中国科学 : 信息科学 2025 年 第 55 卷 第 7 期: 1687–1700
SCIENTIA SINICA Informationis
第二十七届中国科协年会学术论文
《中国科学》杂志社
SCIENCE CHINA PRESS
CrossMark
click for updates
工业智能系统及软件专刊 · 论文

分布式类 Nesterov 加速非凸复合优化算法

李天成, 张坤朋, 徐磊, 高超, 李凡, 杨涛*

东北大学流程工业综合自动化全国重点实验室, 沈阳 110819
* 通信作者. E-mail: yangtao@mail.neu.edu.cn

收稿日期: 2025–04–03; 修回日期: 2025–05–21; 接受日期: 2025–06–05; 网络出版日期: 2025–07–11

国家自然科学基金重点项目 (批准号: 62133003) 和国家重点研发计划课题 (批准号: 2022YFB3305904) 资助

摘要 随着工业智能系统的快速发展以及能源互联网的广泛应用, 工业软件在支持多节点协同计算和解决复杂优化问题中发挥了关键作用. 然而, 能源互联网的分布式、多节点自治特点以及复杂网络结构和实时动态调控需求, 对工业软件的实时性、收敛速率和适用范围提出了更高要求. 这些要求迫切需要设计高效的分布式优化算法作为工业软件的核心支撑. 因此, 本文研究了时变通信拓扑下的分布式非凸复合优化问题, 其中全局目标函数由光滑的非凸部分和非光滑的凸部分组成. 所提出的算法利用逐次凸逼近 (successive convex approximation, SCA) 技术与梯度跟踪机制, 并引入类 Nesterov 动量项以调整每次迭代的更新方向, 从而进一步提升算法的收敛速率. 理论证明了当动量参数低于设定的上界时, 所提算法在固定步长条件下能够渐近收敛至所研究问题的平衡点集. 数值仿真实验进一步验证了算法的有效性.

关键词 工业软件, 分布式非凸优化, 复合优化, 类 Nesterov 加速方法

1 引言

随着全球能源结构的转型和工业智能化的深度发展, 现代工业正迈向以智能化, 互联化为核心特征的工业 4.0 时代 [1,2]. 在这一背景下, 工业智能系统 [3,4] 通过融合物联网、人工智能和大数据分析, 不仅推动了制造业的智能化生产, 还在能源管理、交通运输和资源调度等领域具有广泛应用. 其中, 能源互联网作为工业智能系统的重要组成部分, 通过整合分布式清洁能源、智能电网和信息通信技术, 构建了 “横向多能互补, 纵向源 – 网 – 荷 – 储一体化协调” 的能源管理模式, 显著提升了能源利用效率并推动能源消费的绿色化与智能化 [5].

能源互联网具有显著的分布式特性和多节点自治特点, 其复杂的网络结构以及实时动态调控需求, 对支撑系统的工业软件设计提出了极高要求 [6,7]. 在能源互联网中, 多节点协同优化能力和实时响应性能是实现高效调度与稳定运行的关键. 为了满足能源互联网对于实时性和高效性的需求, 亟需设计高效的底层算法作为工业软件的核心支撑. 尤其是在处理多节点协同计算时, 工业软件必须设计能够

引用格式: 李天成, 张坤朋, 徐磊, 等. 分布式类 Nesterov 加速非凸复合优化算法. 中国科学: 信息科学, 2025, 55: 1687–1700, doi: 10.1360/SSI-2025-0112
Li T C, Zhang K P, Xu L, et al. Distributed Nesterov-like accelerated non-convex composite optimization algorithm. Sci Sin Inform, 2025, 55: 1687–1700, doi: 10.1360/SSI-2025-0112

适应复杂通信拓扑和动态网络环境的优化算法, 才能实现对能源互联网的高效调度和控制. 在这一背景下, 分布式优化算法因其利用节点间的信息交互与本地计算逐步优化全局目标的特点, 逐渐成为能源互联网领域的研究热点 [8~10]. 例如, 文献 [9,10] 所提算法通过节点间的信息交互与本地计算逐步优化系统的全局目标, 已广泛应用于能源互联网的区域自治优化调度和频率快速恢复, 在维持多能源系统实时供需平衡的同时有效降低全网的调节成本. 然而, 这些算法的收敛速度较慢, 难以满足大规模系统对实时性的要求. 为此, 设计能够适应复杂通信拓扑和动态网络环境的高效分布式优化算法, 已成为当前能源互联网领域的研究热点.

近年来, 分布式加速优化算法逐渐成为研究焦点. 文献 [11∼15] 分别提出了分布式比例积分算法、分布式零梯度和算法、加速型去中心化对偶平均算法、分布式精确一阶算法和 Nesterov 动量加速算法, 这些算法旨在加速求解全局目标函数的最优值点. 其中, Nesterov 动量法作为一种广泛应用的梯度加速技术, 能够显著提升算法的收敛速度, 并已被广泛应用于多种分布式算法中 [16~20]. 针对静态通信拓扑, 文献 [18] 提出了一种分布式 Nesterov 加速算法, 并针对强凸和一般凸的目标分别给出了不同的参数设定和理论收敛分析. 文献 [19] 利用 AB/Push-Pull 梯度跟踪技术 [21] 和 Nesterov 动量, 提出了 FROZEN 算法. 在此基础上, 针对时变通信拓扑, 文献 [20] 结合重球动量和 Nesterov 动量, 提出了一种新型的分布式加速优化算法.

上述研究均假设局部目标函数为凸函数, 然而在实际应用中, 局部目标函数也可能呈现光滑非凸 – 非光滑凸的复合形式. 这类具有复合形式的非凸优化问题广泛应用于能源互联网中的分布式经济调度、多能流协同优化和分布式储能优化调度等领域 [22~24]. 针对具有光滑非凸 – 非光滑凸复合结构的非凸优化问题, 文献 [25, 26] 分别提出了分布式近端梯度法 (distributed proximal gradient method, DPGM) 和分布式迭代软阈值算法 (distributed iterative soft-thresholding algorithm, DISTA). 这两种算法均包含光滑非凸部分的梯度下降和非光滑凸部分的近端算子操作, 并且都要求所有智能体的局部目标函数的非光滑凸部分相同. 针对静态通信拓扑, 文献 [27] 利用 SCA (successive convex approximation) 技术和梯度跟踪机制, 引入了类 Nesterov 动量项, 并证明了在局部目标函数为光滑非凸 – 非光滑凸的情况下, 所提出的算法能够渐近收敛至平衡点集. 针对时变通信拓扑, 文献 [28] 提出了基于 SCA 技术的 NEXT 算法, 并证明在衰减步长条件下, 该算法能够有效地处理具有光滑非凸 – 非光滑凸结构的非凸复合优化问题. 文献 [29] 在文献 [28] 的基础上提出了 SONATA 算法, 并证明了算法在固定步长条件下能够渐近收敛至平衡点集. 受文献 [28, 29] 及 Nesterov 动量方法 [20, 27] 的启发, 针对时变通信拓扑, 本文基于 SCA 技术和梯度跟踪机制, 引入了类 Nesterov 动量项, 设计了一种分布式加速非凸复合优化算法.

本文的主要贡献可以总结如下.

(1) 文献 [20] 研究的分布式加速算法要求全局目标函数同时具有强凸性和光滑性, 而本文则聚焦于具有光滑非凸 – 非光滑凸复合形式的优化问题, 其复杂性更高且适用范围更广. 相比文献 [28, 29], 本文通过引入类 Nesterov 动量项, 不仅加速了算法的收敛速度, 还严格证明了该算法在光滑非凸 – 非光滑凸复合优化问题中对平衡点集的渐近收敛性. 尽管文献 [27] 已结合类 Nesterov 动量法, 但其方法仅适用于静态通信拓扑, 而本文的研究则面向更具挑战性的时变通信拓扑.

(2) 针对处理光滑非凸 – 非光滑凸及约束优化问题可能面临的挑战 (例如当目标函数的 (次) 梯度无界时, 传统的优化误差和一致性误差分离处理方法 [28, 30] 不再适用), 本文结合加权平均动力学、一致性和梯度跟踪机制, 构建了一种新的李雅普诺夫 (Lyapunov) 函数, 并提出了相应的下降技术, 用于分析加速算法的收敛性. 值得注意的是, 本文提出的分布式加速非凸复合优化算法不需要目标函数的 (次) 梯度在约束集上一致有界, 从而显著降低了算法的适用条件限制.

符号说明. 在本文中, 非负自然数集 (正自然数集) 记为 $\mathbb{N}_+(\mathbb{N}_{++})$. 向量 $\mathbf{x}$ 被视为列向量; 矩阵用粗体字母表示. 本文在欧几里得 (Euclidean) 空间 $\mathbb{R}^m$ 上工作, 采用标准欧几里得范数, 记作 $\|\cdot\|$; 当 $\|\cdot\|$ 的运算对象为矩阵时, 默认使用谱范数. 对于其他类型的向量或矩阵范数, 例如 ℓ_1-范数或无穷范

数, 用 $\|\cdot\|_p$ 表示, 其中 p 表示对应的范数类型. 向量 $\mathbf{x}$ 的转置记为 $\mathbf{x}^{\mathrm{T}}$. 克罗内克 (Kronecker) 积用符号 $\otimes$ 表示. 本文用 $\mathbf{1}$ 表示所有元素均为 1 的向量, 用 $\mathbf{I}$ 表示单位矩阵; 在不引起歧义的情况下, 向量 $\mathbf{1}$ 和单位矩阵 $\mathbf{I}$ 的维度将根据上下文隐含确定而不显式标注. 给定 $I \in \mathbb{N}_{++}$, 定义集合 $[I] \triangleq \{1,\ldots,I\}$. $\lfloor\cdot\rfloor$ 表示向下取整.

2　问题描述和算法设计

2.1　问题描述

本文研究了多智能体网络下的分布式非凸复合优化问题, 该网络由 I 个智能体组成, 其目标是协作解决以下优化问题:

$$\min_{\mathbf{x}\in\mathcal{K}} U(\mathbf{x}) \triangleq \underbrace{\sum_{i=1}^{I} f_i(\mathbf{x})}_{F(\mathbf{x})} + G(\mathbf{x}), \tag{1}$$

其中 $\mathbf{x}$ 是全局决策变量, 局部目标函数 $f_i: \mathbb{R}^m \to \mathbb{R}$ 为智能体 $i \in \{1,\ldots,I\}$ 的光滑非凸代价函数, G 是一个非光滑凸函数, $\mathcal{K} \subseteq \mathbb{R}^m$ 是闭凸集. 每个智能体 i 仅知晓自身的局部目标函数 f_i (以及 G 和 $\mathcal{K}$), 并通过与邻居节点的信息交互, 协作求解问题 (1).

