arXiv:2304.08051v1 [math.OC] 17 Apr 2023

加速分布式聚合优化

Jiaxu Liu、Song Chen、Shengze Cai、Chao Xu，*IEEE* 会员

摘要—**本文研究网络中的分布式聚合优化问题，其中每个智能体都有自己的局部代价函数，该函数不仅取决于局部状态变量，还取决于所有智能体状态变量的聚合函数。为加快优化过程，我们将重球法和 Nesterov 加速法与分布式聚合梯度跟踪相结合，并提出两种用于求解分布式聚合优化问题的新算法，分别称为 DAGT-HB 和 DAGT-NES。我们分析了 DAGT-HB 和 DAGT-NES 算法在目标函数光滑且强凸，并且参数（例如步长和动量系数）取值在一定范围内时，能够以全局 R–线性收敛率收敛至最优解。我们通过最优布置问题的数值实验验证了所提算法的有效性和优越性。**

索引词—**分布式优化、聚合优化、重球法、Nesterov 加速方法、梯度跟踪、Jury判据。**

I. 引言

分布式优化在机器学习中发挥着关键作用，尤其是在数据集或模型过大而无法装入单台机器的场景中。通过将数据集或模型分布到多台并行机器上并应用分布式优化，可以实现更高效的计算和可扩展的机器学习[1]–[4]。此外，分布式优化因其在编队控制 [5], 、传感器网络 [6], 、资源分配 [7]等领域的广泛应用，也受到了控制领域的极大关注。一般而言，在分布式优化问题中，每个智能体只能访问自身的信息，并与邻居交换数据，以协同最小化全局代价函数。这些智能体可以通过通信与协作并行工作，从而加快优化过程，并能够处理更大的数据集。

为了解决分布式优化问题，研究人员提出了各种卓越的算法。基于一致性方案，一些已知的分布式优化算法通常采用梯度下降，例如分布式次梯度下降[8], 、分布式对偶平均梯度算法 [9], 、push-sum 分布式算法 [10] 等。然而，上述算法都存在收敛速度慢的缺点，这是因为算法设计中采用了逐渐趋于零的步长。采用恒定步长可以提高分布式梯度算法的收敛率，但由于每个智能体使用局部梯度，这类算法只能收敛到最优解的一个小邻域内 [11]。为解决收敛速度慢和无法获得最优解的问题，研究者将梯度跟踪策略融入分布式凸优化算法，以全局平均梯度的估计值替代每个智能体的局部梯度 [12], [13]。在 [14],中，提出了一种名为 EXTRA 的算法，通过引入累积校正项，实现了向全局最优解的几何收敛。

此外，为了提高收敛速度，可以将重球法 [15] 和 Nesterov 加速方法 [16] 等加速算法应用于分布式优化。在[17], 中，提出了分布式 Nesterov 梯度法（D-NG），可将收敛率提高至 $O(\frac{\log k}{k})$。在 [18],中，研究者通过将 AB [19] 与 Nesterov 动量相结合，提出了两种适用于强连通有向网络的分布式 Nesterov 梯度方法。随后，在 [20], 中，作者将AB 与重球法相结合，提出了一种名为 ABm 的分布式重球算法。研究还证明，当步长和动量参数为正且足够小时，ABm 具有全局 R–线性收敛率。Qu 和 Li [21] 进一步将梯度跟踪与分布式 Nesterov 梯度下降方法相结合，提出了两种加速的分布式 Nesterov 方法，分别称为 Acc-DNGD-SC 和 Acc-DNGD-NSC。此外，[22]中还提出了一种名为 D-DNGT 的分布式类 Nesterov 梯度跟踪算法。该算法将梯度跟踪融入带有动量项的分布式 Nesterov 方法，并采用非均匀步长。这些算法在光滑强凸目标函数下具有线性收敛率；与不引入动量项的算法相比，它们显著提高了收敛速度。

在上述分布式优化问题中，主要形成的研究框架称为共识优化。然而，在其他一些场景中，例如多智能体编队控制问题、多机器人监视场景 [23]和聚合博弈[24], ，每个智能体的目标函数不仅取决于其局部状态，还通过聚合变量受其他智能体变量的影响。这类优化问题称为 [25], 中的分布式聚合优化，其中还提出了一种分布式聚合梯度跟踪（DAGT）方法来解决该问题。沿着 [25], 的研究，Li 等人 [26]考虑了带有聚合变量的在线凸优化，以处理时变成本函数。随后，[27]研究了带约束的在线分布式聚合优化问题。Chen 和 Liang [28] 考虑了有限比特通信，并提出了一种新颖的分布式量化算法，称为 D-QAGT。在 [29], 中，分布式凸聚合优化进一步与 Frank-Wolfe 算法相结合，以解决时变通信图上的聚合优化问题。

据我们所知，目前鲜有研究致力于加快分布式聚合优化的收敛速度。因此，本文提出了两种新颖的加速算法，用于

本研究得到科技创新2030——新一代人工智能重大项目（编号：2018AAA0100902）、国家重点研发计划（项目编号：2019YFB1705800）、国家自然科学基金（编号：61973270）以及浙江省自然科学基金（编号：LY21F030003）的资助。

Jiaxu Liu 就职于浙江大学数学科学学院，地址：中国浙江省杭州市310027。（电子邮箱：jiax-uliu@zju.edu.cn）。

Song Chen 就职于浙江大学数学科学学院，地址：中国浙江省杭州市310027。（电子邮箱：math cs@zju.edu.cn）。

Shengze Cai 就职于浙江大学工业控制技术国家重点实验室及网络系统与控制研究所，地址：中国浙江省杭州市310027。（电子邮箱：shengze cai@zju.edu.cn）。

Chao Xu 任职于浙江大学工业控制技术国家重点实验室、网络系统与控制研究所，地址为中国浙江省杭州市 310027；同时任职于浙江大学湖州研究院，地址为中国浙江省湖州市 313000。（电子邮件：cxu@zju.edu.cn）。

分布式聚合优化问题。我们的主要贡献可概括如下。

1）我们将加速算法与分布式聚合梯度跟踪（DAGT）相结合 [25], , 由此提出了两种新算法，分别称为 DAGT-HB 和 DAGT-NES。2）我们从理论上证明，当目标函数光滑且强凸时，算法 DAGT-HB 和 DAGT-NES 能够以全局R–线性收敛率收敛至最优解。此外，我们还给出了参数（例如步长和动量项）的适当取值范围。3）数值仿真验证了所提出的 DAGT-HB 和 DAGT-NES 的有效性和优越性，支持了本文的理论结果。

本文其余部分安排如下。第 2 节介绍基本符号、图的基本定义以及分布式聚合优化问题。第 3 节和第 4 节分别介绍 DAGT-HB 和 DAGT-NES，并分析它们的收敛率。第 5 节给出一个数值示例，以验证所提出的算法。最后，第 6 节对全文进行总结。

II. 预备知识与问题设置

A. 基本符号

实数集和正实数集分别记为 R 和 R^{+}， n 维列向量集合记为 R^{n}。令 $\mathbf{1}_n$ 和 $\mathbf{0}_n$ 分别表示由 n 个 1 和 0 组成的列向量。I_n 表示 n 维单位矩阵。令 $\|\cdot\|$ 和 $x^{\top}$ 分别表示标准欧几里得范数（或诱导矩阵范数）以及 $x \in \mathrm{R}^{n}$ 的转置。对于任意向量 $a = \ a_1\ a_2 \quad a_n \quad b = \ b_1\ b_2 \quad b_n \quad a > (, , . b \iff a_i > b_i \ \ i = 1\ 2 \ \ldots \ n \quad x_1 \cdots \ x_N \in \ ^{n} . . ,)^{\top}$ 和 $(, , \ldots ,)^{\top} \in \mathrm{R}^{n}$，我们使用逐元素序关系 $>$，即 , , , , 。对于向量 , , R，我们用记号 $x = \mathrm{col}\,(x_1, \cdots, \ x_N) = \left(x_1^{\top}, \cdots, x_N^{\top}\right)^{\top}$ 表示一个新的堆叠向量。此外，我们用 $\mathrm{blkdiag}(A_1, A_2, \ldots, A_N)$ 表示分块对角矩阵，其中第 i 个对角块由矩阵 $A_i \in \mathrm{R}^{m_i \times n_i}$, $i = 1, 2, \ldots, n$ 给出。任意矩阵 $A \in \mathrm{R}^{m \times n}$ 和 $B \in \mathrm{R}^{p \times q}$ 的克罗内克积定义为 $A \otimes B \in \mathrm{R}^{mp \times nq}$。令 $\rho(A)$ 表示方阵 A 的谱半径。 $A > 0$ 表示 A 为正，即 A 的每个元素都大于 0。令 $K_N = \ ^{1}_{N} 1_N 1_N^{\top}$。此外， $\nabla f(\cdot)$ 表示可微函数 $f(\cdot)$ 的梯度。

B. 图论

这里给出一些图论的基本定义。令 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ 表示一个加权无向图，其顶点集 $\mathcal{V} = \{1, \ldots, N\}$ 有限，边集为 $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$，加权邻接矩阵为 $A = [a_{ij}] \in \mathbb{R}^{N \times N}$，其中若 $(j, i) \in \mathcal{E}$，则 $a_{ij} > 0$；否则为 $a_{ij} = 0$。 $\mathcal{N}_i = \{j : (j, i) \in \mathcal{E}\}$ 表示智能体 i 的邻居集合， $d_i = \sum_{j=1}^{N} a_{ij}$ 表示顶点 i 的加权度。若对于任意 $i, j \in \mathcal{V}$，都存在一条从 i 到 j 的路径，则称图 $\mathcal{G}$ 为连通图。图 $\mathcal{G}$ 的拉普拉斯矩阵为 $L = D - A$，其中 $D =$ 为对角矩阵 $\{d_1, \ldots, d_N\}$。 L 的实特征值记为 $\lambda_1, ..., \lambda_N$，其中 $\lambda_i \leq \lambda_{i+1}$, $i = 1, \ldots, N-1$。无向图 $\mathcal{G}$ 连通，当且仅当 $\lambda_2 > 0$。

C. 问题表述

本文考虑如下分布式聚合优化问题：

$$\min_{x \in \mathbb{R}^n} f(x) = \sum_{i=1}^{N} f_i\left(x_i, u(x)\right), \quad u(x) = \frac{\sum_{i=1}^{N} \phi_i\left(x_i\right)}{N}, \tag{1}$$

其中，$x = \quad x_1 \quad x_N \quad x_i \in \mathbb{R}^{n_i}\ n \quad \mathbb{R}^{n_i} \to \mathbb{R}$ col (, ……,) 是全局状态变量，且 , : $= \sum_{i=1}^{N} n_i$； f_i : 是局部目标函数。在问题（1）中， $u(x)$ 是聚合变量，可以获取所有智能体的信息，而函数 ϕ_i: $\mathbb{R}^{n_i} \to \mathbb{R}^{d}$ 仅智能体 i 可以访问。此外，每个智能体 i 只知道状态变量 x_i 的信息，无法获取其他状态变量的信息。并且，每个智能体只能私下访问关于 f_i 的信息。本文旨在设计一种分布式优化算法，以求得问题（1）的最优状态变量。

