 l 越大, 梯度信息估计效果和算法收敛性能越好. 经分析, 本文所提出的一阶和零阶分布式镜像下降算法均能获得 $\mathrm{O}(\ln K/\sqrt{K})$ 的收敛速度.

附录 A 引理 2 的证明

证明. 定义 $\boldsymbol{e}_{i,\,k} = \boldsymbol{z}_{i,\,k+1} - \boldsymbol{x}_{i,\,k}$, 则

$$\boldsymbol{z}_{i,\,k+1} = \boldsymbol{x}_{i,\,k} + \boldsymbol{e}_{i,\,k} = \sum_{j=1}^{n} w_{ij,\,k}\boldsymbol{z}_{j,\,k} + \boldsymbol{e}_{i,\,k} = \sum_{j=1}^{n}[\boldsymbol{\Omega}_1^k]_{ij}\boldsymbol{z}_{j,\,1} + \sum_{\kappa=2}^{k+1}\sum_{j=1}^{n}[\boldsymbol{\Omega}_\kappa^k]_{ij}\boldsymbol{e}_{j,\,\kappa-1} \tag{A1}$$

其中, $\boldsymbol{\Omega}_\kappa^k = \boldsymbol{W}_k\boldsymbol{W}_{k-1}\cdots\boldsymbol{W}_\kappa$, $\kappa < k$, $\boldsymbol{\Omega}_\kappa^k = \boldsymbol{W}_k$, $\kappa = k$, $\boldsymbol{\Omega}_\kappa^k = \boldsymbol{I}_n$, $\kappa > k$. 基于式 (A1), 可知 $\bar{\boldsymbol{z}}_{k+1} = \bar{\boldsymbol{z}}_1 + \frac{1}{n}\sum_{\kappa=2}^{k+1}\sum_{j=1}^{n}\boldsymbol{e}_{j,\,\kappa-1}$. 参考文献 [16] 中引理 1 的证明过程, 有 $\|\boldsymbol{e}_{i,\,k}\| \le \frac{L}{\sigma_\phi}\alpha_k$. 从而, 根据文献 [31] 中的引理 3.2, 可知 $\boldsymbol{z}_{i,\,k}$ 的一致性误差有如下上界

$$\|\boldsymbol{z}_{i,\,k+1} - \bar{\boldsymbol{z}}_{k+1}\| \le \gamma\beta^k\sum_{i=1}^{n}\|\boldsymbol{z}_{i,\,1}\| + \frac{\gamma nL}{\sigma_\phi}\sum_{\kappa=2}^{k+1}\beta^{k-\kappa+1}\alpha_{\kappa-1} \tag{A2}$$

此外, 根据 $\boldsymbol{x}_{i,\,k+1} = \sum_{j=1}^{n} w_{ij,\,k+1}\boldsymbol{z}_{j,\,k+1}$ 以及 $\boldsymbol{z}_{i,\,k+1} = \boldsymbol{x}_{i,\,k} + \boldsymbol{e}_{i,\,k}$, 有

$$\begin{cases} \boldsymbol{x}_{i,\,k+1} = \displaystyle\sum_{j=1}^{n}[\boldsymbol{\Omega}_2^{k+1}]_{ij}\boldsymbol{x}_{j,\,1} + \sum_{\kappa=2}^{k+1}\sum_{j=1}^{n}[\boldsymbol{\Omega}_\kappa^{k+1}]_{ij}\boldsymbol{e}_{j,\,\kappa-1} \\ \bar{\boldsymbol{z}}_{k+1} = \bar{\boldsymbol{z}}_1 + \dfrac{1}{n}\displaystyle\sum_{\kappa=2}^{k+1}\sum_{j=1}^{n}\boldsymbol{e}_{j,\,\kappa-1} \end{cases} \tag{A3}$$

参考式 (A2), 可知

$$\|\boldsymbol{x}_{i,\,k+1}-\bar{\boldsymbol{x}}_{k+1}\|\leq \gamma\beta^{k}\sum_{i=1}^{n}\|\boldsymbol{x}_{i,\,1}\|+\frac{\gamma nL}{\sigma_{\phi}}\sum_{\kappa=2}^{k+1}\beta^{k-\kappa+2}\alpha_{\kappa-1}\tag{A4}$$