智能体之间通过通信网络进行信息交换. 在每个离散时间槽 n 下, 通信网络被建模为一个时变无向图 $\mathcal{G}^n = ([I], \mathcal{E}^n, \mathbf{A}^n)$. 其中, $[I] = \{1,\ldots,I\}$ 表示智能体集合; $\mathcal{E}^n$ 表示智能体之间的通信链路集合; 若 $(i,j) \in \mathcal{E}^n$, 则表示智能体 i 和 j 之间可以通信; $\mathbf{A}^n = \left(a_{ij}^n\right)_{i,j=1}^{I}$ 表示对应的加权邻接矩阵.

现对问题 (1) 作如下假设.

假设1　(1) $\mathcal{K}$ 是非空的闭凸集;

(2) 每个 f_i 在包含 $\mathcal{K}$ 的开集上属于 C^1 (一阶连续可微) 类函数;

(3) 每个 ∇f_i 在 $\mathcal{K}$ 上是 L_i-Lipschitz 连续的;

(4) $G: \mathcal{K} \to \mathbb{R}$ 是凸的;

(5) U 在集合 $\mathcal{K}$ 有下界.

假设图 $\{\mathcal{G}^n\}_{n\in N_+}$ 满足如下连通性条件.

假设2　图 $\{\mathcal{G}^n\}_{n\in N_+}$ 是 B 连通的, 即存在一个整数 $B > 0$, 使得对于所有 $k \geqslant 0$, 由边集 $\cup_{t=k}^{k+B-1}\mathcal{E}^t$ 构成的图是连通的.

加权矩阵 $\mathbf{A}^n = \left(a_{ij}^n\right)_{i,j=1}^{I}$ 是一个时变的、双随机且对称的权重矩阵, 其非零元素的分布与图 $\mathcal{G}^n$ 的拓扑结构相一致, 具体满足如下假设.

假设3　(1) $a_{ii}^n \geqslant \kappa > 0, \forall i \in [I]$;

(2) $a_{ij}^n \geqslant \kappa > 0$, 如果 $(i,j) \in \mathcal{E}^n$; 否则 $a_{ij}^n = 0$.

假设4　矩阵 $\mathbf{A}^n$ 是双随机且对称的, 即满足以下条件:

$$\mathbf{1}^{\mathrm{T}}\mathbf{A}^n = \mathbf{1}^{\mathrm{T}},\ \left(\mathbf{A}^n\right)^{\mathrm{T}} = \mathbf{A}^n. \tag{2}$$

2.2　算法设计

为求解问题 (1), 在设计分布式优化算法时, 主要面临两个挑战: (1) 函数 F 具有非凸性; (2) 各智能体无法获得关于函数 F 的全局信息. 为应对上述难题, 本文采用 SCA 技术 (步骤 1) 与梯度跟踪机制 (步骤 2) 相结合的分布式求解框架. 具体如下.

步骤1　由于 F 的非凸性, 直接求解问题 (1) 可能导致计算成本过高. 因此, 可以选择以某种适当方式对问题 (1) 进行近似. 具体地, 每个智能体 i 维护一个关于优化变量 $\mathbf{x}$ 的本地估计 $\mathbf{x}_{(i)}$, 并对 $F(\mathbf{x}_{(i)})=\sum_{j=1}^{I}f_j(\mathbf{x}_{(i)})$ 进行如下近似处理: (1) 在每次迭代 n 中, 将非凸的 $f_i(\mathbf{x}_{(i)})$ 替换为一个强凸的近似函数, 例如 $\widetilde{f}_i(\cdot;\mathbf{x}_{(i)}^n):\mathcal{K}\to\mathbb{R}$, 该函数依赖于当前迭代点 $\mathbf{x}_{(i)}^n$; (2) 将 $\sum_{j\neq i}f_j(\mathbf{x}_{(i)})$ 在 $\mathbf{x}_{(i)}^n$ 处线性化. 具体而言, 其 SCA 更新过程如下.

在每次迭代 n 时, 给定当前迭代点 $\mathbf{x}_{(i)}^n$, 每个智能体 i 通过求解下式更新其状态:

$$\widehat{\mathbf{x}}_{(i)}(\mathbf{x}_{(i)}^n)\triangleq\underset{\mathbf{x}_{(i)}\in\mathcal{K}}{\operatorname{argmin}}\left\{\widetilde{f}_i(\mathbf{x}_{(i)};\mathbf{x}_{(i)}^n)+\widehat{\pi}_{(i)}(\mathbf{x}_{(i)}^n)^{\mathrm{T}}(\mathbf{x}_{(i)}-\mathbf{x}_{(i)}^n)+G(\mathbf{x}_{(i)})\right\},\tag{3}$$

其中, $\widetilde{f}_i(\mathbf{x}_{(i)};\mathbf{x}_{(i)}^n)$ 可被视为函数 $f_i(\mathbf{x}_{(i)})$ 在点 $\mathbf{x}_{(i)}^n$ 处的一种局部强凸近似, 并同时保留 $f_i(\mathbf{x}_{(i)})$ 的一阶性质. 代理函数 $\widetilde{f}_i$ 的选取需满足命题 A1 (见附录 A) 中的条件, 其详细形式可参考文献 [27～29]. $\widehat{\pi}_{(i)}(\mathbf{x}_{(i)}^n)$ 是函数 $\sum_{j\neq i}f_j(\mathbf{x}_{(i)})$ 在 $\mathbf{x}_{(i)}^n$ 处的梯度, 即 $\widehat{\pi}_{(i)}(\mathbf{x}_{(i)}^n)=\sum_{j\neq i}\nabla f_j(\mathbf{x}_{(i)}^n)$.

通过 SCA 技术, 将问题 (1) 的平衡点转化为关于映射 $\widehat{\mathbf{x}}_{(i)}(\cdot)$ 的不动点 (见引理 A1).

步骤2　式 (3) 中的 $\widehat{\mathbf{x}}_{(i)}(\mathbf{x}_{(i)}^n)$ 并非完全分布式的, 因为 $\widehat{\pi}_{(i)}(\mathbf{x}_{(i)}^n)$ 的计算需要依赖于所有的 $\nabla f_j(\mathbf{x}_{(i)}^n)$. 然而, 在智能体 i 的本地环境下这些信息无法获得. 为了解决该问题, 本文采用梯度跟踪机制, 用本地估计 $\widetilde{\pi}_{(i)}(\mathbf{x}_{(i)}^n)$ 代替全局梯度 $\widehat{\pi}_{(i)}(\mathbf{x}_{(i)}^n)$. 具体地, 在每次迭代 n 时, 给定当前迭代值 $\mathbf{x}_{(i)}^n$ 和 $\mathbf{y}_{(i)}^n$, 智能体 i 解决问题 (1) 的一个凸近似形式, 即

$$\widetilde{\mathbf{x}}_{(i)}(\mathbf{x}_{(i)}^n;\mathbf{y}_{(i)}^n)\triangleq\underset{\mathbf{x}_{(i)}\in\mathcal{K}}{\operatorname{argmin}}\left\{\widetilde{f}_i(\mathbf{x}_{(i)};\mathbf{x}_{(i)}^n)+\widetilde{\pi}_{(i)}(\mathbf{x}_{(i)}^n)^{\mathrm{T}}(\mathbf{x}_{(i)}-\mathbf{x}_{(i)}^n)+G(\mathbf{x}_{(i)})\right\},\tag{4}$$

其中 $\widetilde{\pi}_{(i)}(\mathbf{x}_{(i)}^n)\triangleq I\cdot\mathbf{y}_{(i)}^n-\nabla f_i(\mathbf{x}_{(i)}^n)$, $\mathbf{y}_{(i)}^n$ 的目标是跟踪梯度的平均值, 即

$$\lim_{n\to\infty}\left\|I\cdot\mathbf{y}_{(i)}^n-\nabla f_i(\mathbf{x}_{(i)}^n)-\sum_{j\neq i}\nabla f_j(\mathbf{x}_{(i)}^n)\right\|=0.\tag{5}$$

为了仅用本地估计使式 (5) 成立, 需要引入一个包含校正项的分布式一致性算法 (6).

具体地, 用 $\delta_{(i)}^n\in\mathbb{R}^m$ 表示在第 n 次迭代中引入到智能体 i 的更新中的校正项. 给定 $\mathbf{x}_{(i)}^n$, 则引入校正项后的分布式一致性算法描述如下:

$$\mathbf{x}_{(i)}^{n+1}=\sum_{j=1}^{I}a_{ij}^n\mathbf{x}_{(j)}^n+\delta_{(i)}^{n+1}.\tag{6}$$

若假设 3 和 4 成立, 则式 (6) 可通过分布式方式实现: 每个智能体 i 仅使用从其当前邻居 (及自身) 接收到的信息 $a_{ij}^n\mathbf{x}_{(j)}^n$ 来更新其变量. 利用引入校正项后的分布式一致性算法 (6) 更新 $\mathbf{y}_{(i)}^n$, 其中校正项 $\delta_{(i)}^{n+1}=\nabla f_i(\mathbf{x}_{(i)}^{n+1})-\nabla f_i(\mathbf{x}_{(i)}^n)$. 由此得到的梯度跟踪机制可表示为

$$\mathbf{y}_{(i)}^{n+1}=\sum_{j=1}^{I}a_{ij}^n\mathbf{y}_{(j)}^n+\nabla f_i(\mathbf{x}_{(i)}^{n+1})-\nabla f_i(\mathbf{x}_{(i)}^n).\tag{7}$$

在此基础上, 受 Nesterov 动量方法 [27] 的启发, 本文提出了一种分布式加速非凸复合优化算法 (见算法 1). 其中, α 是正步长, β 是正动量参数, 用于在每个时间步中修正梯度. $\widetilde{f}_i(\mathbf{x}_{(i)};\mathbf{s}_{(i)}^n)$ 是根据命题 A1 得到的 f_i 的凸近似函数. 在式 (8) 中, 执行与凸近似函数、非光滑凸函数 $G(\mathbf{x}_{(i)})$ 和全局梯度估计 $\widetilde{\pi}_{(i)}^n$ 相关的局部最小化操作. 随后, 根据式 (8) 中获得的局部估值 $\widetilde{\mathbf{x}}_{(i)}^n(\mathbf{s}_{(i)}^n;\mathbf{y}_{(i)}^n)$ 调整变量 $\mathbf{s}_{(i)}^n$, 并令 $\mathbf{x}_{(i)}^{n+1/2}$ 表示修正后的变量. 式 (10) 是一个常见的平均操作, 确保变量 $\mathbf{x}_{(i)}^{n+1/2}$ 在所有 $i\in[I]$ 中达成一致. 然后, 式 (11) 使用动量项对变量 $\mathbf{x}_{(i)}^{n+1}$ 的更新进行进一步调整, 式 (12) 使用梯度跟踪来估计全局函数的梯度值.