为简化以下分析，对于所有 $i \in \{1, 2, \cdots, N\}$，令 $\nabla_1 f_i(x_i, u(x))$ 和 $\nabla_2 f_i(x_i, u(x))$ 分别表示 $\nabla_{x_i} f_i(x_i, u(x))$ 和 $\nabla_u f_i(x_i, u(x))$。对于 $x = \mathrm{col}(x_1, \ldots, x_N) \in \mathbb{R}^{n}$ 和 $y = \mathrm{col}(y_1, \ldots, y_N) \in \mathbb{R}^{Nd}$，定义 $f(x, y) = \sum_{i=1}^{N} f_i(x_i, y_i)$，$\nabla_1 f(x, y) = \mathrm{col}(\nabla_1 f_1(x_1, y_1) \ldots, \nabla_1 f_N(x_N, y_N))$，以及 $\nabla_2 f(x, y) = \mathrm{col}(\nabla_2 f_1(x_1, y_1) \ldots, \nabla_2 f_N(x_N, y_N))$。此外，记 $\nabla \phi(x) = \mathrm{blkdiag}(\nabla \phi_1(x_1) . \ldots, \nabla \phi_N(x_N)) \in \mathrm{R}^{n \times Nd}$。接下来，对于可微函数 $h(x) = \mathrm{col}\,(h_1(x) \ldots, h_m(x))$，其中 h_i 是实值函数，记 $\nabla h(x) = (\nabla h_1(x) .. , \nabla h_m(x))$，其中 ∇h_i 是 h_i 的梯度（列向量）， $i = 1, 2, ... , m$。

为便于后续分析，有必要给出一些定义和一般性假设。首先，分别给出 L-光滑函数、 m-强凸函数和**R**-线性收敛的定义。

定义 1 (光滑性)：若可微函数 $f : \mathbb{R}^n \to \mathbb{R}$满足对于所有 $x, y \in \mathbb{R}^n$，则称其为 L-光滑函数。

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|. \tag{2}$$

根据 [30], , 这等价于对于所有 $x, y \in \mathbb{R}^n$

$$f(y) \leq f(x) + \nabla f(x)^{\top}(y - x) + \frac{L}{2}\|y - x\|^2. \tag{3}$$

定义 2 (强凸性)：若可微函数 $f : \mathbb{R}^n \to \mathbb{R}$ 满足对于所有 $x, y \in \mathbb{R}^n$，则称其为 m-强凸函数。

$$m\|x - y\|^2 \leq (x - y)^{\top}(\nabla f(x) - \nabla f(y)). \tag{4}$$

这也意味着对于所有 $x, y \in \mathbb{R}^n$

$$f(x) + \nabla f(x)^{\top}(y - x) + \frac{m}{2}\|y - x\|^2 \leq f(y). \tag{5}$$

定义 3 [31]: 若存在常数 $c > 0$，使得序列 $\{x_k\}$ 以速率 $\rho \in$ 收敛**R**-线性收敛到 x_*，其中 0,1

$$||x_k - x_*|| \leq c\rho^k \quad \forall k \in N_+. \tag{6}$$

接下来，我们对图 $\mathcal{G}$ 和目标函数 $f(x)$作一些常见假设。

假设 1：图 $\mathcal{G}$是连通的，且 A是双随机的，即，对于所有 i， $j = 1$， …， N， $\sum_{i=1}^{N} a_{ij} = 1$且 $\sum_{j=1}^{N} a_{ij} = 1$。

假设 2：全局目标函数 f是 L_1-光滑且 $\nabla_1 f\ x\ y\ + \nabla \phi\ x\ \mathbf{1}_N \otimes$-强凸的。这意味着 (,) ( ) $\frac{1}{N}\sum_{i=1}^{m}\nabla_2 f_i\ x_i\ y_i \quad L_1$ (,–) 是 -利普希茨连续的。

假设 3： $\nabla_2 f(x, \ y)$ 是 L_2-利普希茨连续的。

假设 4：对于所有 $i = 1$， 2， …， N，聚合函数 ϕ_i 可微且 L_3-利普希茨连续。

为进行以下分析，下面列出几个重要引理。

引理 1 [25]: 设 F: $\mathbb{R}^n \to \mathbb{R}$ 为 μ-强凸且 L-光滑。则对所有 $x, y \in \mathbb{R}^n$， $\|x - \alpha\nabla F(x) - (y - \alpha\nabla F(y))\| \leq (1 - m\alpha)\|x - y\|$，其中 $\alpha \in (0, \ 1/L]$。

引理 2 [32]: 在假设1下，对于邻接矩阵 A，以下性质成立：

1) $\mathcal{A}\mathcal{K}=\mathcal{K}\mathcal{A}=\mathcal{K}$，其中 $\mathcal{A}=A\otimes I_d$, $\mathcal{K}=K_N\otimes I_{d^\circ}$ 2) 对任意 $x\in\mathbb{R}^{Nd}$，$\|\mathcal{A}x-\mathcal{K}x\|\le\rho\|x-\mathcal{K}x\|$ $\rho=\|A-K_N\|<1$ 且。3) $||A-I_N||\le 2_\circ$

引理 3 [32]: 设 $W\in\mathbb{R}^{n\times n}$ 为非负数，$x\in\mathbb{R}^n$ 为正数。若 $Wx<\lambda x$ 且满足 $\lambda>0$，则 $\rho(W)<\lambda_\circ$

引理 4 [33]: 设 $\mathbb{R}$ 为实数域，$H(z)$表示次数为 n的实系数多项式

$$H(z)=a_nz^n+a_{n-1}z^{n-1}+\cdots+a_2z^2+a_1z+a_0,$$

其中 $a_0,\ a_1,\ \ldots,\ a_n\in\mathbb{R}_\circ$ $H(z)$的Jury矩阵可写为 as

	z^0	z^1	z^2	$\cdots$	z^{n-2}	z^{n-1}	z^n
1	a_0	a_1	a_2	$\cdots$	a_{n-2}	a_{n-1}	a_n
2	a_n	a_{n-1}	a_{n-2}	$\cdots$	a_2	a_1	a_0
3	b_0	b_1	b_2	$\cdots$	b_{n-2}	b_{n-1}	
4	b_{n-1}	b_{n-2}	b_{n-3}	$\cdots$	b_1	b_0	
5	c_0	c_1	c_2	$\cdots$	c_{n-2}		
6	c_{n-2}	c_{n-3}	c_{n-4}	$\cdots$	c_0		
$\vdots$	$\vdots$	$\vdots$	$\vdots$				
$2n-5$	l_0	l_1	l_2	l_3			
$2n-4$	l_3	l_2	l_1	l_0			
$2n-3$	m_0	m_1	m_2				

where

$$b_i=\begin{vmatrix}a_0 & a_{n-i}\\ a_n & a_i\end{vmatrix},\quad i=0,1,2,\ldots,n-1,$$

$$c_j=\begin{vmatrix}b_0 & b_{n-j-1}\\ b_{n-1} & b_j\end{vmatrix},\quad j=0,1,2,\ldots,n-2,$$

$$\cdots$$

$$m_0=\begin{vmatrix}l_0 & l_3\\ l_3 & l_0\end{vmatrix},\quad m_1=\begin{vmatrix}l_0 & l_2\\ l_3 & l_1\end{vmatrix},\quad m_2=\begin{vmatrix}l_0 & l_1\\ l_3 & l_2\end{vmatrix}.$$

实系数多项式$H(z)$（$n\geqslant 3$，$a_n>0$）的所有根的模均小于1，当且仅当以下四个条件成立：

1) $H(1)>0$;
2) $(-1)^nH(-1)>0$;
3) $|a_0|<a_n$;
4) $|b_0|>|b_{n-1}|,|c_0|>|c_{n-2}|,\ldots,|l_0|>|l_3|,|m_0|>|m_2|$.

III. DAGT-HB

为了解决问题 (1)，我们将分布式聚合梯度跟踪（DAGT）[25] 与重球法相结合，提出如下 DAGT-HB算法：

$$\begin{aligned}x_{i,k+1}=x_{i,k}&-\alpha\left[\nabla_1 f_i\left(x_{i,k},u_{i,k}\right)+\nabla\phi_i\left(x_{i,k}\right)s_{i,k}\right]\\&+\beta(x_{i,k}-x_{i,k-1}),\end{aligned}\tag{7}$$

$$u_{i,k+1}=\sum_{j=1}^{N}a_{ij}u_{j,k}+\phi_i\left(x_{i,k+1}\right)-\phi_i\left(x_{i,k}\right),\tag{8}$$

$$\begin{aligned}s_{i,k+1}=&\sum_{j=1}^{N}a_{ij}s_{j,k}+\nabla_2 f_i\left(x_{i,k+1},u_{i,k+1}\right)\\&-\nabla_2 f_i\left(x_{i,k},u_{i,k}\right).\end{aligned}\tag{9}$$

在 DAGT-HB 中，我们引入动量项 $x_{i,\ k}-x_{i,\ k-1}$ 以加快算法的收敛率。由于分布式聚合优化不再依赖一致性协议，而是通过聚合变量将智能体耦合在一起，因此每个智能体只需达到其自身局部目标函数的最优点。直观地说，对于每个智能体，如果算法生成的迭代点始终朝着最优解的方向移动，那么沿同一方向加入动量必然能加快算法的收敛速度。由于 $u(x)$ 是所有智能体都无法直接获取的全局信息，因此引入 $u_{i,\ k}$，供智能体 i 跟踪平均值 $u(x)$。与此同时，$s_{i,\ k}$ 跟踪梯度和 $\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_{i,\ k},\ u(x)\right)$，这一信息同样无法被所有智能体获取。

DAGT-HB算法可以改写为如下紧凑形式：

$$x_{k+1}=x_k-\alpha\left[\nabla_1 f\left(x_k,u_k\right)+\nabla\phi\left(x_k\right)s_k\right]+\beta(x_k-x_{k-1}),\tag{10}$$

$$u_{k+1}=\mathcal{A}u_k+\phi\left(x_{k+1}\right)-\phi\left(x_k\right),\tag{11}$$

$$s_{k+1}=\mathcal{A}s_k+\nabla_2 f\left(x_{k+1},u_{k+1}\right)-\nabla_2 f\left(x_k,u_k\right),\tag{12}$$

其中，$\mathcal{A}=A\otimes I_d$的定义见引理2，$x_k=\mathrm{col}\left(x_{1,\ k},\ \ldots,\ x_{N,\ k}\right)$，$u_k$和 s_k采用类似的记号。

首先注意到，在等式 (11) 和 (12) 两边乘以 $\frac{1}{N}$ 可得

$$\bar{u}_{k+1}=\bar{u}_k+\frac{1}{N}\sum_{i=1}^{N}\phi_i\left(x_{i,k+1}\right)-\frac{1}{N}\sum_{i=1}^{N}\phi_i\left(x_{i,k}\right),\tag{13}$$

$$\bar{s}_{k+1}=\bar{s}_k+\frac{1}{N}\sum_{i=1}^{N}\left[\nabla_2 f_i\left(x_{i,k+1},u_{i,k+1}\right)-\nabla_2 f_i\left(x_{i,k},u_{i,k}\right)\right].\tag{14}$$

然后，若我们将 $\nabla_2 f_i\ x_{i\,0}\ u_{i\,0}$ $i=1\ 2\ \ldots\ N$ $x_{i\,0}$ $s_{i\,0}=$ 以及初始化为 $,_0=\phi_i(,)$ 和 $,(,,,)$（对于 $,\ ,\ ,$），其中 $,$ 是任意的，则可推得

$$\bar{u}_k=\frac{1}{N}\sum_{i=1}^{N}u_{i,k}=\frac{1}{N}\sum_{i=1}^{N}\phi_i\left(x_{i,k}\right),\tag{15}$$

$$\bar{s}_k=\frac{1}{N}\sum_{i=1}^{N}s_{i,k}=\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_{i,k},u_{i,k}\right).\tag{16}$$