□

附录 B　引理 5 的证明

证明. 参考文献 [26] 中的引理 5, 可得

$$\mathrm{E}\left\{\|\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})\|\right\}\leq\frac{\lambda m h_k}{2\sqrt{l}}\tag{B1}$$

此外, 针对 $\mathrm{E}\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})\|^2\}$, 有

$$\begin{aligned}&\mathrm{E}\left\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})\|^2\right\}=\\&\quad\mathrm{E}\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\nabla f_i(\boldsymbol{x}_{i,\,k})+\\&\quad\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\}\overset{(a)}{\leq}\\&\quad2\mathrm{E}\left\{\|\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\right\}+\\&\quad2\mathrm{E}\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\}\overset{(b)}{\leq}\\&\quad2\mathrm{E}\left\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\right\}+2L^2\end{aligned}\tag{B2}$$

其中, $\overset{(a)}{\leq}$ 是基于 $\|\boldsymbol{a}-\boldsymbol{b}\|^2\leq2\|\boldsymbol{a}\|^2+2\|\boldsymbol{b}\|^2$ 所得, $\overset{(b)}{\leq}$ 是根据假设 3 所得. 针对 $\mathrm{E}\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\}$, 有

$$\begin{aligned}&\mathrm{E}\left\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\right\}\overset{(a)}{=}\\&\quad\mathrm{E}\left\{\|\boldsymbol{P}_{i,\,k}\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\right\}=\\&\quad\mathrm{E}\{\langle\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k}),\\&\quad\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\boldsymbol{P}_{i,\,k}[\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})]\rangle\}\overset{(b)}{=}\\&\quad\frac{m}{l}\mathrm{E}\{\|\nabla_k f_i(\boldsymbol{x}_{i,\,k})-\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\nabla f_i(\boldsymbol{x}_{i,\,k})\|^2\}\overset{(c)}{\leq}\\&\quad\frac{\lambda^2m^3h_k^2}{4l^2}\end{aligned}\tag{B3}$$

其中, $\overset{(a)}{=}$ 和 $\overset{(b)}{=}$ 分别由假设 6 中 $\mathrm{E}\{\boldsymbol{P}_{i,\,k}\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\}=\boldsymbol{I}_m$ 和 $\boldsymbol{P}_{i,\,k}^{\mathrm{T}}\boldsymbol{P}_{i,\,k}=\frac{m}{l}\boldsymbol{I}_l$ 所得, $\overset{(c)}{\leq}$ 是根据式 (B1) 所得. □

References

1 Yi Peng, Hong Yi-Guang. Distributed cooperative optimization and its applications. *Scientia Sinica Mathematica*, 2016, **46**(10): 1547−1564
(衣鹏, 洪奕光. 分布式合作优化及其应用. 中国科学: 数学, 2016, **46**(10): 1547−1564)

2 Xie Pei, You Ke-You, Hong Yi-Guang, Xie Li-Hua. A survey of distributed convex optimization algorithms over networks. *Control Theory and Applications*, 2018, **35**(7): 918−927
(谢佩, 游科友, 洪奕光, 谢立华. 网络化分布式凸优化算法研究进展. 控制理论与应用, 2018, **35**(7): 918−927)

3 Nedić A, Liu J. Distributed optimization for control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2018, **1**(1): 77−103

4 Yang T, Yi X L, Wu J F, Yuan Y, Wu D, Meng Z Y, et al. A survey of distributed optimization. *Annual Reviews in Control*, 2019, **47**: 278−305

5 Yang Tao, Chai Tian-You. Research status and prospects of distributed collaborative optimization. *Scientia Sinica Technologica*, 2020, **50**(11): 1414−1425
(杨涛, 柴天佑. 分布式协同优化的研究现状与展望. 中国科学: 技术科学, 2020, **50**(11): 1414−1425)