算法 1 分布式类 Nesterov 加速非凸复合优化算法.

输入: $n=0$, $\mathbf{s}_{(i)}^0=\mathbf{x}_{(i)}^0\in\mathcal{K}$, $\mathbf{y}_{(i)}^0=\nabla f_i(\mathbf{s}_{(i)}^0)$, $\widetilde{\pi}_{(i)}(\mathbf{s}_{(i)}^0)=I\cdot\mathbf{y}_{(i)}^0-\nabla f_i(\mathbf{s}_{(i)}^0)$, $\forall i\in[I]$.

1: 终止条件判断: 若 $\mathbf{s}^n$ 满足终止条件, 则终止算法, 否则继续迭代;

2: 局部 SCA 优化: 每个智能体 i 局部计算

$$\widetilde{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n;\mathbf{y}_{(i)}^n)=\mathop{\mathrm{argmin}}_{\mathbf{x}_{(i)}\in\mathcal{K}}\left\{\widetilde{f}_i(\mathbf{x}_{(i)};\mathbf{s}_{(i)}^n)+(\widetilde{\pi}_{(i)}(\mathbf{s}_{(i)}^n))^{\mathrm{T}}(\mathbf{x}_{(i)}-\mathbf{s}_{(i)}^n)+G(\mathbf{x}_{(i)})\right\};\tag{8}$$

3: 局部变量更新: 每个智能体 i 更新其局部变量

$$\mathbf{x}_{(i)}^{n+1/2}=\mathbf{s}_{(i)}^n+\alpha(\widetilde{\mathbf{x}}_{(i)}^n-\mathbf{s}_{(i)}^n);\tag{9}$$

4: 一致性更新

$$\mathbf{x}_{(i)}^{n+1}=\sum_{j=1}^I a_{ij}^n\mathbf{x}_{(j)}^{n+1/2};\tag{10}$$

5: 分布式类 Nesterov 加速方法

$$\mathbf{s}_{(i)}^{n+1}=\mathbf{x}_{(i)}^{n+1}+\beta(\mathbf{x}_{(i)}^{n+1}-\mathbf{s}_{(i)}^n);\tag{11}$$

6: 梯度跟踪

$$\mathbf{y}_{(i)}^{n+1}=\sum_{j=1}^I a_{ij}^n\mathbf{y}_{(j)}^n+\nabla f_i(\mathbf{s}_{(i)}^{n+1})-\nabla f_i(\mathbf{s}_{(i)}^n);\tag{12}$$

7: 参数 $\widetilde{\pi}_{(i)}(\mathbf{s}_{(i)}^n)$ 更新

$$\widetilde{\pi}_{(i)}(\mathbf{s}_{(i)}^{n+1})=I\cdot\mathbf{y}_{(i)}^{n+1}-\nabla f_i(\mathbf{s}_{(i)}^{n+1});\tag{13}$$

8: 迭代计数更新: $n\leftarrow n+1$, 返回 1;

输出: 算法终止时的解.

3　性能分析

本节进行算法的收敛性分析及其理论证明. 为便于后续分析, 首先对涉及的重要符号和参数进行统一定义. 为简化表述, 式 (8) 中的 $\widetilde{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n;\mathbf{y}_{(i)}^n)$ 将用 $\widetilde{\mathbf{x}}_{(i)}^n$ 代替.

接下来, 定义:

(a) 全局向量

$$\mathbf{g}_{(i)}^n=\nabla f_i(\mathbf{s}_{(i)}^n),\ \mathbf{g}^n=\left[(\mathbf{g}_{(1)}^n)^{\mathrm{T}},\ldots,(\mathbf{g}_{(I)}^n)^{\mathrm{T}}\right]^{\mathrm{T}},\ \mathbf{s}^n=\left[(\mathbf{s}_{(1)}^n)^{\mathrm{T}},\ldots,(\mathbf{s}_{(I)}^n)^{\mathrm{T}}\right]^{\mathrm{T}},$$

$$\mathbf{y}^n=\left[(\mathbf{y}_{(1)}^n)^{\mathrm{T}},\ldots,(\mathbf{y}_{(I)}^n)^{\mathrm{T}}\right]^{\mathrm{T}},\ \widetilde{\mathbf{x}}^n=\left[(\widetilde{\mathbf{x}}_{(1)}^n)^{\mathrm{T}},\ldots,(\widetilde{\mathbf{x}}_{(I)}^n)^{\mathrm{T}}\right]^{\mathrm{T}};$$

(b) 平均估计

$$\overline{\mathbf{s}}^n=\frac{1}{I}\sum_{i=1}^I\mathbf{s}_{(i)}^n,\ \overline{\mathbf{y}}^n=\frac{1}{I}\sum_{i=1}^I\mathbf{y}_{(i)}^n,\ \overline{\mathbf{g}}^n=\frac{1}{I}\sum_{i=1}^I\mathbf{g}_{(i)}^n;$$

(c) 一致性偏差

$$\mathbf{e}_s^n=\mathbf{s}^n-\mathbf{1}_I\otimes\overline{\mathbf{s}}^n,\ \mathbf{e}_y^n=\mathbf{y}^n-\mathbf{1}_I\otimes\overline{\mathbf{y}}^n,$$

$$\Delta\widetilde{\mathbf{x}}_{(i)}^n=\widetilde{\mathbf{x}}_{(i)}^n-\overline{\mathbf{s}}^n,\ \Delta\widetilde{\mathbf{x}}^n=\widetilde{\mathbf{x}}^n-\mathbf{1}_I\otimes\overline{\mathbf{s}}^n,\Delta\mathbf{s}_{(i)}^n=\widetilde{\mathbf{x}}_{(i)}^n-\mathbf{s}_{(i)}^n,\ \Delta\mathbf{s}^n=\widetilde{\mathbf{x}}^n-\mathbf{s}^n;$$

(d) 基于 L_i, $\widetilde{L}_i$ 和 τ_i 的定义, 引入参数

$$L=\sum_{i=1}^I L_i,\ L_{\max}=\max_{1\leqslant i\leqslant I}L_i,\ \widetilde{L}_{\max}=\max_{1\leqslant i\leqslant I}\widetilde{L}_i,\ c_\tau=\min_{1\leqslant i\leqslant I}\tau_i,\ c_L=\left(L\sqrt{I}+L_{\max}+\widetilde{L}_{\max}\right)/I.$$

设 $\{\mathbf{s}^n \triangleq (\mathbf{s}_{(i)}^n)_{i=1}^I\}_{n\in\mathbb{N}_+}$ 为算法 1 生成的序列, 其收敛性通过以下两个方面衡量: 一是平均序列 $\overline{\mathbf{s}}^n \triangleq (1/I)\cdot\sum_{i=1}^I \mathbf{s}_{(i)}^n$ 与最优解之间的距离; 二是局部变量 $\mathbf{s}_{(i)}^n$ 之间的一致性误差. 距离平衡点的度量由以下函数定义:

$$J(\overline{\mathbf{s}}^n) \triangleq \left\|\overline{\mathbf{s}}^n - \underset{\mathbf{z}\in\mathcal{K}}{\operatorname{argmin}}\left\{\nabla F(\overline{\mathbf{s}}^n)^{\mathrm{T}}(\mathbf{z}-\overline{\mathbf{s}}^n) + \frac{1}{2}\|\mathbf{z}-\overline{\mathbf{s}}^n\|^2 + G(\mathbf{z})\right\}\right\|. \tag{14}$$

J 是一个连续的平稳性度量 [31], 且当且仅当 $\overline{\mathbf{s}}^\infty$ 为问题 (1) 的平衡点集时, $J(\overline{\mathbf{s}}^\infty)=0$. 在第 n 次迭代中, 一致性误差定义为 $D(\mathbf{s}^n) \triangleq \|\mathbf{s}^n - \mathbf{1}_I\otimes\overline{\mathbf{s}}^n\|$, 该误差衡量了当前节点的状态 $\mathbf{s}^n$ 与平均状态 $\overline{\mathbf{s}}^n$ 之间的偏离程度, 反映了系统的一致性收敛情况. 当且仅当所有的 $\mathbf{s}_{(i)}^n$ 达到一致时, D 的值为零. 将度量指标 J 和 D 结合起来, 构造出一个综合的度量函数, 即 $M(\mathbf{s}^n) \triangleq J(\overline{\mathbf{s}}^n)^2 + D(\mathbf{s}^n)^2$. 基于上述度量函数, 本文将给出算法 1 的主要收敛结果.

定理1　若假设 1~4 成立, 且算法 1 的固定步长 α 满足:

$$\alpha \leqslant \min\left\{\frac{(\sigma^2(1-\widetilde{\rho})/B-(B+1/\varepsilon+B/\varepsilon)(\beta+\beta\sqrt{I})^2)^{1/2}}{(2(B+1/\varepsilon+B\varepsilon))^{1/2}(1+\beta\sqrt{I})},\right.$$
$$\left.\frac{c_\tau}{I}\left(\frac{(1+\beta)L}{I}+2\sqrt{\frac{c_\triangle}{1-\sigma^2}}c_L+2\sqrt{c_\perp+\frac{c_\triangle c_\wedge}{1-\sigma^2}}\right)^{-1}\right\}, \tag{15}$$

其中

$$\sigma\in(0,1), \rho_B \triangleq \left(1-\kappa/2I^2\right)^{\frac{1}{2}}, \varepsilon\in\left(0,(1-\rho_B^2)/\left(2B\rho_B^2\right)\right),$$
$$\widetilde{\rho}=\rho_B^2+2B\rho_B^2\varepsilon, \beta\in\left(0,\sigma\sqrt{(1-\widetilde{\rho})}/\left((1+\sqrt{I})\sqrt{B\left(B+1/\varepsilon+B/\varepsilon\right)}\right)\right).$$

则有 $\lim_{n\to\infty} M(\mathbf{s}^n)=0$.