接下来，我们证明问题 (1) 的最优解与 DAGT-HB算法的不动点之间的等价性。

引理5: 在假设1和假设2下，(10)-(12)的定点是问题(1)的最优解。

证明： 将(11)-(13)的平衡点记为 $x^*=\mathrm{col}\left(x_1^*,\ldots,x_N^*\right)$、$u^*=\mathrm{col}(u_1^*,\ldots,u_N^*)$和 $s^*=\mathrm{col}\left(s_1^*,\ldots,s_N^*\right)$。由(10)-(12)，可得

$$\nabla_1 f\left(x^*,u^*\right)+\nabla\phi\left(x^*\right)s^*=\mathbf{0}_{Nd},\tag{17}$$

$$\mathcal{L}u^*=\mathbf{0}_{Nd},\quad \mathcal{L}s^*=\mathbf{0}_{Nd},\tag{18}$$

其中，$\mathcal{L}=L\otimes I_d$和 L是图 $\mathcal{G}$的拉普拉斯矩阵。根据拉普拉斯矩阵的性质，容易推得，对于所有 $i\neq j$，$u_i^*=u_j^*$和 $s_i^*=s_j^*$。由公式 (15)-(16)，可得

$$u_i^*=\frac{1}{N}\sum_{i=1}^{N}\phi_i\left(x_i^*\right)=u(x^*),\tag{19}$$

$$s_i^*=\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_i^*,u(x^*)\right).\tag{20}$$

将(19)和(20)代入(17)，可得 $\nabla f(x^*)=0_\circ$ 由于 f 是 $m-$ 强凸的，x^* 是问题(1)的唯一最优解。

■

A. 辅助结果

为了分析该算法的收敛性和收敛率，我们采用压缩状态向量方法，并收集以下四个量：

1) $||x_{k+1}-x^*||$，网络中的状态误差；2) $||x_{k+1}-x_k||$，状态差；3) $||u_{k+1}-\mathcal{K}u_{k+1}||$，聚合变量跟踪误差；4) $||s_{k+1}-\mathcal{K}s_{k+1}||$，梯度和跟踪误差。

在接下来的引理6–9中，我们推导上述四个量之间的关系。首先，我们推导$||x_{k+1}-x^*||$（网络中的状态误差）的上界。

引理6：在假设 1-4下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}&\left\|x_{k+1}-x^*\right\|\\&\leq(1-m\alpha)\left\|x_k-x^*\right\|+\alpha L_1\left\|u_k-\mathcal{K}u_k\right\|\\&\quad+\alpha L_3\left\|s_k-\mathcal{K}s_k\right\|+\beta\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{21}$$

证明：对于 $\|x_{k+1}-x^*\|$，由式（10）可得

$$\begin{aligned}&\left\|x_{k+1}-x^*\right\|=\left\|x_k-x^*-\alpha\left[\nabla_1 f(x_k,u_k)+\nabla\phi(x_k)\ s_k\right]+\beta(\right.\\&\left.x_k-x_{k-1})\right\|\leq\left\|x_k-x^*-\alpha\left[\nabla_1 f(x_k,u_k)+\nabla\phi(x_k)\right.\right.\\&\left.\left.s_k\right]\right\|+\beta\left\|x_k-x_{k-1}\right\|\leq\Big\|x_k-x^*-\alpha\left[\nabla_1 f(x_k,\mathbf{1}_N\otimes\bar{u}_k)\right.\\&\left.+\nabla\phi(x_k)\,\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\right]+\alpha\nabla f\left(x^*\right)\Big\|\\&+\alpha\Big\|\nabla_1 f\left(x_k,u_k\right)+\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\bar{s}_k-\nabla_1 f\left(x_k,\mathbf{1}_N\otimes\bar{u}_k\right)\\&-\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\Big\|\\&+\alpha\left\|\nabla\phi\left(x_k\right)s_k-\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\bar{s}_k\right\|+\beta\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{22}$$

根据引理 1，我们可以对（22）右侧第一项作如下界定:

$$\begin{aligned}&\Big\|x_k-\alpha[\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\\&\quad+\nabla_1 f\left(x_k,\mathbf{1}_N\otimes\bar{u}_k\right)]-\left[x^*-\alpha\nabla f\left(x^*\right)\right]\Big\|\\&\leq(1-m\alpha)\left\|x_k-x^*\right\|.\end{aligned}\tag{23}$$

对于第二项，由于 $f(x)$ 是 L_1–光滑的，且 $\mathbf{1}_N\otimes\bar{u}_k=\mathcal{K}u_k$，我们可以得到

$$\begin{aligned}&\alpha\Big\|\nabla_1 f\left(x_k,u_k\right)+\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\bar{s}_k-\nabla_1 f\left(x_k,\mathbf{1}_N\otimes\bar{u}_k\right)\\&\quad-\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\Big\|\\&\leq\alpha L_1\left\|u_k-\mathcal{K}u_k\right\|.\end{aligned}\tag{24}$$

对于第三项，利用假设4和 $\mathbf{1}_N\otimes\bar{s}_k=\mathcal{K}s_k$，可得以下不等式:

$$\alpha\left\|\nabla\phi\left(x_k\right)s_k-\nabla\phi\left(x_k\right)\mathbf{1}_N\otimes\bar{s}_k\right\|\leq\alpha L_3\left\|s_k-\mathcal{K}s_k\right\|.\tag{25}$$

将（23）—（25）代入（22）后，即完成证明。其次，我们推导 $\|x_{k+1}-x_k\|$的一个界。■

引理 7：在假设 1-4 下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}\left\|x_{k+1}-x_k\right\|\leq&\alpha L_1\left(1+L_3\right)\left\|x_k-x^*\right\|+\alpha L_1\left\|u_k-\mathcal{K}u_k\right\|\\&+\alpha L_3\left\|s_k-\mathcal{K}s_k\right\|+\beta\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{26}$$

证明：注意到 $\nabla f(x^*)=0$，于是我们有

$$\begin{aligned}&\left\|x_{k+1}-x_k\right\|\\&=\left\|\beta(x_k-x_{k-1})-\alpha(\nabla_1 f\left(x_k,u_k\right)+\nabla\phi\left(x_k\right)s_k)\right\|\\&\leq\alpha\Big\|\nabla_1 f\left(x_k,u_k\right)+\nabla\phi\left(x_k\right)\mathcal{K}s_k-\nabla_1 f\left(x^*,\mathbf{1}_N\otimes u^*\right)\\&\quad-\nabla\phi\left(x^*\right)\left[\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x^*,\mathbf{1}_N\otimes u^*\right)\right]\Big\|\\&\quad+\alpha\left\|\nabla\phi\left(x_k\right)\left(s_k-\mathcal{K}s_k\right)\right\|+\beta\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{27}$$

利用假设2和范数的三角不等式，我们可以得到以下公式:

$$\begin{aligned}&\Big\|\nabla_1 f\left(x_k,u_k\right)+\nabla\phi\left(x_k\right)\mathcal{K}s_k-\nabla_1 f\left(x^*,\mathbf{1}_N\otimes u^*\right)\\&\quad-\nabla\phi\left(x^*\right)\left[\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(x^*,\mathbf{1}_N\otimes u^*\right)\right]\Big\|\\&\leq L_1\left(\left\|x_k-x^*\right\|+\left\|u_k-\mathbf{1}_N\otimes u^*\right\|\right)\\&\leq L_1\left(\left\|x_k-x^*\right\|+\left\|u_k-\mathcal{K}u_k\right\|+\left\|\mathcal{K}u_k-\mathbf{1}_N\otimes u^*\right\|\right).\end{aligned}\tag{28}$$

对于 $\|\mathcal{K}u_k-\mathbf{1}_N\otimes u^*\|$，我们可以推导出

$$\begin{aligned}\left\|\mathcal{K}u_k-1_N\otimes u^*\right\|^2&=\left\|1_N\otimes\left(\bar{u}_k-u^*\right)\right\|^2\\&=N\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i\left(x_{i,k}\right)-\phi_i\left(x_i^*\right)\right)\right\|^2\\&\leq\frac{1}{N}\left(\sum_{i=1}^{N}\left\|\phi_i\left(x_{i,k}\right)-\phi_i\left(x_i^*\right)\right\|\right)^2\\&\leq\frac{1}{N}\left(\sum_{i=1}^{N}L_3\left\|x_{i,k}-x_i^*\right\|\right)^2\\&\leq L_3^2\sum_{i=1}^{N}\left\|x_{i,k}-x_i^*\right\|^2\\&=L_3^2\left\|x_k-x^*\right\|^2,\end{aligned}\tag{29}$$

其中，利用 ϕ_i 是 L_3-利普希茨连续这一性质可得到第二个不等式；应用以下事实：对于任意非负标量 s,

$\left(\sum_{i=1}^{N}a_i\right)^2\leq$，即可轻松得到最后一个不等式。因此，我们有 $N\sum_{i=1}^{N}a_i^2$ a_i

$$\left\|\mathcal{K}u_k-1_N\otimes u^*\right\|\leq L_3\left\|x_k-x^*\right\|.\tag{30}$$

然后，利用假设4可得

$$\left\|\nabla\phi\left(x_k\right)\left(s_k-\mathcal{K}s_k\right)\right\|\leq L_3\left\|s_k-\mathcal{K}s_k\right\|.\tag{31}$$

最后，将 (28)、(30) 和 (31) 代入 (27)，即可得到结果 (26)。■

下一步是对聚合变量跟踪误差$||u_{k+1}-\mathcal{K}u_{k+1}||$进行界定。

引理8：在假设 1-4 下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}&\left\|u_{k+1}-\mathcal{K}u_{k+1}\right\|\\&\leq(\rho+\alpha L_1L_3)\left\|u_k-\mathcal{K}u_k\right\|+\alpha L_1L_3\left(1+L_3\right)\left\|x_k-x^*\right\|\\&\quad+\alpha L_3^2\left\|s_k-\mathcal{K}s_k\right\|+\beta L_3\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{32}$$

Proof: For $\|u_{k+1}-\mathcal{K}u_{k+1}\|$, by invoking (11), it leads to

$$\begin{aligned}&\|u_{k+1}-\mathcal{K}u_{k+1}\|\\&=\|\mathcal{A}u_k+\phi(x_{k+1})-\phi(x_k)-\mathcal{K}\mathcal{A}u_k-\mathcal{K}[\phi(x_{k+1})-\phi(x_k)]\|\\&\le\rho\|u_k-\mathcal{K}u_k\|+\|I-\mathcal{K}\|\|\phi(x_{k+1})-\phi(x_k)\|\\&\le\rho\|u_k-\mathcal{K}u_k\|+L_3\|I-\mathcal{K}\|\|x_{k+1}-x_k\|,\end{aligned}\tag{33}$$

其中，利用引理2得到了第一个不等式；利用假设4可得到最后一个不等式。注意到$\|I-\mathcal{K}\|=1$，将 (26) 代入 (33) 即可完成证明。

■

最后，我们推导出界 $\|s_{k+1}-\mathcal{K}s_{k+1}\|$，即梯度和 $\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i(x_i,\ u(x_{k+1}))$ 的跟踪误差。

引理 9：在假设 1-4 下，以下不等式成立，$\forall k\ge 0$:

$$\begin{aligned}&\|s_{k+1}-\mathcal{K}s_{k+1}\|\\&\le(\rho+\alpha L_2L_3(1+L_3))\|s_k-\mathcal{K}s_k\|\\&\quad+(\alpha L_1L_2(1+L_3)+2L_2)\|u_k-\mathcal{K}u_k\|\\&\quad+\alpha L_1L_2(1+L_3)^2\|x_k-x^*\|+\beta L_2(1+L_3)\|x_k-x_{k-1}\|.\end{aligned}\tag{34}$$

证明：对于 $\|s_{k+1}-\mathcal{K}s_{k+1}\|$，利用式（13），可得

$$\begin{aligned}&\|s_{k+1}-\mathcal{K}s_{k+1}\|\\&=\|\mathcal{A}s_k+\nabla_2 f(x_{k+1},u_{k+1})-\nabla_2 f(x_k,u_k)\\&\qquad-\mathcal{K}\mathcal{A}s_k-\mathcal{K}[\nabla_2 f(x_{k+1},u_{k+1})-\nabla_2 f(x_k,u_k)]\|\\&\le\|\mathcal{A}s_k-\mathcal{K}s_k\|\\&\quad+\|I-\mathcal{K}\|\|\nabla_2 f(x_{k+1},u_{k+1})-\nabla_2 f(x_k,u_k)\|\\&\le\rho\|s_k-\mathcal{K}s_k\|+\|\nabla_2 f(x_{k+1},u_{k+1})-\nabla_2 f(x_k,u_k)\|\\&\le\rho\|s_k-\mathcal{K}s_k\|+L_2(\|x_{k+1}-x_k\|+\|u_{k+1}-u_k\|),\end{aligned}\tag{35}$$