6 Deng Wen, Li Wei-Jian, Zeng Xian-Lin, Hong Yi-Guang. A survey of distributed algorithms for solving matrix equations. *Control Theory and Applications*, 2021, **38**(11): 1695−1706
(邓文, 李伟健, 曾宪琳, 洪奕光. 矩阵方程的分布式求解算法研究概述. 控制理论与应用, 2021, **38**(11): 1695−1706)

7 Yang Tao, Xu Lei, Yi Xin-Lei, Zhang Sheng-Jun, Chen Rui-Juan, Li Yu-Zhe. Event-triggered distributed optimization algorithms. *Acta Automatica Sinica*, 2022, **48**(1): 133−143
(杨涛, 徐磊, 易新蕾, 张圣军, 陈蕊娟, 李渝哲. 基于事件触发的分布式优化算法. 自动化学报, 2022, **48**(1): 133−143)

8 Yang S F, Liu Q S, Wang J. A multi-agent system with a proportional-integral protocol for distributed constrained optimization. *IEEE Transactions on Automatic Control*, 2016, **62**(7): 3461−3467

9 Zhu Y N, Yu W W, Wen G H, Chen G R. Projected primal-dual dynamics for distributed constrained nonsmooth convex optimization. *IEEE Transactions on Cybernetics*, 2018, **50**(4): 1776−1782

10 Yuan K, Ling Q, Yin W T. On the convergence of decentralized gradient descent. *SIAM Journal on Optimization*, 2016, **26**(3): 1835−1854

11 Pu S, Shi W, Xu J M, Nedić A. Push-pull gradient methods for distributed optimization in networks. *IEEE Transactions on Automatic Control*, 2021, **66**(1): 1−16

12 Nedić A, Ozdaglar A, Parrilo P A. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 2010, **55**(4): 922−938

13 Lei J L, Chen H F, Fang H T. Primal-dual algorithm for distributed constrained optimization. *Systems and Control Letters*, 2016, **96**: 110−117

14 Cheng S S, Liang S, Fan Y, Hong Y G. Distributed gradient tracking for unbalanced optimization with different constraint sets. *IEEE Transactions on Automatic Control*, 2023, **68**(6): 3633−3640

15 Beck A, Teboulle M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 2003, **31**(3): 167−175

16 Yuan D M, Hong Y G, Ho D W, Jiang G P. Optimal distributed stochastic mirror descent for strongly convex optimization. *Automatica*, 2018, **90**: 196−203

17 Yang Y, Jia Q S, Xu Z B, Guan X H, Spanos C J. Proximal ADMM for nonconvex and nonsmooth optimization. *Automatica*, 2022, **146**: Article No. 110551

18 Hou J, Zeng X L, Wang G, Sun J, Chen J. Distributed momentum-based Frank-Wolfe algorithm for stochastic optimization. *IEEE/CAA Journal of Automatica Sinica*, 2022, **10**(3): 685−699

19 Chen S X, Garcia A, Shahrampour S. On distributed nonconvex optimization: Projected subgradient method for weakly convex problems in networks. *IEEE Transactions on Automatic Control*, 2021, **67**(2): 662−675

20 Wei M L, Yu W W, Liu H Z, Xu Q. Distributed weakly convex optimization under random time-delay interference. *IEEE*

Transactions on Network Science and Engineering, 2024, **11**(1): 212−224

21 Wang Y H, Zhao W X, Hong Y G, Zamani M. Distributed subgradient-free stochastic optimization algorithm for nonsmooth convex functions over time-varying networks. *SIAM Journal on Control and Optimization*, 2019, **57**(4): 2821−2842

22 Pang Y P, Hu G Q. Randomized gradient-free distributed optimization methods for a multiagent system with unknown cost function. *IEEE Transactions on Automatic Control*, 2019, **65**(1): 333−340

23 Pang Y P, Hu G Q. Gradient-free distributed optimization with exact convergence. *Automatica*, 2022, **144**: Article No. 110474

24 Yi X L, Zhang S J, Yang T, Johansson K H. Zeroth-order algorithms for stochastic distributed nonconvex optimization. *Automatica*, 2022, **142**: Article No. 110353