证明　算法 1 (参见式 (11) 和 (12)) 可以被简化为以下紧凑形式:

$$\mathbf{s}^{n+1} = \widehat{\mathbf{A}}^n(\mathbf{s}^n+\alpha\Delta\mathbf{s}^n)+\beta\Delta\theta^n,\ \mathbf{y}^{n+1}=\widehat{\mathbf{A}}^n\mathbf{y}^n+\mathbf{g}^{n+1}-\mathbf{g}^n, \tag{16}$$

其中, $\widehat{\mathbf{A}}^n=\mathbf{A}^n\otimes I_m$, $\Delta\theta^n=\mathbf{x}^{n+1}-\mathbf{s}^n$. 由算法 1 生成的序列 $\overline{\mathbf{s}}^n$ 和 $\overline{\mathbf{y}}^n$ 的动态形式如下所示:

$$\overline{\mathbf{s}}^{n+1}=\overline{\mathbf{s}}^n+\frac{(1+\beta)\,\alpha}{I}\left(\mathbf{1}_I^{\mathrm{T}}\otimes I_m\right)\Delta\widetilde{\mathbf{x}}^n,\ \overline{\mathbf{y}}^{n+1}=\overline{\mathbf{y}}^n+\overline{\mathbf{g}}^{n+1}-\overline{\mathbf{g}}^n. \tag{17}$$

本文研究了 U 在 $\overline{\mathbf{s}}^n$ 轨迹上的动态特性. 为此, 定义优化输入 $\alpha\|\Delta\widetilde{\mathbf{x}}^n\|$, 以及一致性误差 $\|\mathbf{e}_s^n\|$ 和跟踪误差 $\|\mathbf{e}_y^n\|$ 在连续 B 次迭代中的累计平方和如下:

$$E_{\Delta\widetilde{\mathbf{x}}}^n \triangleq \sum_{t=0}^{B-1}\alpha^2\left\|\Delta\widetilde{\mathbf{x}}^{n+t}\right\|^2,\ E_{s_\perp}^n \triangleq \sum_{t=0}^{B-1}\left\|\mathbf{e}_s^{n+t}\right\|^2,\ E_{y_\perp}^n \triangleq \sum_{t=0}^{B-1}\left\|\mathbf{e}_y^{n+t}\right\|^2. \tag{18}$$

为分析序列 $\left\{U\left(\overline{\mathbf{s}}^{n+B+k}\right)\right\}_{n\in\mathbb{N}_+}$ 的收敛性, 首先将 $E_{s\perp}^n$ 和 $E_{y\perp}^n$ (即式 (A7) 中的项 iii, 见引理 A5) 作为 $E_{\Delta\widetilde{\mathbf{x}}}^n$ 的函数加以界定 (见引理 A7). 在引理 A7 的基础上, 进一步证明序列 $\{E_{s\perp}^n\}_{n\in\mathbb{N}_+}$ 和 $\{E_{y\perp}^n\}_{n\in\mathbb{N}_+}$ 的可加性 (见引理 A9). 然后利用引理 A7 和 A9, 构建一个新的李亚普诺夫函数, 其收敛性意味着 $\{E_{\Delta\widetilde{\mathbf{x}}}^n\}_{n\in\mathbb{N}_+}$ 的可加性, 从而进一步说明所有误差序列的收敛性, 正如定理 1 所述.

将式 (A7) 和 (A18) (分别乘以 $(1+\beta)/2$) 相加后, 得到

$$\sum_{k=0}^{B-1}U(\overline{\mathbf{s}}^{n+B+k})+\frac{1+\beta}{2}\xi_y^{-1}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\,\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_y^{n+B+k}\right\|^2$$
$$+\frac{1+\beta}{2\mu_{\min}}\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\,\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+B+k}\right\|^2$$

$$
\begin{aligned}
&\leqslant \sum_{k=0}^{B-1} U(\overline{\mathbf{s}}^{n+k}) + \frac{1+\beta}{2}\xi_y^{-1} \sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_y^{n+k}\right\|^2 \\
&\quad + \frac{1+\beta}{2\mu_{\min}} \left(c_L\xi_x^{-1} + \xi_y^{-1}c_\wedge\right) \sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_s^{n+k}\right\|^2 - \frac{(1+\beta)c_\tau}{2I} \sum_{k=0}^{B-1}\sum_{t=0}^{B-1} \alpha\|\Delta\widetilde{\mathbf{x}}^{n+k+t}\|^2 \\
&\quad + \frac{1+\beta}{2}\left(\frac{(1+\beta)L}{I} + c_L\xi_x + \xi_y + \xi_y^{-1}c_\perp\right) \sum_{k=0}^{B-1} E_{\Delta\widetilde{x}}^{n+k} + \frac{1+\beta}{2\mu_{\min}}\left(c_L\xi_x^{-1} + \xi_y^{-1}c_\wedge\right) c_\triangle \sum_{k=0}^{B-1} E_{\Delta\widetilde{\mathbf{x}}}^{n+k}, \quad (19)
\end{aligned}
$$

定义

$$
\begin{aligned}
V^n \triangleq & \sum_{k=0}^{B-1} U(\overline{\mathbf{s}}^{n+k}) + \frac{1+\beta}{2}\xi_y^{-1} \sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_y^{n+k}\right\|^2 \\
& + \frac{1+\beta}{2\mu_{\min}} \left(c_L\xi_x^{-1} + \xi_y^{-1}c_\wedge\right) \sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_s^{n+k}\right\|^2, \quad (20)
\end{aligned}
$$

和

$$
\gamma = \frac{(1+\beta)c_\tau}{2I} - \frac{(1+\beta)\,\alpha}{2}\left(\frac{(1+\beta)L}{I} + c_L\xi_x + \xi_y + \xi_y^{-1}c_\perp + \frac{c_\triangle}{\mu_{\min}}\left(c_L\xi_x^{-1} + \xi_y^{-1}c_\wedge\right)\right). \quad (21)
$$

将式 (20) 和 (21) 代入式 (19), 即可得到 V^n 所需的下降性质, 对于足够大的 n 有以下成立:

$$
V^{n+B} \leqslant V^n - \sum_{k=0}^{B-1}\sum_{t=0}^{B-1} \alpha\gamma \left\|\Delta\widetilde{\mathbf{x}}^{n+k+t}\right\|^2. \quad (22)
$$

定理 1 的证明分为两个步骤, 具体如下.

在步骤 1 中, 目标是证明 $\lim_{n\to\infty} D(\mathbf{s}^n) = 0$. 为此, 首先利用 Lyapunov 函数的下降性质和引理 A4, 证明 $\lim_{n\to\infty} \|\Delta\widetilde{\mathbf{x}}^n\| = 0$, 然后再结合引理 A8, 进一步证明 $\lim_{n\to\infty} D(\mathbf{s}^n) = 0$, 其中 α 满足式 (15). 在步骤 2 中, 目标是证明 $\lim_{n\to\infty} J(\overline{\mathbf{s}}^n) = 0$. 为此, 将 $J(\overline{\mathbf{s}}^n)$ 拆解为 $\left\|\widehat{\mathbf{x}}_{(i)}(\overline{\mathbf{s}}^n) - \overline{\mathbf{s}}^n\right\|$ 和 $\left\|\overline{\mathbf{s}}(\overline{\mathbf{s}}^n) - \widehat{\mathbf{x}}_{(i)}(\overline{\mathbf{s}}^n)\right\|$ 两部分, 并分别分析其收敛性. 首先利用引理 A1、引理 A3 和步骤 1 的结论, 证明 $\lim_{n\to\infty} \left\|\widehat{\mathbf{x}}_{(i)}(\overline{\mathbf{s}}^n) - \overline{\mathbf{s}}^n\right\| = 0$. 然后利用一阶最优条件, 证明 $\lim_{n\to\infty} \left\|\overline{\mathbf{s}}(\overline{\mathbf{s}}^n) - \widehat{\mathbf{x}}_{(i)}(\overline{\mathbf{s}}^n)\right\| = 0$. 通过以上两部分的分析, 最终可以得出 $\lim_{n\to\infty} J(\overline{\mathbf{s}}^n) = 0$.

步骤1 为了使 V^n 具有所需的下降性质, 需选择 α 使得 $\gamma > 0$. 将 $\alpha_{\max}$ (参见式 (A15)) 和 $\mu_{\min} = 1-\sigma^2$ 的表达式代入 (21), 可以验证, 当 $\alpha \leqslant \alpha_{\max}$ 且 α 还满足以下条件时, 有 $\gamma > 0$.

$$
\alpha \leqslant \frac{c_\tau}{I}\left(\frac{(1+\beta)L}{I} + c_L\xi_x + \frac{c_\triangle c_L}{1-\sigma^2}\xi_x^{-1} + \xi_y + \left(c_\perp + \frac{c_\triangle c_\wedge}{1-\sigma^2}\right)\xi_y^{-1}\right)^{-1}, \quad (23)
$$

其中 $\xi_x, \xi_y > 0$ 为自由参数. 上述上界的最大值由以下公式确定:

$$
\xi_x = \sqrt{\frac{c_\triangle}{1-\sigma^2}},\ \ \xi_y = \sqrt{c_\perp + \frac{c_\triangle c_\wedge}{1-\sigma^2}}, \quad (24)
$$

将式 (24) 代入式 (23) 可得

$$
\alpha \leqslant \frac{c_\tau}{I}\left(\frac{(1+\beta)L}{I} + 2\sqrt{\frac{c_\triangle}{1-\sigma^2}}c_L + 2\sqrt{c_\perp + \frac{c_\triangle c_\wedge}{1-\sigma^2}}\right)^{-1}. \quad (25)
$$

结合 $\alpha \leqslant \alpha_{\max}$ 和式 (25), 可以得到 α 的最终界限 (15). 在满足 (15) 的条件下, 利用式 (22) 和引理 A4, 可以得出 $\lim_{n\to\infty} \|\Delta\widetilde{\mathbf{x}}^n\| = 0$. 根据引理 A8, 进一步得到 $\lim_{n\to\infty} \|\mathbf{e}_s^n\| = 0$ 和 $\lim_{n\to\infty} \|\mathbf{e}_y^n\| = 0$, 即 $\lim_{n\to\infty} D(\mathbf{s}^n) = 0$.