其中，利用引理2得到了第一个不等式，并在最后一个不等式中运用了假设2。注意，

$$\begin{aligned}&\|u_{k+1}-u_k\|\\&=\|\mathcal{A}u_k-u_k+\phi(x_{k+1})-\phi(x_k)\|\\&=\|\mathcal{A}u_k-\mathcal{A}\mathcal{K}u_k+\mathcal{K}u_k-u_k+\phi(x_{k+1})-\phi(x_k)\|\\&\le\|(\mathcal{A}-I_N\otimes I_d)(u_k-\mathcal{K}u_k)\|+\|\phi(x_{k+1})-\phi(x_k)\|\\&\le\|A-I_N\|\|u_k-\mathcal{K}u_k\|+L_3\|x_{k+1}-x_k\|\\&\le 2\|u_k-\mathcal{K}u_k\|+L_3\|x_{k+1}-x_k\|.\end{aligned}\tag{36}$$

然后，将式（36）和式（26）代入式（35），即可完成证明。

■

B. 主要结果

现在我们给出本节的主要结果。基于引理6—9，我们在以下定理中给出DAGT-HB算法的收敛性和收敛率。

定理 1：在假设 $1-4$ 下，若

$$(\alpha,\beta)\in\bigcap_{i=1}^{6}\mathcal{S}_i,\tag{37}$$

其中 $\mathcal{S}_i, i=1,\cdots,6$ 在以下证明中定义，则由 DAGT-HB 生成的 $x_k=\text{col}(x_{1,k},\ldots,x_{N,k})$ 能以 R–线性收敛率收敛到问题 (1) 的优化器。

证明：记

$$V_k=\text{col}\left(\|x_k-x^*\|,\|x_k-x_{k-1}\|,\|u_k-\mathcal{K}u_k\|,\|s_k-\mathcal{K}s_k\|\right).\tag{38}$$

由引理6—9可得

$$V_{k+1}\le PV_k,\tag{39}$$

其中

$$P=\begin{bmatrix}1-m\alpha & \beta & \alpha L_1 & \alpha L_3\\ \alpha L_1(1+L_3) & \beta & \alpha L_1 & \alpha L_3\\ \alpha L_1L_3(1+L_3) & \beta L_3 & \rho+\alpha L_1L_3 & \alpha L_3^2\\ \alpha L_1L_3(1+L_3)^2 & \beta L_2(1+L_3) & p_{43} & p_{44}\end{bmatrix},\tag{40}$$

$p_{43}=\alpha L_1L_3(1+L_3)+2L_2$ 和 $p_{44}=\rho+\alpha L_2L_3(1+L_3)$. 首先，基于引理 3，我们寻找 α 和 β 的一个较小范围，使其满足 $\rho(P)<1$。我们定义一个正向量 $z=[z_1,z_2,z_3,z_4]^T$，使得

$$Pz<z,\tag{41}$$

这等于

$$\begin{aligned}&0<\beta<\frac{\alpha(mz_1-L_1z_3-L_3z_4)}{z_2}=M_1,\\&0<\beta<\frac{z_2-\alpha L_1(1+L_3)z_1-\alpha L_1z_3-\alpha L_3z_4}{z_2}=M_2,\\&0<\beta<\\&\frac{(1-\rho-\alpha L_1L_3)z_3-\alpha L_1L_3(1+L_3)z_1-\alpha L_3^2z_4}{L_3z_2}=M_3,\\&0<\beta<\frac{[1-\rho-\alpha L_2L_3(1+L_3)]z_4-\alpha L_1L_2(1+L_3)^2z_1}{L_2(1+L_3)z_2},\\&-\frac{[\alpha L_1L_2(1+L_3)+2L_2]z_3}{L_2(1+L_3)z_2}=M_4.\end{aligned}\tag{42}$$

由上述不等式可得

$$\begin{aligned}&0<\alpha<\frac{z_2}{L_1(1+L_3)z_1+L_1z_3+L_3z_4}=J_1,\\&0<\alpha<\frac{1-\rho}{L_1L_3}=J_2,\\&0<\alpha<\frac{(1-\rho)z_3}{L_1L_3(1+L_3)z_1+L_1L_3z_3+L_3^2z_4},\\&0<\alpha<\frac{1-\rho}{L_2L_3(1+L_3)}=J_3,\\&0<\alpha<\frac{(1-\rho)z_4-2L_2z_3}{L_2(1+L_3)[L_1(1+L_3)z_1+L_1z_3+L_3z_4]}=J_4,\\&z_1>\frac{L_1z_3+L_3z_4}{m},\\&z_4>\frac{2L_2z_3}{1-\rho}.\end{aligned}\tag{43}$$

也就是说，当时，我们可以任意选取 z_2 和 z_3

$$0<\alpha<\bar{\alpha}=\min\{J_1,J_2,J_3,J_4,\frac{1}{L_1}\},\tag{44}$$

and

$$0<\beta<\bar{\beta}=\min\{M_1,M_2,M_3,M_4\},\tag{45}$$

其中 $z_2>0,\ z_3>0,\ z_1>\frac{L_1z_3+L_3z_4}{\mu}$和 $z_4>\frac{2L_2z_3}{1-\rho}$，我们有 $\rho(P)<1$。

接下来，我们使用 Jury判据来求得满足 $\rho(P)<1$ 的 α 和 β 的精确范围。通过计算，我们可以得到 P 的特征多项式：

$$H(\lambda)=|\lambda I-P|=a_0+a_1\lambda+a_2\lambda^2+a_3\lambda^3+\lambda^4,\tag{46}$$

其中

$$
\begin{aligned}
a_0 &= \beta d_1\rho(\rho+2\alpha d_2),\\
a_1 &= \beta[-d_1\rho+(d_1+\rho)\times(-\rho+2\alpha d_2)]+(m\alpha-1)\rho(\rho+\alpha d_2)\\
&\quad -\alpha L_3 d_1\times[\alpha L_1(\rho+\alpha L_3 d_2)+\alpha L_3 d_3],\\
a_2 &= \beta(d_1+2\rho+\alpha d_2)+(1-m\alpha)(2\rho+\alpha d_2)+\rho(\rho+\alpha d_2)\\
&\quad +\alpha L_3 d_1[L_1+L_3(1+L_3)]+L_3[\alpha L_1(\rho+\alpha d_2)+\alpha L_3 d_3],\\
a_3 &= -\beta+(m-2)\alpha-1-\alpha L_3[L_2(1+L_3)+L_1],
\end{aligned}
\tag{47}
$$

且 $d_1=1-m\alpha-\alpha L_1(1+L_3)$、$d_2=L_3(1+L_3)(L_2-L_3)$ 和 $d_3=-2\alpha L_2+\rho(1+L_3)$。于是可以得到：

$$
\begin{aligned}
&b_0=a_0^2-1, b_1=a_0a_1-a_3, b_2=a_0a_2-a_2, b_3=a_0a_3-a_1\\
&c_0=b_0^2-b_3^2, c_1=b_0b_1-b_2b_3, c_2=b_0b_2-b_1b_3.
\end{aligned}
\tag{48}
$$

下面记

$$
\begin{aligned}
&\mathcal{S}_1=\{(\alpha,\beta)|a_0+a_1+a_2+a_3+1>0\},\\
&\mathcal{S}_2=\{(\alpha,\beta)|a_0-a_1+a_2-a_3+1>0\},\\
&\mathcal{S}_3=\{(\alpha,\beta)||a_0|<1\},\quad \mathcal{S}_4=\{(\alpha,\beta)||b_0|>|b_3|\}\\
&\mathcal{S}_5=\{(\alpha,\beta)||c_0|>|c_2|\},\quad \mathcal{S}_6=\{(\alpha,\beta)|\alpha>0,\beta>0\}.
\end{aligned}
\tag{49}
$$

因此，根据引理4，当 $(\alpha,\beta)\in\bigcap_{i=1}^{6}\mathcal{S}_i$ 时，矩阵 P 的谱半径小于 1。鉴于上述分析，我们知道 $\bigcap_{i=1}^{6}\mathcal{S}_i$ 非空。最后，记 $\rho_1=\rho(P)$ 且 $0<\rho_1<1$，则可得

$$
||V_{k+1}||\le\rho_1||V_k||. \tag{50}
$$

此外，可得

$$
||x_k-x^*||\le||V_k||\le C_1\rho_1^k, \tag{51}
$$

其中 $C_1=||V_0||$。因此，DAGT-HB 能够实现 **R**-线性收敛速率。至此，证明完成。■

注 1： 在定理 1 中，我们已经证明，当步长 α 且动量参数 β 遵循（37）时，DAGT-HB 具有 **R**-线性收敛速率。但我们承认，定理 1 中 α 和 β 的理论界较为保守。如何获得理论边界，乃至最优参数，将在我们未来的工作中予以研究。

推论 1： 在定理 1 的相同假设下，有以下等式：

$$
|f(x_k)-f(x^*)|\le\frac{L_1}{2}C_1^2\rho_1^{2k}. \tag{52}
$$

证明：由于 $f(x)$ 是 L_1-光滑的，且 $\nabla f(x^*)=0$，我们可以得到

$$
f(x_k)-f(x^*)\le\frac{L_1}{2}||x_k-x^*||^2. \tag{53}
$$

将 (51) 代入 (53)，即可完成证明。■

作为本节的总结，DAGT-HB 方法可表述为如下算法1。

IV. DAGT-NES

除了重球法，Nesterov算法也是一种广为人知的加速方法，可以与 DAGT 结合，以

算法1： DAGT-HB

Input : initial point $x_{i,-1}$ and $x_{i,0}\in\mathbb{R}^{n_i}$, $u_{i,0}=\phi_i(x_{i,0})$ and $s_{i,0}=\nabla_2 f_i(x_{i,0},u_{i,0})$ for $i=1,2,\dots,N$, let $(\alpha,\beta)\in\bigcap_{i=1}^{6}\mathcal{S}_i$, set $k=0$, select appropriate $\epsilon>0$ and $K_{max}\in\mathbb{N}^+$.

Output: optimal x^* and $f(x^*)$.

1 虽然 $k<K_{max}$ 和 $||\nabla f(x_k)||\ge\epsilon$ 确实

2 迭代：对每个 $i\in\{1,2,$ 进行更新 $\dots N\}$:

$x_{i,k+1}$ iterated by formula (7),

$u_{i,k+1}$ iterated by formula (8),

$s_{i,k+1}$ iterated by formula (9).