25 Xu L, Yi X L, Deng C, Shi Y, Chai T Y, Yang T. Quantized zeroth-order gradient tracking algorithm for distributed nonconvex optimization under Polyak-Łojasiewicz condition. *IEEE Transactions on Cybernetics*, 2024, **54**(10): 5746−5758

26 Wang R Y, Fan Y, Cheng S S. Zeroth-order algorithm design with orthogonal direction for distributed weakly convex optimization. In: Proceedings of the 63rd IEEE Conference on Decision and Control (CDC2024). Milan, Italy: IEEE, 2024.

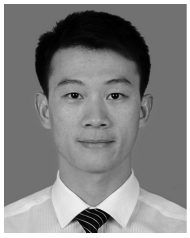
27 Juditsky A, Kwon J, Moulines É. Unifying mirror descent and dual averaging. *Mathematical Programming*, 2023, **199**: 793−830

28 Moreau J J. Proximitè et dualitè dans un espace hilbertien. *Bull. De La Sociètè Mathèmatique De France*, 1965, **93**: 273−299

29 Liu Hao-Yang, Hu Jiang, Li Yong-Feng, Wen Zai-Wen. *Optimization: Modeling, Algorithm and Theory*. Beijing: Higher Education Press, 2020.
(刘浩洋, 户将, 李勇锋, 文再文. 最优化: 建模、算法与理论. 北京: 高等教育出版社, 2020.)

30 Zhou W X, Zhou Y, Chen X L, Ning T T, Chen H Y, Guo Q, et al. Pancreatic cancer-targeting exosomes for enhancing immunotherapy and reprogramming tumor microenvironment. *Biomaterials*, 2021, **268**: Article No. 120546

31 Ram S S, Nedić A, Veeravalli V V. Distributed stochastic subgradient projection algorithms for convex optimization. *Journal of Optimization Theory and Applications*, 2010, **147**: 516−545



程松松 安徽大学电气工程与自动化学院副教授. 2018 年获得中国科学技术大学控制科学与工程专业博士学位. 主要研究方向为分布式计算, 优化与博弈.
E-mail: sscheng@ahu.edu.cn
(**CHENG Song-Song** Associate professor at the School of Electrical Engineering and Automation, Anhui University. He received his Ph.D. degree in control science and engineering from University of Science and Technology of China in 2018. His research interest covers distributed computation, optimization, and games.)

陈 茹 安徽大学电气工程与自动化学院硕士研究生. 主要研究方向为基于对偶理论的分布式非凸优化与博弈. E-mail: ruchen@stu.ahu.edu.cn
(**CHEN Ru** Master student at the School of Electrical Engineering and Automation, Anhui University. Her research interest covers distributed nonconvex optimization and game based on duality theory.)

樊 渊 安徽大学电气工程与自动化学院教授. 2011 年获得中国科学技术大学和香港城市大学控制科学与工程专业博士学位. 主要研究方向为分布式控制与优化, 机器人导航与控制. 本文通信作者.
E-mail: yuanf@ahu.edu.cn
(**FAN Yuan** Professor at the School of Electrical Engineering and Automation, Anhui University. He received his Ph.D. degree in control science and engineering from University of Science and Technology of China and the City University of Hong Kong in 2011. His research interest covers distributed control and optimization, and robot navigation and control. Corresponding author of this paper.)

邱剑彬 哈尔滨工业大学智能控制与系统研究所教授. 2009 年获得中国科学技术大学和香港城市大学机械电子工程专业博士学位. 主要研究方向为智能控制理论与应用, 人工智能大模型, 航天器智能自主控制, 智能机器人. E-mail: jbqiu@hit.edu.cn
(**QIU Jian-Bin** Professor at the Research Institute of Intelligent Control and Systems, Harbin Institute of Technology. He received his Ph.D. degree in mechatronics engineering from University of Science and Technology of China and the City University of Hong Kong in 2009. His research interest covers intelligent control theory and application, artificial intelligence large model, intelligent autonomous control of spacecraft, and intelligent robotics.)