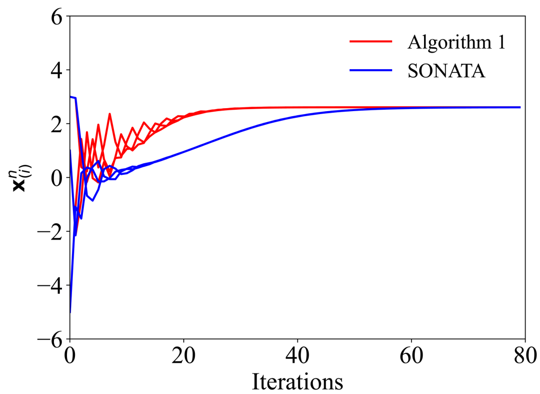

图 1 (网络版彩图) $\mathbf{x}_{(1)}^n$, $\mathbf{x}_{(2)}^n$ 和 $\mathbf{x}_{(3)}^n$ 的值.

Figure 1 (Color online) Values of $\mathbf{x}_{(1)}^n$, $\mathbf{x}_{(2)}^n$ and $\mathbf{x}_{(3)}^n$.

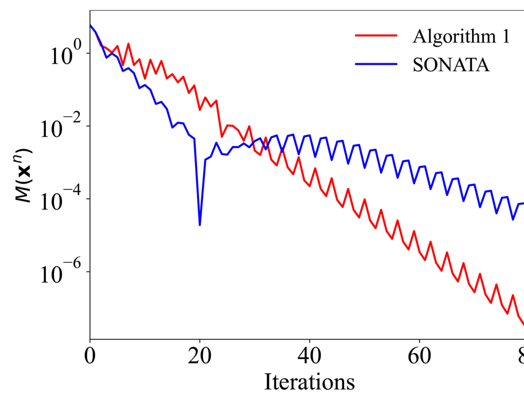

图 2 (网络版彩图) $M(\mathbf{x}^n)$ 随迭代次数的变化.

Figure 2 (Color online) Variation of $M(\mathbf{x}^n)$ with the number of iterations.

步骤2 在步骤 1 中, 本文证明了以下 3 点: (i) $\lim_{n\to\infty}\|\Delta\widetilde{\mathbf{x}}^n\| = 0$; (ii) $\lim_{n\to\infty}\|\mathbf{e}_s^n\| = 0$; (iii) $\lim_{n\to\infty}\|\mathbf{e}_y^n\| = 0$. 接下来证明 $\lim_{n\to\infty} J(\bar{\mathbf{s}}^n) = 0$. 回忆 $J(\bar{\mathbf{s}}^n) \triangleq \|\bar{\mathbf{s}}(\bar{\mathbf{s}}^n) - \bar{\mathbf{s}}^n\|$ 的定义, 为了简化记号, 设 $\bar{\mathbf{s}}(\bar{\mathbf{s}}^n) \triangleq \operatorname{argmin}_{\mathbf{z}\in\mathcal{K}}\{\nabla F(\bar{\mathbf{s}}^n)^{\mathrm{T}}(\mathbf{z}-\bar{\mathbf{s}}^n) + \frac{1}{2}\|\mathbf{z}-\bar{\mathbf{s}}^n\|^2 + G(\mathbf{z})\}$. 由于 $J(\bar{\mathbf{s}}^n) \leqslant \|\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n) - \bar{\mathbf{s}}^n\| + \|\bar{\mathbf{s}}(\bar{\mathbf{s}}^n) - \widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n)\|$, 因此, 只需证明右侧的两个项在渐近意义下均趋于零即可. 下文对此进行证明.

(1) $\lim_{n\to\infty}\|\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n) - \bar{\mathbf{s}}^n\| = 0$. 为证明 $\|\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n) - \bar{\mathbf{s}}^n\|$ 在渐近意义下趋于零, 对其进行如下界定 $\|\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n) - \bar{\mathbf{s}}^n\| \leqslant \|\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n) - \widehat{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n)\| + \|\widehat{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n) - \widetilde{\mathbf{x}}_{(i)}^n\| + \|\widetilde{\mathbf{x}}_{(i)}^n - \bar{\mathbf{s}}^n\|$. 结合引理 A1、引理 A3 以及结论 (i)~(iii), 可以得出 $\lim_{n\to\infty}\|\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n) - \bar{\mathbf{s}}^n\| = 0$.

(2) $\lim_{n\to\infty}\|\bar{\mathbf{s}}(\bar{\mathbf{s}}^n) - \widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n)\| = 0$. 利用 $\bar{\mathbf{s}}(\bar{\mathbf{s}}^n)$ 和 $\widehat{\mathbf{x}}_{(i)}(\bar{\mathbf{s}}^n)$ 的一阶最优条件, 可以对它们的差异进行如下界定:

$$
\begin{aligned}
&\left\|\bar{\mathbf{s}}\left(\bar{\mathbf{s}}^n\right) - \widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right)\right\| \leqslant \left\|\nabla\widetilde{f}_i\left(\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right);\bar{\mathbf{s}}^n\right) - \nabla f_i\left(\bar{\mathbf{s}}^n\right) - \widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right) + \bar{\mathbf{s}}^n\right\| \\
&\leqslant \left\|\nabla\widetilde{f}_i\left(\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right);\bar{\mathbf{s}}^n\right) - \nabla\widetilde{f}_i\left(\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right);\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right)\right)\right\| + \left\|\nabla f_i\left(\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right)\right) - \nabla f_i\left(\bar{\mathbf{s}}^n\right)\right\| + \left\|\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right) - \bar{\mathbf{s}}^n\right\| \\
&\leqslant \left(\widetilde{L}_i + L_i + 1\right)\left\|\widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right) - \bar{\mathbf{s}}^n\right\|. \qquad (26)
\end{aligned}
$$

利用结论 (1), 可以得到 $\lim_{n\to\infty}\left\|\bar{\mathbf{s}}\left(\bar{\mathbf{s}}^n\right) - \widehat{\mathbf{x}}_{(i)}\left(\bar{\mathbf{s}}^n\right)\right\| = 0$, 进而得到 $\lim_{n\to\infty} J\left(\bar{\mathbf{s}}^n\right) = 0$.

4 数值仿真

为便于说明, 本文考虑一个简单的分布式非凸优化问题, 即 $U(\mathbf{x}) = \sum_{i=1}^3 f_i(\mathbf{x}) + \|\mathbf{x}\|_1$. 该问题发生在包含 $I=3$ 个智能体的拓扑上. 其中局部目标函数 f_i 具体描述如下:

$$
f_1(\mathbf{x}) = \begin{cases} \left(\mathbf{x}^3 - 16\mathbf{x}\right)(\mathbf{x}+2), & \text{if } |\mathbf{x}| \leqslant 10, \\ 4248\mathbf{x} - 32400, & \text{if } \mathbf{x} > 10, \\ -3112\mathbf{x} - 25040, & \text{if } \mathbf{x} < -10, \end{cases} \quad
f_2(\mathbf{x}) = \begin{cases} \left(0.5\mathbf{x}^3 + \mathbf{x}^2\right)(\mathbf{x}-4), & \text{if } |\mathbf{x}| \leqslant 10, \\ 1620\mathbf{x} - 12600, & \text{if } \mathbf{x} > 10, \\ -2220\mathbf{x} - 16600, & \text{if } \mathbf{x} < -10, \end{cases}
$$

$$
f_3(\mathbf{x}) = \begin{cases} (\mathbf{x}+2)^2(\mathbf{x}-4), & \text{if } |\mathbf{x}| \leqslant 10, \\ 228\mathbf{x} - 2016, & \text{if } \mathbf{x} > 10, \\ 228\mathbf{x} - 1984, & \text{if } \mathbf{x} < -10. \end{cases}
$$

通信拓扑 $\mathcal{G}^n$ 的选择方式如下: 每两个连续时间段的连通图被选择为以下两种随机设定的信息交换图组合: $([3], 1 \leftrightarrow 2, 2 \leftrightarrow 3)$, $([3], 1 \leftrightarrow 2, 1 \leftrightarrow 3)$ 或者 $([3], 1 \leftrightarrow 2, 1 \leftrightarrow 3)$, $([3], 1 \leftrightarrow 2, 2 \leftrightarrow 3)$. 显然, 这样的通信拓扑序列是 4 强连通的. 此外, 假设 1∼4 对于函数 U 成立, 本文设置 $\alpha = 0.003$ 和 $\beta = 0.2$, 初始代理估计值 $\mathbf{x}^0_{(1)} = -5$, $\mathbf{x}^0_{(2)} = 1$ 和 $\mathbf{x}^0_{(3)} = 3$, 然后应用算法 1 来保证渐近收敛到平衡点, 仿真效果如图 1 和 2 所示. 从图 1 可以看出算法 1 收敛到平衡点 $\mathbf{x} = 2.45$. 图 2 显示了算法 1 相比于未进行加速的 SONATA 算法 [29] 有着更快的收敛速度.

5　总结

本文设计了一种分布式类 Nesterov 加速非凸复合优化算法, 适用于通信拓扑时变的多智能体网络. 该算法结合 SCA 技术与梯度跟踪机制, 并引入类 Nesterov 动量项, 以提升算法的收敛效率. 理论证明了当固定步长和动量参数小于给定的上界时, 所提算法渐近收敛至所考虑问题的平衡点集. 未来研究将聚焦于如何进一步提高算法的收敛速度, 以拓展其在实际工程中的应用价值.