3 更新：$k=k+1$。

求解问题（1）。为此，我们提出以下 DAGT-NES 算法：

$$
x_{i,k+1}=y_{i,k}-\alpha\left[\nabla_1 f_i\left(y_{i,k},u_{i,k}\right)+\nabla\phi_i\left(y_{i,k}\right)s_{i,k}\right], \tag{54}
$$

$$
y_{i,k+1}=x_{i,k+1}+\gamma(x_{i,k+1}-x_{i,k}), \tag{55}
$$

$$
u_{i,k+1}=\sum_{j=1}^{N}a_{ij}u_{j,k}+\phi_i\left(y_{i,k+1}\right)-\phi_i\left(y_{i,k}\right), \tag{56}
$$

$$
s_{i,k+1}=\sum_{j=1}^{N}a_{ij}s_{j,k}+\nabla_2 f_i\left(y_{i,k+1},u_{i,k+1}\right)-\nabla_2 f_i\left(y_{i,k},u_{i,k}\right). \tag{57}
$$

同样，在 DAGT-NES 中，$u_{i,k}$ 被引入，使智能体 i 能够跟踪平均 $u(y)$，而$s_{i,k}$则跟踪梯度和$\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i(y_i,u(y))$。DAGT-NES 算法可改写为以下紧凑形式：

$$
x_{k+1}=y_k-\alpha\left[\nabla_1 f\left(y_k,u_k\right)+\nabla\phi\left(y_k\right)s_k\right], \tag{58}
$$

$$
y_{k+1}=x_{k+1}+\gamma(x_{k+1}-x_k), \tag{59}
$$

$$
u_{k+1}=\mathcal{A}u_k+\phi\left(y_{k+1}\right)-\phi\left(y_k\right), \tag{60}
$$

$$
s_{k+1}=\mathcal{A}s_k+\nabla_2 f\left(y_{k+1},u_{k+1}\right)-\nabla_2 f\left(y_k,u_k\right). \tag{61}
$$

其中 $\mathcal{A}=A\otimes I_d$ 的定义见引理2，$x_k=\mathrm{col}(x_{1,k},\dots,x_{N,k})$，$u_k$、$y_k$ 和 s_k 的记法类似。

注意，如果我们将 u 和 s 初始化为 $u_{i,0}=\phi_i(y_{i,0})$ 和 $s_{i,0}=\nabla_2 f_i(y_{i,0},u_{i,0})$，其中 $i=1,2,\dots,N$，且 $y_{i,0}$ 为任意值，那么类似于 DAGT-HB，我们可以得到

$$
\bar{u}_k=\frac{1}{N}\sum_{i=1}^{N}u_{i,k}=\frac{1}{N}\sum_{i=1}^{N}\phi_i\left(y_{i,k}\right), \tag{62}
$$

$$
\bar{s}_k=\frac{1}{N}\sum_{i=1}^{N}s_{i,k}=\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y_{i,k},u_{i,k}\right). \tag{63}
$$

接下来，我们证明问题 (1) 的最优解与 DAGT-NES 算法的不动点等价。

引理 10：在假设1和假设2下，(58)-(61) 的均衡点是问题 (1) 的最优解。

证明：见附录 A。■

A. 辅助结果

与 DAGT-HB 类似，我们采用压缩状态向量方法推导 DAGT-NES 的收敛性和收敛率，并仍然整理出以下四个量：

1) $||x_{k+1}-x^*||$，网络中的状态误差；

2) $||x_{k+1}-x_k||$，状态差；3) $||u_{k+1}-\mathcal{K}u_{k+1}||$，聚合变量跟踪误差；4) $||s_{k+1}-\mathcal{K}s_{k+1}||$，梯度和跟踪误差；

在接下来的引理 11–14 中，我们推导上述四个量之间的关系。首先，我们推导$||x_{k+1}-x^*||$（网络中的状态误差）的界。

引理 11： 在假设 1-4 下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}&\left\|x_{k+1}-x^*\right\|\\&\leq(1-m\alpha)\left\|x_k-x^*\right\|+\alpha L_1\left\|u_k-\mathcal{K}u_k\right\|\\&\quad+\alpha L_3\left\|s_k-\mathcal{K}s_k\right\|+(1-m\alpha)\gamma\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{64}$$

证明：见附录 B ∎

其次，我们推导 $\|x_{k+1}-x_k\|$的界。

引理12： 在假设 $1-4$ 下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}&\|x_{k+1}-x_k\|\\&\leq\alpha L_1\left(1+L_3\right)\left\|x_k-x^*\right\|+\alpha L_1\left\|u_k-\mathcal{K}u_k\right\|\\&\quad+\alpha L_3\left\|s_k-\mathcal{K}s_k\right\|+\gamma(1+\alpha L_1+\alpha L_1L_3)\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{65}$$

证明：见附录 C。 ∎

下一步是界定聚合变量跟踪误差$||u_{k+1}-\mathcal{K}u_{k+1}||$。

引理 13： 在假设 1-4 下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}&\|u_{k+1}-\mathcal{K}u_{k+1}\|\\&\leq[\rho+\alpha L_1L_3(\gamma+1)]\left\|u_k-\mathcal{K}u_k\right\|+\alpha L_3^2(\gamma+1)\left\|s_k-\mathcal{K}s_k\right\|\\&\quad+\alpha L_1L_3\left(1+L_3\right)(\gamma+1)\left\|x_k-x^*\right\|\\&\quad+\gamma L_3[(1+\gamma)(1+\alpha L_1+\alpha L_1L_3)+1]\left\|x_k-x_{k-1}\right\|.\end{aligned}\tag{66}$$

证明：见附录 D。 ∎

最后，我们推导梯度和 $\frac{1}{N}\sum_{i=1}^N\nabla_2 f_i\left(x_i,\ u(x_{k+1})\right)$ 跟踪误差的界$\dashv\|s_{k+1}-\mathcal{K}s_{k+1}\|$。

引理 14： 在假设 1-4 下，以下不等式成立，$\forall k\geq 0$:

$$\begin{aligned}&\|s_{k+1}-\mathcal{K}s_{k+1}\|\\&<[\rho+\alpha L_2L_3\left(1+L_3\right)(1+\gamma)]\left\|s_k-\mathcal{K}s_k\right\|\\&\quad+\alpha L_1L_2\left(1+L_3\right)^2(\gamma+1)\left\|x_k-x^*\right\|\\&\quad+[\alpha L_1L_2\left(1+L_3\right)(r+1)+2L_2]\left\|u_k-\mathcal{K}u_k\right\|\\&\quad+\gamma L_2(L_3+1)[(1+\gamma)(1+\alpha L_1+\alpha L_1L_3)+1]||x_k-x_{k-1}||.\end{aligned}\tag{67}$$

Proof: See Appendix E. ∎

B. 主要结果

总结引理 11-14，我们在下述定理中给出 DAGT-NES 算法的收敛性和收敛率。

定理 2： 在假设 1-4 下，若

$$(\alpha,\gamma)\in\bigcap_{i=1}^{6}\mathcal{T}_i,\tag{68}$$

其中，$\mathcal{T}_i$、$i=1,\cdots,6$ 在下文证明中定义，则由 DAGT-NES 生成的 $x_k=\text{col}(x_{1,k},..,x_{N,k})$ 能以 R–线性收敛率收敛到问题 (1) 的优化器。

证明：记

$$V_k=\text{col}\left(\left\|x_k-x^*\right\|,\|x_k-x_{k-1}\|,\|u_k-\mathcal{K}u_k\|,\|s_k-\mathcal{K}s_k\|\right).\tag{69}$$

由引理 11-14 可得

$$V_{k+1}\leq QV_k,\tag{70}$$

其中

$$Q=\begin{bmatrix}1-m\alpha&(1-m\alpha)\gamma&\alpha L_1&\alpha L_3\\\alpha L_1(1+L_3)&q_{22}&\alpha L_1&\alpha L_3\\\alpha L_1L_3\left(1+L_3\right)(\gamma+1)&q_{32}&q_{33}&q_{34}\\\alpha L_1L_2(1+L_3)^2(\gamma+1)&q_{42}&q_{43}&q_{44}\end{bmatrix},\tag{71}$$

and

$$\begin{cases}q_{22}=\gamma(1+\alpha L_1+\alpha L_1L_3),\\q_{32}=\gamma L_3[(1+\gamma)(1+\alpha L_1+\alpha L_1L_3)+1],\\q_{33}=\rho+\alpha L_1L_3(\gamma+1),\\q_{34}=\alpha L_3^2(\gamma+1),\\q_{42}=\gamma L_2(L_3+1)[(1+\gamma)(1+\alpha L_1+\alpha L_1L_3)+1],\\q_{43}=\alpha L_1L_2\left(1+L_3\right)(\gamma+1)+2L_2,\\q_{44}=\rho+\alpha L_2L_3\left(1+L_3\right)(1+\gamma).\end{cases}\tag{72}$$

首先，根据引理 3，我们希望找到一个较小的范围，使 α 和 γ 满足 $\rho(Q)<1$。我们定义一个正向量 $t=[t_1,t_2,t_3,t_4]^T$，使得

$$Qt<t.\tag{73}$$

然而，由于矩阵 Q中，α 和 γ 之间存在很强的非线性关系，因此难以确定使矩阵 Q 的谱半径小于 1 的 α 和 γ 的取值范围。为简化计算，我们令

$$\alpha\leq\frac{1}{L_1}\quad\text{and}\quad\gamma\leq\min\{\frac{1}{L_2},\frac{1}{L_3}\}\tag{74}$$

以消去矩阵 Q 中包含 α 和 γ之间非线性关系的一些元素。于是可得

$$V_{k+1}<RV_k,\tag{75}$$

其中

$$R=\begin{bmatrix}1-m\alpha&(1-m\alpha)\gamma&\alpha L_1&\alpha L_3\\\alpha L_1(1+L_3)&\gamma(2+L_3)&\alpha L_1&\alpha L_3\\\alpha L_1L_3\left(2+L_3\right)&r_{32}&r_{33}&\alpha L_3(L_3+1)\\r_{41}&r_{42}&r_{43}&r_{44}\end{bmatrix},\tag{76}$$

and

$$\begin{cases}r_{32}=\gamma(L_3^2+4L_3+2),\\r_{33}=\rho+\alpha L_1(L_3+1),\\r_{41}=\alpha L_1(1+L_2)(1+L_3)^2,\\r_{42}=\gamma(L_3+1)(L_2L_3+2L_2+L_3+1),\\r_{43}=\alpha L_1(L_2+1)\left(1+L_3\right)+2L_2,\\r_{44}=\rho+\alpha L_2\left(1+L_3\right)^2.\end{cases}\tag{77}$$

然后求解以下等式：

$$Rt<t,\tag{78}$$

这等价于

$$\begin{aligned}&0<\gamma<\frac{\alpha(mt_1-L_1t_3-L_3t_4)}{(1-m\alpha)t_2}=\Gamma_1,\\&0<\gamma<\frac{t_2-\alpha L_1(1+L_3)t_1-\alpha L_1t_3-\alpha L_3t_4}{(2+L_3)t_2}=\Gamma_2,\\&0<\gamma<\frac{[1-\rho-\alpha L_1(L_3+1)]t_3-\alpha L_1L_3(2+L_3)t_1}{(L_3^2+4L_3+2)t_2}\\&\quad-\frac{\alpha L_3(L_3+1)t_4}{(L_3^2+4L_3+2)t_2}=\Gamma_3,\\&0<\gamma<\frac{[1-\rho-\alpha L_2(1+L_3)^2]t_4-\alpha L_1(L_2+1)(1+L_3)^2t_1}{L_2(1+L_3)t_2}\\&\quad-\frac{[\alpha L_1(L_2+1)(1+L_3)+2L_2]t_3}{L_2(1+L_3)t_2}=\Gamma_4.\end{aligned}\tag{79}$$

由上述不等式可得

$$
\begin{aligned}
&0<\alpha<\frac{t_2}{L_1(1+L_3)t_1+L_1t_3+L_3t_4}=\Theta_1,\\
&0<\alpha<\frac{1-\rho}{L_1(L_3+1)}=\Theta_2,\\
&0<\alpha<\frac{(1-\rho)t_3}{L_1L_3(2+L_3)t_1+L_1(L_3+1)t_3+(L_3^2+L_3)t_4}=\Theta_3,\\
&0<\alpha<\frac{1-\rho}{L_2L_3(1+L_3)}=\Theta_4,\\
&0<\alpha<\\
&\frac{(1-\rho)t_4-2L_2t_3}{L_1(L_2+1)(1+L_3)[(1+L_3)t_1+t_3]+L_2(1+L_3)^2t_4}=\Theta_5,\\
&t_1>\frac{L_1t_3+L_3t_4}{m},\\
&t_4>\frac{2L_2t_3}{1-\rho}.
\end{aligned}
\tag{80}
$$

也就是说，当.....时，我们可以任意选择 t_2 和 t_3

$$
0<\alpha<\bar{\alpha}=\min\{\Theta_1,\Theta_2,\Theta_3,\Theta_4,\Theta_5,\frac{1}{L_1}\} \tag{81}
$$

and

$$
0<\gamma<\bar{\gamma}=\min\{\Gamma_1,\Gamma_2,\Gamma_3,\Gamma_4,\frac{1}{L_2},\frac{1}{L_3}\}, \tag{82}
$$

其中 $t_2>0$、 $t_3>0$、 $t_1>\frac{L_1t_3+L_3t_4}{\mu}$和 $t_4>\frac{2L_2t_3}{1-\rho}$，我们有 $\rho(R)<1$。

接下来，我们使用 Jury判据来确定满足 $\rho(Q)<1$ 的 α 和 β 的精确范围。通过计算，我们可以得到 Q 的特征多项式:

$$
G(\lambda)=|\lambda I-Q|=a_0+a_1\lambda+a_2\lambda^2+a_3\lambda^3+\lambda^4, \tag{83}
$$

其中

$$
\begin{aligned}
a_0&=(e_1-1)[\alpha L_3(\rho\alpha L_1-e_2)+\rho^2e_3]+\rho^2\alpha\gamma e_1L_1(1+L_3),\\
a_1&=(e_1-1)[L_3e_2(1+r)-e_4\gamma+\rho(2e_3+\rho-\alpha L_3^2)]\\
&\quad+\gamma L_3[e_2+\alpha L_3(e_1+\rho e_1-1-2\rho)]\\
&\quad-\rho^2e_3-\alpha L_1(1+L_3)\rho(2\gamma e_1+\rho),\\
a_2&=(1+\gamma)L_3e_2+(e_1-1)[(1+\gamma)e_4+2\rho+e_3]\\
&\quad+\alpha L_1[(1+L_3)(2\rho+\gamma e_1)+L_3\gamma+L_3(1+\gamma)(e_1-1-\rho)]\\
&\quad-\gamma e_4+\rho(2e_3+\rho),\\
a_3&=e_1-e_2-1-2\rho-(1+\gamma)e_4-\alpha L_1[L_3(2+\gamma)+1],
\end{aligned}
\tag{84}
$$

以及 $e_1=\alpha[m+L_1(1+L_3)]$, $e_2=\alpha L_2[\rho(1+L_3)-2L_3]$, $e_3=\gamma(1+\alpha L_1+\alpha L_1L_3)$ $e_4=\alpha L_2L_3(1+L_3)$ ( ) 和 ( )。因此我们可以得到:

$$
\begin{aligned}
&b_0=a_0^2-1, b_1=a_0a_1-a_3, b_2=a_0a_2-a_2, b_3=a_0a_3-a_1\\
&c_0=b_0^2-b_3^2, c_1=b_0b_1-b_2b_3, c_2=b_0b_2-b_1b_3.
\end{aligned}
\tag{85}
$$

接下来记

$$
\begin{aligned}
&\mathcal{T}_1=\{(\alpha,\gamma)|a_0+a_1+a_2+a_3+1>0\},\\
&\mathcal{T}_2=\{(\alpha,\gamma)|a_0-a_1+a_2-a_3+1>0\},\\
&\mathcal{T}_3=\{(\alpha,\gamma)||a_0|<1\},\quad \mathcal{T}_4=\{(\alpha,\gamma)||b_0|>|b_3|\}\\
&\mathcal{T}_5=\{(\alpha,\gamma)||c_0|>|c_2|\},\quad \mathcal{T}_6=\{(\alpha,\gamma)|\alpha>0,\gamma>0\}.
\end{aligned}
\tag{86}
$$

6因此，根据引理4，当 $(\alpha,\ \gamma)\in\bigcap_{i=1}^{6}\mathcal{T}_i$ 时，矩阵 Q 的谱半径小于 1。根据上述分析，我们知道 $\bigcap_{i=1}^{6}\mathcal{T}_i$ 非空。最后，记 $\rho_2=\rho(Q)$ 和 $0<\rho_2<1$ 那么我们可以得到

$$
||V_{k+1}||\le\rho_2||V_k||. \tag{87}
$$

此外，可得

$$
||x_k-x^*||\le||V_k||\le C_2\rho_2^k, \tag{88}
$$

其中 $C_2=||V_0||$。因此，DAGT-NES 可以达到**R**线性收敛率。证明完毕。■

注2： 在定理 2 中，我们已经证明，当步长为 α，且最大动量参数 γ 满足式 (68) 时，DAGT-NES 具有 **R** 线性收敛速率。与 DAGT-HB 类似，定理 2 中 α 和 γ 的理论界是保守的。如何获得理论界乃至最优参数，将作为未来工作加以研究。

推论2： 在与定理 2 相同的假设下，以下等式成立:

$$
|f(x_k)-f(x^*)|\le\frac{L_1}{2}C_2^2\rho_2^{2k}. \tag{89}
$$

证明: 由于 $f(x)$ 是 L_1-光滑的，且 $\nabla f(x^*)=0$，我们可以得到

$$
f(x_k)-f(x^*)\le\frac{L_1}{2}||x_k-x^*||^2. \tag{90}
$$

将 (88) 代入 (90)，即可完成证明。 ■

作为本节的总结，DAGT-NES 方法表述如下算法2。

算法2： DAGT-NES

Input : initial point $x_{i,0}$ and $y_{i,0}\in\mathbb{R}^{n_i}$, $u_{i,0}=\phi_i(y_{i,0})$ and $s_{i,0}=\nabla_2 f_i(y_{i,0},u_{i,0})$ for $i=1,2,\ldots,N$, let $(\alpha,\gamma)\in\bigcap_{i=1}^{6}\mathcal{T}_i$, set $k=0$, select appropriate $\epsilon>0$ and $K_{max}\in\mathbb{N}^+$.

Output: optimal x^* and $f(x^*)$.

1 当 $k<K_{max}$ 且 $||\nabla f(x_k)||\ge\epsilon$ 时

2 迭代: 对每个 $i\in\{1,2,\ldots,N\}$进行更新:

$x_{i,k+1}$ iterated by formula (58),
$y_{i,k+1}$ iterated by formula (59),
$u_{i,k+1}$ iterated by formula (60),
$s_{i,k+1}$ iterated by formula (61).

3 更新: $k=k+1$。

V. 数值仿真

为了验证我们提出的 DAGT-HB 和 DAGT-NES 方法的有效性，我们通过以下最优布置进行数值仿真。在 $\mathbf{R}^2$,中的一个最优布置问题中，假设有 5 个实体，分别位于 $r_1=(10,\ 4)$、$r_2=(1,\ 3)$、 $r_3=(2,\ 7)$、 $r_4=(8,\ 10)$ 和 $r_5=(3,\ 9)$。此外，还有 5 个自由实体，每个自由实体都可以单独与某个固定实体建立连接。我们的目的是确定自由实体的最优位置x_i, $i\in\{1,2,3,4,5\}$，从而最小化每个自由实体当前位置到相应固定实体位置的距离之和，以及每个实体到所有自由实体加权中心的距离。因而，每个自由实体的成本函数可建模如下:

$$
f_i(x_i,u(x_i))=\omega_i||x_i-r_i||^2+||x_i-u(x)||^2,\quad i=1,2,\ldots,5, \tag{91}
$$

其中, $\sum_{i=1}^{5}x_i$ ϕ_i $i=$ 表示权重,

其值设为20。我们设为 0 。因此，在此条件下， 是 的恒等映射

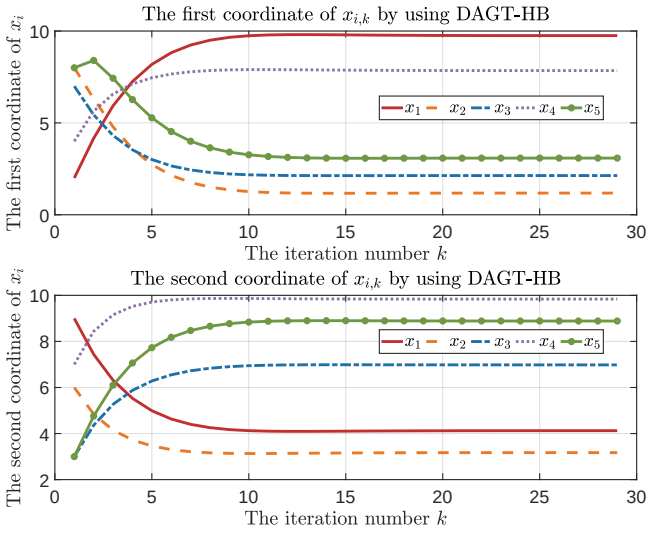

图 1. 使用 DAGT-HB 时$\boldsymbol{x_{i,\ k}}$ 的演化。

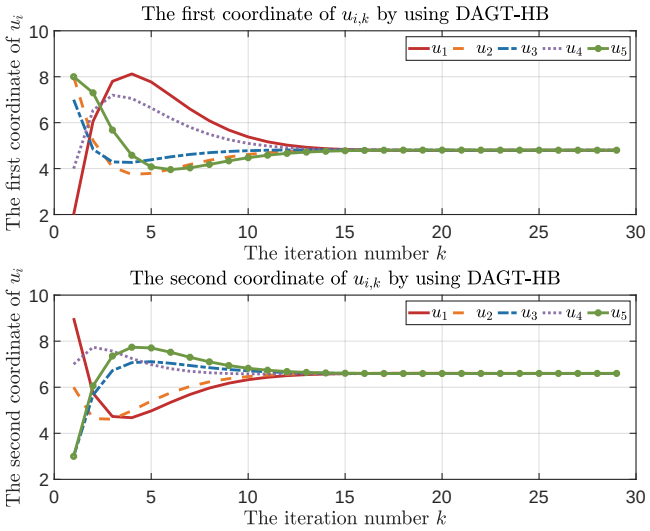

图 2. 使用 DAGT-HB 时$\boldsymbol{u_{i,\ k}}$ 的演化。

1,. .., 5。通信图随机选取为强连通且双随机。

我们选择初始点 $x_{1,-1}$ =(0, 11)、$x_{2,-1}$ =(9, 8)、$x_{3,-1}$ = (9, 1)、$x_{4,-1}$ =(1, 4) 和 $x_{5,-1}$ =(3, 1)；$x_{1,0}$ =$y_{1,0}$ =(2, 9)、$x_{2,0}$ = $y_{2,0}$ =(8, 6)、$x_{3,0}$ = $y_{3,0}$ =(7, 3)、$x_{4,0}$ = $y_{4,0}$ =(4, 7) 和 $x_{5,0}$ = $y_{5,0}$ =(8, 3)。对于 i = 1, 2,. .. , 5，我们将 $u_{i,0}$ = ϕ_i (x_i,0) 和 $s_{i,0}$ = $\nabla_2 f_i(x_i,0,u_i,0)$ 初始化，并设置步长 α = 0.005，分别选择动量项 $\beta = 0.28$ 和 $\gamma = 0.25$。

然后，我们使用 DAGT-HB 方法求解最优布置问题，结果如图 1 和图 2 所示。在图 1 中可以看到，所有智能体都能收敛到其最优位置，最优位置分别为 x_1^* =(9.7524, 4.1248)、x_2^* = (1.810,3.1714)、x_3^* =(2.1333,6.9810)、x_4^* =(7.8416, 9.8381) 和 x_5^* =(3.0857, 8.8857)。图 2 展示了 $u_{i,k}$ 的演化过程，表明每个自由实体的估计值 $u_{i,k}$ 都以较快的速度收敛到最优聚合位置 $u(x^*)$ =(4.8,6.6)。此外，图 1 和图 2 表明收敛率较快，从而支持我们的理论分析。同时，DAGT-NES 的 $x_{i,k}$ 和 $u_{i,k}$ 的演化过程与 DAGT-HB 的类似，结果如图 3 和图 4 所示。