参考文献

1 Lasi H, Fettke P, Kemper H G, et al. Industry 4.0. Bus Inf Syst Eng, 2014, 6: 239–242

2 Wang X W. Industry 4.0: intelligent industry. IoT Technol, 2013, 3: 3–4 [王喜文. 工业 4.0: 智能工业. 物联网技术, 2013, 3: 3–4]

3 Meziane F, Vadera S, Kobbacy K, et al. Intelligent systems in manufacturing: current developments and future prospects. Integr Manufact Syst, 2000, 11: 218–238

4 Ding J L, Yang C E, Chen Y D, et al. Current status and prospects of intelligent optimization decision systems for complex industrial processes. Acta Automat Sin, 2018, 44: 1931–1943 [丁进良, 杨翠娥, 陈远东, 等. 复杂工业过程智能优化决策系统的现状与展望. 自动化学报, 2018, 44: 1931–1943]

5 Yin S R, Ai Q, Zeng S Q, et al. Challenges and prospects of distributed multi-energy optimization research in energy Internet. Power Grid Technol, 2018, 42: 1359–1369 [殷爽睿, 艾芊, 曾顺奇, 等. 能源互联网多能分布式优化研究挑战与展望. 电网技术, 2018, 42: 1359–1369]

6 Zhong W, Yu R, Xie S, et al. Software defined networking for flexible and green energy Internet. IEEE Commun Mag, 2016, 54: 68–75

7 Wang K, Yu J, Yu Y, et al. A survey on energy Internet: architecture, approach, and emerging technologies. IEEE Syst J, 2017, 12: 2403–2416

8 Yang T, Yi X, Wu J, et al. A survey of distributed optimization. Annu Rev Control, 2019, 47: 278–305

9 Xu Y, Zhang W, Liu W, et al. Distributed subgradient-based coordination of multiple renewable generators in a microgrid. IEEE Trans Power Syst, 2013, 29: 23–33

10 Wang Z, Wu W, Zhang B. A fully distributed power dispatch method for fast frequency recovery and minimal generation cost in autonomous microgrids. IEEE Trans Smart Grid, 2015, 7: 19–31

11 Wang J, Elia N. Control approach to distributed optimization. In: Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing, 2010. 557–561

12 Lu J, Tang C Y. Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case. IEEE Trans Autom Control, 2012, 57: 2348–2354

13 Liu C, Shi Y, Li H, et al. Accelerated dual averaging methods for decentralized constrained optimization. IEEE Trans Autom Control, 2022, 68: 2125–2139

14 Shi W, Ling Q, Wu G, et al. EXTRA: an exact first-order algorithm for decentralized consensus optimization. SIAM J Optim, 2015, 25: 944–966

15 Jakovetic D, Xavier J, Moura J M F. Fast distributed gradient methods. IEEE Trans Autom Control, 2014, 59: 1131–1146

16 Nesterov Y. Introductory Lectures on Convex Optimization: A Basic Course. Berlin: Springer, 2013. 87

17 Nguyen D T A, Nguyen D T, Nedić A. Geometric convergence of distributed heavy-ball Nash equilibrium algorithm over time-varying digraphs with unconstrained actions. IEEE Control Syst Lett, 2023, 7: 1963–1968

18 Qu G, Li N. Accelerated distributed Nesterov gradient descent. IEEE Trans Autom Control, 2019, 65: 2566–2581

19 Xin R, Jakovetic D, Khan U A. Distributed Nesterov gradient methods over arbitrary graphs. IEEE Signal Process Lett, 2019, 26: 1247–1251
20 Nguyen D T A, Nguyen D T, Nedich A. Accelerated AB/push-pull methods for distributed optimization over time-varying directed networks. 2023. ArXiv:2302.01214
21 Nedich A, Nguyen D T A, Nguyen D T. AB/push-pull method for distributed optimization in time-varying directed networks. 2022. ArXiv:2209.06974
22 Wen G, Yu X, Liu Z. Recent progress on the study of distributed economic dispatch in smart grid: an overview. Front Inform Technol Electron Eng, 2021, 22: 25–39
23 Wu F, Huang S, Li R, et al. Research on the collaborative optimization of multi-energy flow microgrids. Energy Procedia, 2016, 103: 345–350
24 Ho W S, Macchietto S, Lim J S, et al. Optimal scheduling of energy storage for renewable energy distributed energy generation system. Renew Sustain Energy Rev, 2016, 58: 1100–1107
25 Chen A. Fast distributed first-order methods. Dissertation for Ph.D. Degree. Cambridge: Massachusetts Institute of Technology, 2012
26 Ravazzi C, Fosson S, Magli E. Distributed soft thresholding for sparse signal recovery. In: Proceedings of IEEE Global Communications Conference (GLOBECOM), 2013. 3429–3434
27 Zheng L, Li H, Li J, et al. A distributed Nesterov-like gradient tracking algorithm for composite constrained optimization. IEEE Trans Signal Inf Process over Netws, 2023, 9: 60–73
28 Lorenzo P D, Scutari G. NEXT: in-network nonconvex optimization. IEEE Trans Signal Inf Process over Netws, 2016, 2: 120–136
29 Scutari G, Sun Y. Distributed nonconvex constrained optimization over time-varying digraphs. Math Program, 2019, 176: 497–544
30 Wai H T, Lafond J, Scaglione A, et al. Decentralized Frank-Wolfe algorithm for convex and nonconvex problems. IEEE Trans Autom Control, 2017, 62: 5522–5537
31 Facchinei F, Scutari G, Sagratella S. Parallel selective algorithms for nonconvex big data optimization. IEEE Trans Signal Process, 2015, 63: 1874–1889

附录 A

代理函数 $\widetilde{f}_i$ 的选取需满足以下条件.

命题A1　设 $\widetilde{f}_i:\mathcal{K}\times\mathcal{K}\to\mathbb{R}$ 是关于第 1 个变量的 $\mathbf{C}^1$ 函数, 并满足以下要求:

(1) $\nabla\widetilde{f}_i(\mathbf{x};\mathbf{x})=\nabla f_i(\mathbf{x}), \forall\mathbf{x}\in\mathcal{K}$;

(2) $\widetilde{f}_i(\cdot;\mathbf{y})$ 在 $\mathcal{K}$ 上关于 τ_i 是一致强凸的;

(3) $\nabla\widetilde{f}_i(\mathbf{x};\cdot)$ 在 $\mathcal{K}$ 上是关于常数 $\widetilde{L}_i$ 一致 Lipschitz 连续的;

其中, $\nabla\widetilde{f}_i(\mathbf{x};\mathbf{y})$ 表示函数 $\widetilde{f}_i$ 关于第 1 个变量在点 $(\mathbf{x},\mathbf{y})$ 处的偏导数.

映射 $\widehat{\mathbf{x}}_{(i)}(\cdot)$ 具有以下性质.

引理A1 (文献 [31] 中的命题 8)　每个映射 $\widehat{\mathbf{x}}_{(i)}(\cdot)$ (参见式 (3)) 满足以下性质:

(1) $\widehat{\mathbf{x}}_{(i)}(\cdot)$ 在 $\mathcal{K}$ 上是 $\widehat{L}$-Lipschitz 连续的, 即存在一个有限的常数 $\widehat{L}>0$, 使得

$$\left\|\widehat{\mathbf{x}}_{(i)}(\mathbf{z})-\widehat{\mathbf{x}}_{(i)}(\mathbf{w})\right\|\leqslant\widehat{L}\left\|\mathbf{z}-\mathbf{w}\right\|,\ \forall\mathbf{z},\mathbf{w}\in\mathcal{K}; \tag{A1}$$

(2) 映射 $\widehat{\mathbf{x}}_{(i)}(\cdot)$ 的不动点集合与问题 (1) 的平衡点集合一致.

为了便于引出引理 A2, 先将包含校正项的分布式一致性算法 (6) 以矩阵 – 向量形式重写为

$$\mathbf{x}^{n+1}=\widehat{\mathbf{A}}^n\mathbf{x}^n+\delta^{n+1}, \tag{A2}$$

其中

$$\mathbf{x}^n\triangleq\left[(\mathbf{x}^n_{(1)})^{\mathrm{T}},\ldots,(\mathbf{x}^n_{(I)})^{\mathrm{T}}\right]^{\mathrm{T}},\ \delta^n\triangleq\left[(\delta^n_{(1)})^{\mathrm{T}},\ldots,(\delta^n_{(I)})^{\mathrm{T}}\right]^{\mathrm{T}},\ \widehat{\mathbf{A}}^n=\mathbf{A}^n\otimes I_m. \tag{A3}$$

引理 A2 为后续引理 A6 的证明提供了基础, 现表述如下.

引理A2 (文献 [29] 中的命题 1)　设 $\{\mathcal{G}^n\}_{n\in N_+}$ 为满足假设 2 的一系列时变无向图, $\{\mathbf{x}^n\}_{n\in N_+}$ 是通过包含校正项的分布式一致性算法 (A2) 生成的序列, 其中权重矩阵 $\{\mathbf{A}^n\}_{n\in N_+}$ 满足假设 3 和 4. 那么, 一致性偏差的动态衰减满足

$$\|\mathbf{e}^n_x\|\leqslant\lambda^k\left\|\mathbf{e}^{n-k}_x\right\|+\lambda^t\sum_{t=0}^{k-1}\left\|\delta^{n-t}\right\|,\ \forall n,k\in\mathbb{N}_+,\ n\geqslant k, \tag{A4}$$

其中 $\mathbf{e}^n_x\triangleq\mathbf{x}^n-\mathbf{1}_I\otimes\sum_{i=1}^I\mathbf{x}^n_{(i)}$, $\lambda^t\triangleq\min\{1,(\rho)^{\lfloor t/B\rfloor}\}$, $\rho\triangleq\left(1-\kappa/2I^2\right)^{1/2}$. 引理 A2 提供了一组关于包含校正项的分布式一致性方案的统一收敛条件, 这些条件适用于任意给定的校正项序列 $\{\delta^n\}_{n\in\mathbb{N}_+}$.

下一个引理表明如果一致性误差 $\mathbf{e}_s^n$ 和跟踪误差 $\mathbf{e}_y^n$ 都趋于零, 则代理 i 的解 $\widetilde{\mathbf{x}}_{(i)}^n$ 与其无误差对应解 $\widehat{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n)$ 之间的分歧也将渐近消失.

引理A3 (文献 [28] 中的命题 9)　$\widetilde{\mathbf{x}}_{(i)}^n$ 和 $\widehat{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n)$ 满足以下关系式:

$$\left\|\widetilde{\mathbf{x}}_{(i)}^n - \widehat{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n)\right\| \leqslant \frac{I}{\tau_i}\|\mathbf{e}_y^n\| + \frac{L_{\max}\sqrt{I}+L}{\tau_i}\|\mathbf{e}_s^n\|. \tag{A5}$$

因此, 有 $\|\mathbf{e}_s^n\|, \|\mathbf{e}_y^n\| \underset{n\to\infty}{\to} 0 \Rightarrow \|\widetilde{\mathbf{x}}_{(i)}^n - \widehat{\mathbf{x}}_{(i)}(\mathbf{s}_{(i)}^n)\| \underset{n\to\infty}{\to} 0$.