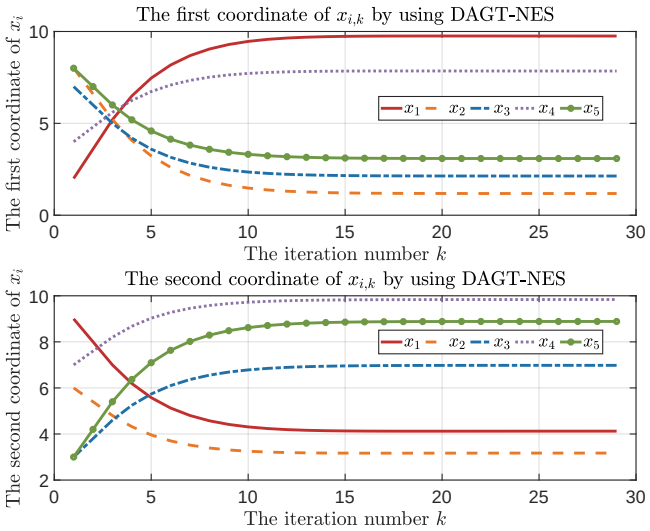

图 3. 使用 DAGT-NES 时$\boldsymbol{x_{i,\ k}}$ 的演化。

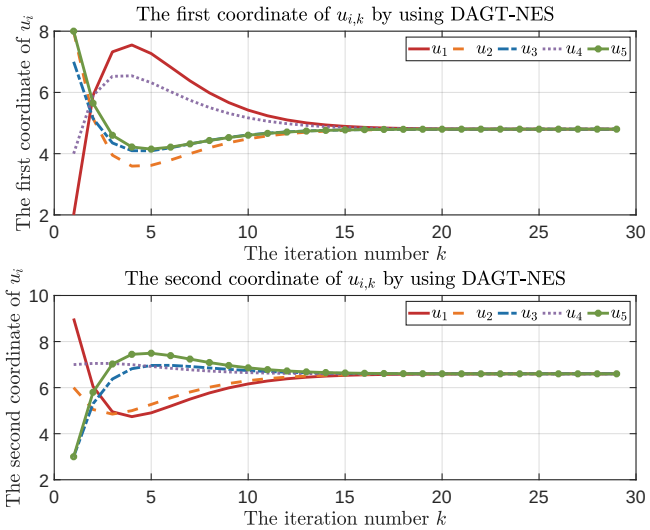

图 4. 使用 DAGT-NES 时$\boldsymbol{u_{i,\ k}}$ 的演化。

为了展示 DAGT-HB 和 DAGT-NES 的优越性，我们在相同的初始条件下将它们与 DAGT 进行比较。以第一个状态 x_1 为例，第一个状态的误差如图 5 所示。我们发现，DAGT-HB 和 DAGT-NES 的收敛速度明显快于 DAGT，这表明引入的动量项 $x_{i,k} - x_{i,k-1}$ 能够提高算法的收敛率。此外，为了与 DAGT 的收敛情况进行比较，图 6 展示了 DAGT、DAGT-HB 和 DAGT-NES 的损失 $(f(x_k) - f(x_*))^2$。这一结果进一步体现了我们算法的优越性。同时，我们发现 DAGT-HB 的收敛率更快，但如图 5 所示，可能会引起更多振荡。振荡产生的原因以及如何减少振荡，值得在未来研究中进一步探讨。

VI. 结论

本文提出了两种新颖的算法，称为 DAGT-HB 和 DAGT-NES，用于求解网络中的分布式聚合优化问题。受加速算法的启发，我们将重球法和 Nesterov 加速方法与分布式聚合梯度跟踪方法相结合。此外，我们证明，

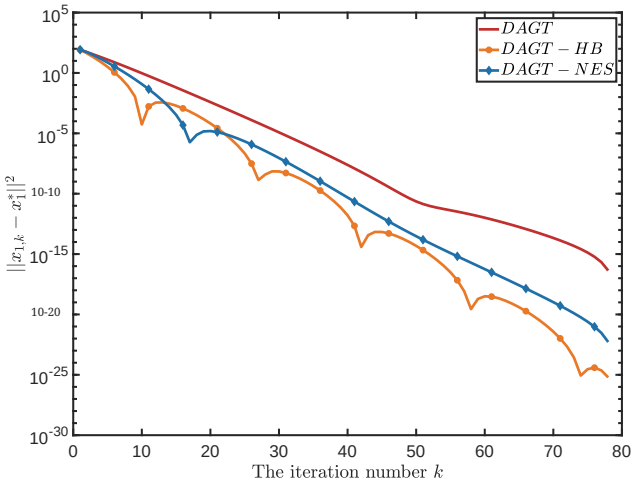

图5. 状态误差 $||x_{1,k}-x_1^*||^2$ DAGT、DAGT- HB和DAGT-NES之间的比较。

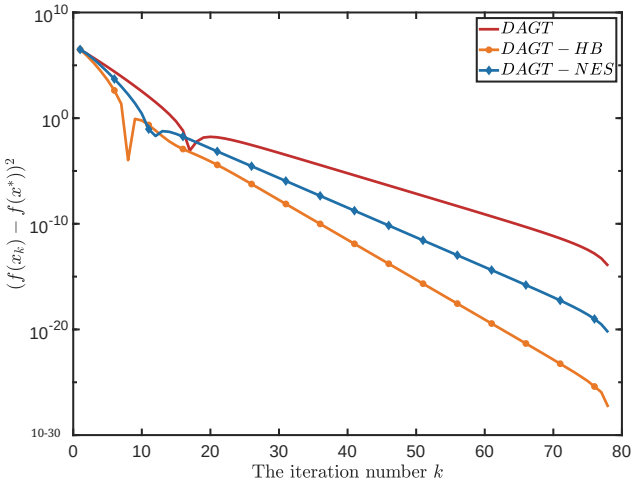

图6. DAGT、DAGT-HB和DAGT-NES的代价函数误差 $||f(x_k)-f(x^*)||^2$ 比较。

当目标函数光滑且强凸，并且步长和动量项选择得当时，DAGT-HB 和 DAGT-NES 算法能够以全局$\mathbf{R}$–线性收敛率收敛到最优解。最后，我们以最优布置问题为例，验证 DAGT-HB 和 DAGT-NES 的有效性和优越性。在相同条件下，与原始 DAGT 方法相比，DAGT-HB 和 DAGT-NES 能使状态误差和成本函数更快收敛。此外，我们发现 DAGT-HB 的收敛率高于 DAGT-NES，但可能会引起更多振荡。未来的研究可关注振荡产生的原因、DAGT-HB 和 DAGT-NES 中参数选择的敏感性，以及将 DAGT-HB 和 DAGT-NES 扩展到非平衡图、非凸目标函数和带约束的分布式聚合优化问题。

VII. 附录

A. 引理 10 的证明

将（58）-（61）的平衡点记为 x^* =col（x_1^*，……，x_N^*）、y^* = col（y_1^*，……，y_N^*）、u^* =col（u_1^*，……，u_N^*）和s^* =col（s_1^*，……，s_N^*）。根据（59），我们知道 $y^*=x^*$。接下来，由（58）可得，(60) 和 (61)，我们可以推导出

$$\nabla_1 f\left(y^*,u^*\right)+\nabla\phi\left(y^*\right)s^*=\mathbf{0}_{Nd}, \tag{92}$$
$$\mathcal{L}u^*=\mathbf{0}_{Nd},\quad \mathcal{L}s^*=\mathbf{0}_{Nd}, \tag{93}$$

其中，$\mathcal{L}=L\otimes I_d$ 和 L 是图 $\mathcal{G}$的拉普拉斯矩阵。由于拉普拉斯矩阵的性质，对于所有 $i\neq j$，u_i^* 必等于 u_j^*，且s_i^* 必等于 s_j^*。由等式 (62) 和 (63)，可得

$$u_i^*=\frac{1}{N}\sum_{i=1}^{N}\phi_i\left(y_i^*\right)=u(y^*), \tag{94}$$

$$s_i^*=\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y_i^*,u(y^*)\right). \tag{95}$$

将 (94) 和 (95) 代入 (92)，可得 $\nabla f(x^*)$ $=$ $\nabla f\ y^*\ =0$ f $m-$ x^* ()。由于具有强凸性，因此它是问题 (1) 的唯一最优解。

B. 引理 11 的证明

对于 $\|x_{k+1}-x^*\|$，利用 (58)，可得

$$\begin{aligned}
&\left\|x_{k+1}-x^*\right\|\\
&=\left\|y_k-y^*-\alpha\left[\nabla_1 f\left(y_k,u_k\right)+\nabla\phi\left(y_k\right)s_k\right]\right\|\\
&\leq\Bigg\|y_k-y^*-\alpha\Big[\nabla_1 f\left(y_k,\mathbf{1}_N\otimes\bar{u}_k\right)\\
&\quad+\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\Big]+\alpha\nabla f\left(y^*\right)\Bigg\|\\
&\quad+\alpha\|\nabla_1 f\left(y_k,u_k\right)+\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\bar{s}_k-\nabla_1 f\left(y_k,\mathbf{1}_N\otimes\bar{u}_k\right)\\
&\quad-\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\|\\
&\quad+\alpha\left\|\nabla\phi\left(y_k\right)s_k-\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\bar{s}_k\right\|.
\end{aligned} \tag{96}$$

根据引理 1，我们可以对 (96) 右侧第一项作如下界定：

$$\begin{aligned}
&\Bigg\|y_k-y^*-\alpha\Big[\nabla_1 f\left(y_k,\mathbf{1}_N\otimes\bar{u}_k\right)\\
&\quad+\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\Big]+\alpha\nabla f\left(y^*\right)\Bigg\|\\
&\leq(1-m\alpha)\left\|y_k-y^*\right\|.
\end{aligned} \tag{97}$$

然后，将 (59) 代入 (97)，并注意到 $x^*=y^*$，可得

$$\left\|y_k-y^*\right\|\leq\left\|x_k-x^*\right\|+\gamma\left\|x_k-x_{k-1}\right\|. \tag{98}$$

对于第二项，由于 $f(x)$ 是 L_1–光滑的，且 $\mathbf{1}_N\otimes\bar{u}_k=\mathcal{K}u_k$，我们可以得到

$$\begin{aligned}
&\alpha\|\nabla_1 f\left(y_k,u_k\right)+\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\bar{s}_k-\nabla_1 f\left(y_k,\mathbf{1}_N\otimes\bar{u}_k\right)\\
&\quad-\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y_{i,k},\mathbf{1}_N\otimes\bar{u}_k\right)\|\\
&\leq\alpha L_1\left\|u_k-\mathcal{K}u_k\right\|.
\end{aligned} \tag{99}$$

对于最后一项，利用假设4和 $\mathbf{1}_N\otimes\bar{s}_k=\mathcal{K}s_k$，可得以下不等式：

$$\alpha\left\|\nabla\phi\left(y_k\right)s_k-\nabla\phi\left(y_k\right)\mathbf{1}_N\otimes\bar{s}_k\right\|\leq\alpha L_3\left\|s_k-\mathcal{K}s_k\right\|. \tag{100}$$

然后，将 (97)–(100) 与 (96) 结合，我们便完成了证明。

C. 引理12的证明

对于 $\|x_{k+1}-x_k\|$，我们应用 (58)，并注意到 $\nabla f(x^*)=0$，于是得到

$$\begin{aligned}&\|x_{k+1}-x_k\|\\&=\|\gamma(x_k-x_{k-1})-\alpha(\nabla_1 f(y_k,u_k)+\nabla\phi(y_k)s_k)\|\\&\le\alpha\Big\|\nabla_1 f(y_k,u_k)+\nabla\phi(y_k)\mathcal{K}s_k-\nabla_1 f\left(y^*,\mathbf{1}_N\otimes u^*\right)\\&\quad-\nabla\phi(x^*)\left[\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y^*,\mathbf{1}_N\otimes u^*\right)\right]\Big\|\\&\quad+\alpha\|\nabla\phi(y_k)(s_k-\mathcal{K}s_k)\|+\gamma\|x_k-x_{k-1}\|.\end{aligned}\tag{101}$$

利用假设2和范数的三角不等式，可得以下公式:

$$\begin{aligned}&\Big\|\nabla_1 f(y_k,u_k)+\nabla\phi(y_k)\mathcal{K}s_k-\nabla_1 f\left(y^*,\mathbf{1}_N\otimes u^*\right)\\&\quad-\nabla\phi(y^*)\left[\mathbf{1}_N\otimes\frac{1}{N}\sum_{i=1}^{N}\nabla_2 f_i\left(y^*,\mathbf{1}_N\otimes u^*\right)\right]\Big\|\\&\le L_1\left(\|y_k-x^*\|+\|u_k-\mathbf{1}_N\otimes u^*\|\right)\\&\le L_1\left(\|x_k-x^*\|+\|u_k-\mathcal{K}u_k\|\right)\\&\quad+\gamma L_1\|x_k-x_{k-1}\|+L_1\|\mathcal{K}u_k-\mathbf{1}_N\otimes u^*\|.\end{aligned}\tag{102}$$

由（30）可知

$$\begin{aligned}\|\mathcal{K}u_k-1_N\otimes u^*\|&\le L_3\|y_k-y^*\|\\&\le L_3\|x_k-x^*\|+L_3\gamma\|x_k-x_{k-1}\|.\end{aligned}\tag{103}$$

然后利用假设4可得

$$\|\nabla\phi(y_k)(s_k-\mathcal{K}s_k)\|\le L_3\|s_k-\mathcal{K}s_k\|.\tag{104}$$

最后，将（102）–（104）代入（101），即可得到引理12。

D. 引理 13 的证明

对于 $\|u_{k+1}-\mathcal{K}u_{k+1}\|$，通过调用 (60)，可得

$$\|u_{k+1}-\mathcal{K}u_{k+1}\|=\|\mathcal{A}u_k+\phi(y_{k+1})-\phi(y_k)-\mathcal{K}\mathcal{A}u_k-\mathcal{K}[\phi(y_{k+1})-\phi(y_k)]\|\le\rho\|u_k-\mathcal{K}u_k\|+\|I-\mathcal{K}\|\|\phi(y_{k+1})-\phi(y_k)\|\le\rho\|u_k-\mathcal{K}u_k\|+L_3\|y_{k+1}-y_k\|=\rho\|u_k-\mathcal{K}u_k\|+L_3\|(1+\gamma)(x_{k+1}-x_k)-\gamma(x_k-x_{k-1})\|\le\rho\|u_k-\mathcal{K}u_k\|+L_3(1+\gamma)\|x_{k+1}-x_k\|+\gamma L_3\|x_k-x_{k-1}\|,\tag{105}$$

其中第一条不等式利用了引理2，而利用假设4可得第三条不等式。然后

substituting (65) i nto (105) 我们可以 o btain the Lemma 13.

E. 引理 14 的证明

对于 $\|s_{k+1}-\mathcal{K}s_{k+1}\|$，利用 (61) 可得

$$\begin{aligned}&\|s_{k+1}-\mathcal{K}s_{k+1}\|\\&=\|\mathcal{A}s_k+\nabla_2 f(y_{k+1},u_{k+1})-\nabla_2 f(y_k,u_k)\\&\quad-\mathcal{K}\mathcal{A}s_k-\mathcal{K}[\nabla_2 f(y_{k+1},u_{k+1})-\nabla_2 f(y_k,u_k)]\|\\&\le\|\mathcal{A}s_k-\mathcal{K}s_k\|\\&\quad+\|I-\mathcal{K}\|\|\nabla_2 f(y_{k+1},u_{k+1})-\nabla_2 f(y_k,u_k)\|\\&\le\rho\|s_k-\mathcal{K}s_k\|+\|\nabla_2 f(y_{k+1},u_{k+1})-\nabla_2 f(y_k,u_k)\|\\&\le\rho\|s_k-\mathcal{K}s_k\|+L_2\left(\|y_{k+1}-y_k\|+\|u_{k+1}-u_k\|\right),\end{aligned}\tag{106}$$

其中，最后一个不等式用到了假设2。然后由（36）可知

$$\|u_{k+1}-u_k\|\le 2\|u_k-\mathcal{K}u_k\|+L_3\|y_{k+1}-y_k\|.\tag{107}$$

将（107）代入（106），可得

$$\begin{aligned}&\|s_{k+1}-\mathcal{K}s_{k+1}\|\\&<\rho\|s_k-\mathcal{K}s_k\|+L_2(L_3+1)\|y_{k+1}-y_k\|+2L_2\|u_k-\mathcal{K}u_k\|\\&\le\rho\|s_k-\mathcal{K}s_k\|+L_2(L_3+1)(\gamma+1)\|x_{k+1}-x_k\|\\&\quad+L_2(L_3+1)\gamma\|x_k-x_{k-1}\|+2L_2\|u_k-\mathcal{K}u_k\|.\end{aligned}\tag{108}$$

然后将 (65) 代入 (108)，即可得到引理 14 .

参考文献

[1]Jeffrey Dean、Greg Corrado、Rajat Monga、Kai Chen、Matthieu Devin、Mark Mao、Marc'aurelio Ranzato、Andrew Senior、Paul Tucker、Ke Yang 等。大规模分布式深度网络。神经信息处理系统进展，25，2012。[2] Sergio Barbarossa、Stefania Sardellitti 和 Paolo Di Lorenzo。边计算边通信：5G异构网络上的分布式移动云计算。IEEE信号处理杂志，31(6):45–55，2014。[3] Mu Li、David G Andersen、Alexander J Smola 和 Kai Yu。利用参数服务器实现通信高效的分布式机器学习。神经信息处理系统进展，27，2014。[4] Joel B. Predd、Sanjeev R. Kulkarni 和 H. Vincent Poor。分布式学习的协同训练算法。IEEE信息论汇刊，55:1856–1871，2009。[5] Wei Wang、Jiangshuai Huang、Changyun Wen 和 Huijin Fan。用于一致性跟踪的分布式自适应控制及其在非完整移动机器人编队控制中的应用。Automatica，50(4):1254–1263，2014。[6] Shanying Zhu、Cailian Chen、Wenshuang Li、Bo Yang 和 Xinping Guan。异构传感器网络中用于目标跟踪的分布式最优一致性滤波器。IEEE控制论汇刊，43(6):1963–1976，2013。[7] Zhenhua Deng、Shu Liang 和 Yiguang Hong。权重平衡有向图上资源分配问题的分布式连续时间算法。IEEE控制论汇刊，48:3116–3125，2018。[8] Angelia Nedic 和 Asuman Ozdaglar。多智能体优化的分布式次梯度方法。IEEE 自动控制汇刊，54(1):48–61，2009。[9] John C. Duchi、Alekh Agarwal 和 Martin J. Wainwright。分布式优化的对偶平均：收敛性分析与网络扩展。IEEE 自动控制汇刊，57:592–606，2010。[10] Konstantinos I. Tsianos、Sean F. Lawlor 和 Michael G. Rabbat。用于凸优化的 push-sum 分布式对偶平均。2012年 IEEE 第51届 IEEE决策与控制会议（CDC），第5453–5458页，2012。[11] Kun Yuan、Qing Ling 和 Wotao Yin。论去中心化梯度下降的收敛性。SIAM优化杂志，26:1835–1854，2016。[12] Jinming Xu、Shanying Zhu、Yeng Chai Soh 和 Lihua Xie。非协调常数步长下多智能体优化的增广分布式梯度方法。载于：2015年 IEEE 第54届 IEEE决策与控制会议（CDC），第2055–2060页。IEEE，2015。[13] Angelia Nedic、Alexander Olshevsky 和 Wei Shi。时变图上分布式优化的几何收敛实现。SIAM优化杂志，27:2597–2633，2016。[14] Wei Shi、Qing Ling、Gang Wu 和 Wotao Yin。EXTRA：用于去中心化一致性优化的精确一阶算法。SIAM优化杂志，25(2):944–966，2015。[15] Boris T Polyak。加速迭代方法收敛的一些方法。苏联计算数学与数学物理，4(5):1–17，1964。[16] Yurii Evgen'evich Nesterov。求解具有收敛率 o($\frac{1}{k^2}$) 的凸规划问题的一种方法。载于《科学院院报》，第269卷，第543–547页。俄罗斯科学院，1983。[17] Du˘san Jakoveti´c、Joao Xavier 和 Jos´e MF Moura。快速分布式梯度方法。IEEE 自动控制汇刊，59(5):1131–1146，2014。[18] Ran Xin、Du˘san Jakoveti´c 和 Usman A Khan。任意图上的分布式 Nesterov 梯度方法。IEEE信号处理快报，26(8):1247–1251，2019。[19] Ran Xin 和 Usman A. Khan。用于有向图优化且具有几何收敛性的线性算法。IEEE控制系统快报

，2(3):315–320 ，2018。

[20] Ran Xin 和 Usman A. Khan。分布式重球法：一种结合梯度跟踪的一阶方法推广与加速。*IEEE* 自动控制汇刊，65(6):2627–2633，2020。[21] Guannan Qu 和 Na Li。加速分布式 Nesterov 梯度下降法。*IEEE* 自动控制汇刊，65 2581，2020 6 2566( ): – 。[22] Qingguo L¨u、Xiaofeng Liao、Huaqing Li 和 Tingwen Huang。一种用于有向网络分布式优化的类 Nesterov 梯度跟踪算法。*IEEE* 系统、人和控制论汇刊：系统，51(10):6258–6270，2020。[23] Guido Carnevale、Nicola Mimmo 和 Giuseppe Notarstefano。用于分布式协作机器人学的聚合反馈优化。*IFAC- PapersOnLine*，55(13):7–12，2022。[24] Jayash Koshal、Angelia Nedi´c 和 Uday V Shanbhag。图上的聚合博弈分布式算法。《运筹学》，64(3):680–704，2016。[25] Xiuxian Li、Lihua Xie 和 Yiguang Hong。多智能体网络上的分布式聚合优化。*IEEE* 自动控制汇刊，67(6):3165–3171，2022。[26] Xiuxian Li、Xinlei Yi 和 Lihua Xie。具有聚合变量的分布式在线凸优化。*IEEE* 网络系统控制汇刊，9(1):438–449，2022。[27] Guido Carnevale、Andrea Camisa 和 Giuseppe Notarstefano。动态多机器人协调的分布式在线聚合优化。*IEEE* 自动控制汇刊，第 1–8 页，2022。[28] Ziqin Chen 和 Shu Liang。采用量化通信的分布式聚合优化。*arXiv* 预印本 *arXiv:2201.01969*，2022。[29] Tongyu Wang 和 Peng Yi。约束聚合优化的分布式无投影算法。*arXiv* 预印本 *arXiv:2207 11885*，2022。。[30] Yurii Nesterov。《凸优化导论：基础课程》，第 87 卷。施普林格科学与商业媒体，2003。[31] James M Ortega 和 Werner C Rheinboldt。《多变量非线性方程的迭代解法》。工业与应用数学学会，2000。[32] Roger A Horn 和 Charles R Johnson。《矩阵分析》。剑桥大学出版社，2012。[33] Baodong Zheng、Lijie Liang 和 Chunrui Zhang。扩展 Jury 判据。《中国科学：数学》，53:1133–1150，2010。