引理A4 (脚注1) 中的引理 1)　设 $\{\mathbf{X}^n\}_{n\in\mathbb{N}_+}$, $\{\mathbf{Y}^n\}_{n\in\mathbb{N}_+}$ 和 $\{\mathbf{Z}^n\}_{n\in\mathbb{N}_+}$ 是 3 个序列, 其中 $\mathbf{X}^n$ 和 $\mathbf{Z}^n$ 对于所有 $n \in \mathbb{N}_+$ 都是非负的. 假设

$$\sum_{k=0}^{B-1}\mathbf{Y}^{n+B+k} \leqslant \sum_{k=0}^{B-1}\mathbf{Y}^{n+k} - \sum_{k=0}^{B-1}\mathbf{X}^{n+k} + \sum_{k=0}^{B-1}\mathbf{Z}^{n+k}, \tag{A6}$$

其中 $n = 0,1,\ldots,\sum_{n=0}^{\infty}\mathbf{Z}^n < +\infty$. 那么, 要么 $\sum_{k=0}^{B-1}\mathbf{Y}^{n+k} \to -\infty$, 要么 $\sum_{k=0}^{B-1}\mathbf{Y}^{n+k}$ 收敛到一个有限值, 并且 $\sum_{n=0}^{\infty}\mathbf{X}^n < +\infty$.

回顾 c_τ 的定义, 有以下结论.

引理A5 (文献 [29] 中的引理 10)　令 $\{(\mathbf{s}^n, \mathbf{y}^n)\}_{n\in\mathbb{N}_+}$ 为算法 1 生成的序列, 则满足以下条件:

$$\begin{aligned}\sum_{k=0}^{B-1}U(\bar{\mathbf{s}}^{n+B+k}) \leqslant{}& \sum_{k=0}^{B-1}U(\bar{\mathbf{s}}^{n+k}) - \frac{(1+\beta)c_\tau}{2I}\sum_{k=0}^{B-1}\sum_{t=0}^{B-1}\alpha\|\Delta\widetilde{\mathbf{x}}^{n+k+t}\|^2 \\ &+ \frac{1+\beta}{2}\underbrace{\sum_{k=0}^{B-1}\left(\frac{c_L}{\xi_x}E_{s_\perp}^{n+k} + \frac{1}{\xi_y}E_{y_\perp}^{n+k}\right)}_{\text{term iii}} + \frac{1+\beta}{2}\left(\frac{(1+\beta)L}{I} + c_L\xi_x + \xi_y\right)\sum_{k=0}^{B-1}E_{\Delta\widetilde{x}}^{n+k},\end{aligned} \tag{A7}$$

其中, $\xi_x > 0$ 和 $\xi_y > 0$.

下个引理刻画了一致性误差 $\|\mathbf{e}_s^n\|$ 和跟踪误差 $\|\mathbf{e}_y^n\|$ 的动态演化过程.

引理A6　$\|\mathbf{e}_s^n\|$ 和 $\|\mathbf{e}_y^n\|$ 满足以下递推不等式:

$$\left\|\mathbf{e}_s^{n+B}\right\| \leqslant \rho_B\|\mathbf{e}_s^n\| + \sum_{t=0}^{B-1}\beta(\sqrt{I}+1)\left\|\mathbf{e}_s^{n+t}\right\| + \sum_{t=0}^{B-1}(1+\beta\sqrt{I})\left\|\alpha\Delta\mathbf{s}^{n+t}\right\|, \tag{A8}$$

$$\left\|\mathbf{e}_y^{n+B}\right\| \leqslant \rho_B\|\mathbf{e}_y^n\| + L_{\max}\sum_{t=0}^{B-1}\left(2+\beta(\sqrt{I}+1)\right)\left\|\mathbf{e}_s^{n+t}\right\| + L_{\max}\sum_{t=0}^{B-1}(1+\beta\sqrt{I})\left\|\alpha\Delta\mathbf{s}^{n+t}\right\|. \tag{A9}$$

证明　回忆算法 1 的更新公式以向量 – 矩阵形式表示在式 (16) 中. 注意, 其中 $\mathbf{s}$ 的更新是分布式一致性算法 (A2) 的一个特例, 校正项为 $\delta^{n+1} = \alpha\widehat{\mathbf{A}}^n\Delta\mathbf{s}^n + \beta\Delta\theta^n$. 然后应用引理 A2, 可以得到

$$\begin{aligned}\left\|\mathbf{e}_s^{n+B}\right\| &\leqslant \rho_B\|\mathbf{e}_s^n\| + \sum_{t=0}^{B-1}\lambda^{B-1-t}\left\|\alpha\widehat{\mathbf{A}}^{n+t}\Delta\mathbf{s}^{n+t} + \beta\Delta\theta^{n+t}\right\| \\ &\leqslant \rho_B\|\mathbf{e}_s^n\| + \sum_{t=0}^{B-1}\left(\left\|\alpha\Delta\mathbf{s}^{n+t}\right\| + \left\|\beta\Delta\theta^{n+t}\right\|\right) \\ &= \rho_B\|\mathbf{e}_s^n\| + \sum_{t=0}^{B-1}\left\|\alpha\Delta\mathbf{s}^{n+t}\right\| + \sum_{t=0}^{B-1}\beta\left\|\widehat{\mathbf{A}}^{n+t}\left(\mathbf{s}^{n+t}+\alpha\Delta\mathbf{s}^{n+t}\right) - \mathbf{s}^{n+t}\right\| \\ &\leqslant \rho_B\|\mathbf{e}_s^n\| + \sum_{t=0}^{B-1}\left\|\alpha\Delta\mathbf{s}^{n+t}\right\| + \sum_{t=0}^{B-1}\beta\left\|\widehat{\mathbf{A}}^{n+t}\left(\mathbf{s}^{n+t} - 1_I\otimes\bar{\mathbf{s}}^{n+t}\right)\right\| \\ &\quad + \sum_{t=0}^{B-1}\beta\left(\left\|\alpha\widehat{\mathbf{A}}^{n+t}\Delta\mathbf{s}^{n+t}\right\| + \left\|1_I\otimes\bar{\mathbf{s}}^{n+t} - \mathbf{s}^{n+t}\right\|\right) \\ &\leqslant \rho_B\|\mathbf{e}_s^n\| + \sum_{t=0}^{B-1}\left\|\alpha\Delta\mathbf{s}^{n+t}\right\| + \sum_{t=0}^{B-1}\beta\left(\sum_{i=1}^{I}\left\|\mathbf{s}_{(i)}^{n+t} - \bar{\mathbf{s}}^{n+t}\right\| + \sum_{i=1}^{I}\alpha\left\|\Delta\mathbf{s}_{(i)}^{n+t}\right\| + \left\|\mathbf{e}_s^{n+t}\right\|\right) \\ &\leqslant \rho_B\|\mathbf{e}_s^n\| + (1+\beta\sqrt{I})\sum_{t=0}^{B-1}\left\|\alpha\Delta\mathbf{s}^{n+t}\right\| + \beta(1+\sqrt{I})\sum_{t=0}^{B-1}\left\|\mathbf{e}_s^{n+t}\right\|.\end{aligned} \tag{A10}$$

1) Bertsekas D P, Tsitsiklis J N. Gradient convergence in gradient methods with errors. SIAM J Optim, 2000, 10: 627–642.

为了证明式 (A9), 采用类似的方法: 注意到式 (16) 中 $\mathbf{y}$ 的更新是 $\mathbf{x}^{n+1} = \widehat{\mathbf{A}}^n \mathbf{x}^n + \delta^{n+1}$ 的一个特例, 其校正项为 $\delta^{n+1} = \mathbf{g}^{n+1} - \mathbf{g}^n$, 因此可以写为

$$\begin{aligned}
\left\|\mathbf{e}_y^{n+B}\right\| &\leqslant \rho_B \|\mathbf{e}_y^n\| + \sum_{t=0}^{B-1} \lambda^{B-1-t} \left\|\mathbf{g}^{n+t+1} - \mathbf{g}^{n+t}\right\| \\
&\leqslant \rho_B \|\mathbf{e}_y^n\| + \sum_{t=0}^{B-1} L_{\max} \left\|\mathbf{s}^{n+t+1} - \mathbf{s}^{n+t}\right\| \\
&\leqslant \rho_B \|\mathbf{e}_y^n\| + \sum_{t=0}^{B-1} L_{\max} \left\|\beta \Delta\theta^{n+t} + J\mathbf{s}^{n+t} - \mathbf{s}^{n+t}\right\| + \sum_{t=0}^{B-1} L_{\max} \left\|\widehat{\mathbf{A}}^{n+t} \left(\mathbf{s}^{n+t} - J\mathbf{s}^{n+t} + \alpha\Delta\mathbf{s}^{n+t}\right)\right\| \\
&\leqslant \rho_B \|\mathbf{e}_y^n\| + \sum_{t=0}^{B-1} L_{\max}(1+\beta\sqrt{I}) \left\|\alpha\Delta\mathbf{s}^{n+t}\right\| + \sum_{t=0}^{B-1} L_{\max} \left(2 + \beta(1+\sqrt{I})\right) \left\|\mathbf{e}_s^{n+t}\right\|.
\end{aligned} \tag{A11}$$

至此, 证明完成.

利用引理 A6, 可以得到一致性误差 $\|\mathbf{e}_s^n\|$ 和跟踪误差 $\|\mathbf{e}_y^n\|$ 在连续 B 次迭代中的累计平方和的动态特性.

引理A7 (文献 [29] 中的命题 12)　序列 $\{\|\mathbf{e}_s^n\|^2\}_{n\in\mathbb{N}_+}$ 和 $\{\|\mathbf{e}_y^n\|^2\}_{n\in\mathbb{N}_+}$ 满足以下条件:

$$\begin{aligned}
&\sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_s^{n+B+k}\right\|^2 \\
&\leqslant \sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_s^{n+k}\right\|^2 - \underbrace{\left(1 - \frac{B}{1-\widetilde{\rho}}\widetilde{m}_1\right)}_{\mu} \sum_{k=0}^{B-1} E_{s\perp}^{n+k} + \underbrace{\frac{B}{1-\widetilde{\rho}}\widetilde{m}_2}_{c\Delta} \sum_{k=0}^{B-1} E_{\Delta\widetilde{\mathbf{x}}}^{n+k},
\end{aligned} \tag{A12}$$

$$\begin{aligned}
&\sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_y^{n+B+k}\right\|^2 \\
&\leqslant \sum_{k=0}^{B-1} \frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}} \left\|\mathbf{e}_y^{n+k}\right\|^2 - \sum_{k=0}^{B-1} E_{y\perp}^{n+k} + \underbrace{\frac{B}{1-\widetilde{\rho}}\widetilde{m}_3}_{c\square} \sum_{k=0}^{B-1} E_{s\perp}^{n+k} + \underbrace{\frac{B}{1-\widetilde{\rho}}\widetilde{m}_4}_{c\perp} \sum_{k=0}^{B-1} E_{\Delta\widetilde{\mathbf{x}}}^{n+k},
\end{aligned} \tag{A13}$$

其中

$$\begin{aligned}
&\widetilde{\rho} = \rho_B^2 + 2B\rho_B^2\varepsilon, \quad \varepsilon \in \left(0, \left(1-\rho_B^2\right) / \left(2B\rho_B^2\right)\right), \\
&\widetilde{m}_1 = \left(B + \frac{1}{\varepsilon} + B\frac{1}{\varepsilon}\right)\left(\beta + \beta\sqrt{I}\right)^2 + 2\left(B + \frac{1}{\varepsilon} + B\varepsilon\right)(1+\beta\sqrt{I})^2\alpha^2, \\
&\widetilde{m}_2 = 2\left(B + \frac{1}{\varepsilon} + B\varepsilon\right)(1+\beta\sqrt{I})^2, \\
&\widetilde{m}_3 = L_{\max}^2\left(B + \frac{1}{\varepsilon} + B\frac{1}{\varepsilon}\right)\left(2 + \beta + \beta\sqrt{I}\right)^2 + 2L_{\max}^2\left(B + \frac{1}{\varepsilon} + B\varepsilon\right)(1+\beta\sqrt{I})^2\alpha^2, \\
&\widetilde{m}_4 = 2L_{\max}^2\left(B + \frac{1}{\varepsilon} + B\frac{1}{\varepsilon}\right)(1+\beta\sqrt{I})^2.
\end{aligned} \tag{A14}$$

设 $\sigma \in (0,1)$, 令 $\mu \geqslant \mu_{\min} = 1 - \sigma^2 > 0$, 则可以得到步长 α 的其中一个上限, 即

$$\alpha \leqslant \alpha_{\max} = \frac{(\sigma^2(1-\widetilde{\rho})/B - (B + 1/\varepsilon + B/\varepsilon)(\beta + \beta\sqrt{I})^2)^{1/2}}{(2(B + 1/\varepsilon + B\varepsilon))^{1/2}(1+\beta\sqrt{I})}, \tag{A15}$$

其中

$$\beta \in \left(0, \sigma\sqrt{(1-\widetilde{\rho})} / \left((1+\sqrt{I})\sqrt{B(B + 1/\varepsilon + B/\varepsilon)}\right)\right). \tag{A16}$$

引理A8　假设 (i) $\sum_{n=0}^{\infty} \alpha^2 \|\Delta\widetilde{\mathbf{x}}^n\|^2 < \infty$; (ii) $\alpha \leqslant \alpha_{\max}$. 那么, 一致性误差和跟踪误差分别满足 $\sum_{n=0}^{\infty} \|\mathbf{e}_s^n\|^2 < \infty$ 和 $\sum_{n=0}^{\infty} \|\mathbf{e}_y^n\|^2 < \infty$.

证明　由式 (18) 可知, 为了证明 $\sum_{n=0}^{\infty} E_{\mathbf{s}_\perp}^n < \infty$ (对应于 $\sum_{n=0}^{\infty} \|\mathbf{e}_s^n\|^2 < \infty$) 以及 $\sum_{n=0}^{\infty} E_{\mathbf{y}_\perp}^n < \infty$ (对应于 $\sum_{n=0}^{\infty} \|\mathbf{e}_y^n\|^2 < \infty$), 仅需证明前者即可.

将引理 A4 应用于式 (A12) (见引理 A7), 可以得到 $\sum_{n=0}^{\infty} E_{\Delta\widetilde{\mathbf{x}}}^n < +\infty \Longrightarrow \sum_{n=0}^{\infty} E_{\mathbf{s}_\perp}^n < +\infty$. 这可以通过以下不等式推出:

$$\sum_{k=0}^{n} E_{\Delta\widetilde{\mathbf{x}}}^k = \sum_{k=0}^{n} \sum_{t=0}^{B-1} \alpha^2 \left\|\Delta\widetilde{\mathbf{x}}^{k+t}\right\|^2 \leqslant B \sum_{k=0}^{n+B-1} \alpha^2 \left\|\Delta\widetilde{\mathbf{x}}^k\right\|^2. \tag{A17}$$

对式 (A7) 中的项 iii 进行界定, 具体如下所述.

引理A9 假设 $\alpha \leqslant \alpha_{\max}$, 则

$$\begin{aligned}
&\xi_y^{-1}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_y^{n+B+k}\right\|^2\\
&+\frac{1}{\mu_{\min}}\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+B+k}\right\|^2\\
&\leqslant \xi_y^{-1}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_y^{n+k}\right\|^2+\frac{c_L\xi_x^{-1}}{\mu_{\min}}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+k}\right\|^2\\
&\underbrace{-\sum_{k=0}^{B-1}\left(c_L\xi_x^{-1}E_{s\perp}^{n+k}+\xi_y^{-1}E_{y\perp}^{n+k}\right)}_{\text{term iii}}+\frac{\xi_y^{-1}c_\wedge}{\mu_{\min}}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+k}\right\|^2\\
&+\left(\frac{c_\triangle}{\mu_{\min}}\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)+\xi_y^{-1}c_\perp\right)\sum_{k=0}^{B-1}E_{\Delta\widetilde{\mathbf{x}}}^{n+k},
\end{aligned}\tag{A18}$$

其中

$$\mu_{\min}=1-\frac{B}{1-\widetilde{\rho}}\left(\left(B+\frac{1}{\varepsilon}+\frac{B}{\varepsilon}\right)\left(\beta+\beta\sqrt{I}\right)^2+2\left(B+\frac{1}{\varepsilon}+B\varepsilon\right)(1+\beta\sqrt{I})^2\left(\alpha_{\max}\right)^2\right),\tag{A19}$$

$$c_\wedge=\frac{B}{1-\widetilde{\rho}}\left(L_{\max}^2\left(B+\frac{1}{\varepsilon}+B\frac{1}{\varepsilon}\right)\left(2+\beta+\beta\sqrt{I}\right)^2+2L_{\max}^2\left(B+\frac{1}{\varepsilon}+B\varepsilon\right)(1+\beta\sqrt{I})^2\left(\alpha_{\max}\right)^2\right).\tag{A20}$$

证明 将式 (A13) 的两边同时乘以 ξ_y^{-1}, 可得

$$\begin{aligned}
&\xi_y^{-1}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_y^{n+B+k}\right\|^2\\
&\leqslant \xi_y^{-1}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_y^{n+k}\right\|^2-\xi_y^{-1}\sum_{k=0}^{B-1}E_{y\perp}^{n+k}+\xi_y^{-1}c_\Box\sum_{k=0}^{B-1}E_{s\perp}^{n+k}+\xi_y^{-1}c_\perp\sum_{k=0}^{B-1}E_{\Delta\widetilde{\mathbf{x}}}^{n+k}\\
&\leqslant \xi_y^{-1}\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_y^{n+k}\right\|^2\\
&-\sum_{k=0}^{B-1}\left(c_L\xi_x^{-1}E_{s\perp}^{n+k}+\xi_y^{-1}E_{y\perp}^{n+k}\right)+\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}E_{s\perp}^{n+k}+\xi_y^{-1}c_\perp\sum_{k=0}^{B-1}E_{\Delta\widetilde{\mathbf{x}}}^{n+k}.
\end{aligned}\tag{A21}$$

由于 $\alpha \leqslant \alpha_{\max}$, $\mu \geqslant \mu_{\min}$. 所以, 由式 (A12) 可得

$$\begin{aligned}
&\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+B+k}\right\|^2\\
&\leqslant \sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+k}\right\|^2-\mu_{\min}\sum_{k=0}^{B-1}E_{s\perp}^{n+k}+c\Delta\sum_{k=0}^{B-1}E_{\Delta\widetilde{\mathbf{x}}}^{n+k}.
\end{aligned}\tag{A22}$$

将上述不等式的两边同时乘以 $(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge)/\mu_{\min}$, 并利用 $\alpha \leqslant \alpha_{\max}$ 的事实, 可得

$$\begin{aligned}
&\frac{1}{\mu_{\min}}\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+B+k}\right\|^2\\
&\leqslant \frac{1}{\mu_{\min}}\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}\frac{k+1+(B-k-1)\widetilde{\rho}}{1-\widetilde{\rho}}\left\|\mathbf{e}_s^{n+k}\right\|^2\\
&-\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}E_{s\perp}^{n+k}+\frac{c_\triangle}{\mu_{\min}}\left(c_L\xi_x^{-1}+\xi_y^{-1}c_\wedge\right)\sum_{k=0}^{B-1}E_{\Delta\widetilde{\mathbf{x}}}^{n+k}.
\end{aligned}\tag{A23}$$

将式 (A23) 加入式 (A21) 后, 即可得到所需结果.

Distributed Nesterov-like accelerated non-convex composite optimization algorithm

Tiancheng LI, Kunpeng ZHANG, Lei XU, Chao GAO, Fan LI & Tao YANG*

State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China
* Corresponding author. E-mail: yangtao@mail.neu.edu.cn

Abstract With the rapid development of intelligent industrial systems and the widespread application of the energy internet, industrial software has played a critical role in supporting multi-node collaborative computation and solving complex optimization problems. However, the distributed and multi-node autonomous characteristics of the energy internet, along with its complex network structures and real-time dynamic regulation requirements, impose higher demands on the real-time performance, convergence rate, and applicability of industrial software. These requirements urgently necessitate the design of efficient distributed optimization algorithms as the core support for industrial software. Therefore, this paper investigates distributed non-convex composite optimization problems under time-varying communication topologies, where the global objective function comprises a smooth non-convex part and a non-smooth convex part. The proposed algorithm leverages the successive convex approximation (SCA) technique and gradient tracking mechanism, while incorporating a Nesterov-like momentum term to adjust the update direction in each iteration, thereby further enhancing the convergence rate of the algorithm. Theoretical analysis proves that, under a fixed step size and when the momentum parameter is below a specified upper bound, the proposed algorithm asymptotically converges to the equilibrium point set of the studied problem. Numerical simulations further validate the effectiveness of the algorithm.

Keywords industrial software, distributed non-convex optimization, composite optimization, Nesterov-like acceleration method